INTEGRATING LINKED OPEN DATA IN MOBILE AUGMENTED REALITY APPLICATIONS

Teză destinată obţinerii
titlului ştiinţific de doctor inginer
la
Universitatea Politehnica Timişoara
în domeniul
INGINERIE ELECTRONICĂ ȘI TELECOMUNICAȚII
de către

Ing. Silviu Vert

Conducător ştiinţific:	prof.univ.dr.ing. Radu Vasiu
Referenţi ştiinţifici:	prof.univ.dr.ing. Marco Sajeva
	prof.univ.dr.ing. Aurel Vlaicu
	prof.univ.dr.ing. Vladimir Crețu

Ziua susţinerii tezei: 18.09.2015

Seriile Teze de doctorat ale UPT sunt:

1. Automatică
2. Chimie
3. Energetică
4. Ingineria Chimică
5. Inginerie Civilă
6. Inginerie Electrică
7. Inginerie Electronică şi Telecomunicaţii
8. Inginerie Industrială
9. Inginerie Mecanică
10. Ştiinţa Calculatoarelor
11. Ştiinţa şi Ingineria Materialelor
12. Ingineria sistemelor
13. Inginerie energetică
14. Calculatoare şi tehnologia informaţiei
15. Ingineria materialelor
16. Inginerie şi Management

Universitatea Politehnica Timişoara a iniţiat seriile de mai sus în scopul diseminării expertizei, cunoştinţelor şi rezultatelor cercetărilor întreprinse în cadrul Şcolii doctorale a universităţii. Seriile conţin, potrivit H.B.Ex.S Nr. 14 / 14.07.2006, tezele de doctorat susţinute în universitate începând cu 1 octombrie 2006.

România, 300159 Timişoara, Bd. Republicii 9,
Tel./fax 0256 403823
e-mail: editura@edipol.upt.ro

Foreword

This thesis was written during my doctoral studies, which took place within the Department of Communications from the Politehnica University of Timisoara.

The research tackles fresh aspects regarding the accessing, the processing and the integration of various open data that one can find on the World Wide Web today. The particular purpose of the resulting integrated data is to be exploited in mobile augmented reality applications for tourism.

Based on the critical study of the research domains that are approached in this thesis, I propose a model and describe the implementation of such a mobile augmented reality application for tourists in Romania and Trento, Italy. I explore the challenges that show up during the integration of open government data and user-generated data for such a purpose and analyze the profile of this type of data. The exploited technologies and tools, the used approaches and the drawn conclusions are described in detail in the thesis.

I enjoyed delving into these issues, as they perfectly overlap with my core interests, which I also explored in "extracurricular" activities. As Ambassador for Romania on behalf of the Open Knowledge Foundation, I promote open data and open knowledge in my country; as co-founder of the Smart City Association, I get engaged in activities promoting the use of open government data for smart city applications, including augmented reality-based ones; and as part of the team that is working for the candidacy of Timisoara as European Capital of Culture in 2021, I work on projects which combine digital, culture and tourism.

As part of my doctoral studies, I got a 3-month Erasmus internship in Palermo, Italy, where I explored in more depth the area of mobile applications for tourism. I am grateful to Professor Marco Sajeva for his hospitality and guidance during my staying there, which was one of the most productive periods of my doctoral research.

I acknowledge the financial support that I received towards the end of my doctoral studies from the strategic grant POSDRU/159/1.5/S/137070 of the Ministry of National Education, Romania, co-financed by the European Union.

Several thanks go to the team of the Multimedia Research Centre, with whom I worked closely and from whom I learned a lot during the years spent at the university. Special thanks go to my colleague Bogdan Drăgulescu, whose expertise and willingness to help have empowered me to overcome a lot of blocking points in my research.

I am extremely grateful to Professor Radu Vasiu, my PhD scientific coordinator, for his patience, valuable guidance and openness towards doing research on issues that I profoundly care about, while keeping an eye on the scientific relevance of my endeavor.

Finally, I am deeply grateful to my parents, Silvia and Constantin, and my brother, Sorin, for their understanding and support, and to my fiancée, Adriana, for her love and encouragements, and for sacrificing countless moments of quality time together for me to be able to write *just* another paragraph for the thesis.

Timişoara, september 2015 Silviu Vert

For my family.

This work was partially supported by the strategic grant POSDRU/159/1.5/S/137070 (2014) of the Ministry of National Education, Romania, co-financed by the European Social Fund – Investing in People, within the Sectoral Operational Programme Human Resources Development 2007-2013.

Vert, Silviu

Integrating linked open data in mobile augmented reality applications

Teze de doctorat ale UPT, Seria 7, Nr. 80, Editura Politehnica, 2015, 132 pagini, 22 figuri, 9 tabele.

ISSN:1842-7014

ISBN:978-606-554-998-2

Cuvinte cheie:
Realitate augmentată, date guvernamentale deschise, date interconectate, aplicații mobile turistice

Rezumat:
Teza de față abordează aspecte actuale privind preluarea, prelucrarea și integrarea diverselor date deschise existente în acest moment pe World Wide Web. Datele obținute sunt destinate exploatării în aplicații mobile, create în scop turistic, ce folosesc tehnologii de realitate augmentată.

Studiul critic al evoluției și stadiului actual al tehnologiilor de realitate augmentată, împreună cu analiza maturității principiilor și uneltelor de publicare și consum a datelor interconectate, au scos la iveală avantajele integrării datelor deschise interconectate în aplicații mobile de realitate augmentată pentru turism, dar și provocările ridicate de o astfel de integrare.

Autorul propune un model și descrie implementarea unui prototip de aplicație mobilă, care integrează atât date deschise guvernamentale cât și date deschise generate de utilizatori și care folosește realitatea augmentată bazată pe tehnologii web pentru a ajuta un turist să descopere puncte de interes în jurul său.

Este inclusă de asemenea și o analiză a profilului datelor integrate, împreună cu o comparație a aplicației dezvoltate cu alte aplicații similare.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AR	Augmented Reality
CSV	Comma-Separated Values
DOI	Digital Object Identifier
HTTP	Hypertext Transfer Protocol
LOD	Linked Open Data
OWL	Ontology Web Language
POI	Point of Interest
RDFS	RDF Schema
RDF	Resource Description Framework
SEO	Search Engine Optimization
SKOS	Simple Knowledge Organization System
URI	Uniform Resource Identifier
WWW	World Wide Web
XML	Extensible Markup Language

1. MOTIVATION

The first chapter is an exposition of my motivation towards dealing with the subject of the thesis. Accordingly, it shows the relevance and the timeliness of the explored issues and it puts forward the assumptions made within the thesis. The chapter also lists the scientific publications (conference articles, journal articles, book chapters) that support my research activity and it concludes by explaining the structure of the thesis.

1.1. General overview of the selected subject

Featuring Science Fiction-like characteristics, augmented reality (AR) is a set of technologies that already has a solid history of research and that is undoubtedly here to stay. Its strength and novelty lies in its ability to improve our perception of the real world, ideally seemingly, by extending it with digital information, which enables us to perceive hidden information around us, to interact with our sorroundings in a novel and useful way, and possibily doing this in a collaborative way.

To get a grasp of the extent that AR can and is influencing our lives, I list here some domains in which AR has proven useful: military (seeing hidden information on the enemy territory), medicine (AR-based surgery), entertainment (AR-based games), tourism (exploring an unknown place), marketing (AR-based ads) and so on.

Aside from very focused implementations, AR has also developed towards general public use, in the form of AR browsers, which aim to help regular users in their various daily tasks: showing the route to a new place, finding a restaurant nearby to eat, getting information for a product that one buys at the supermarket or interactively reading a magazine. However, for various reasons, including privacy and legal issues, toghether with reluctance to invasive technology and to awkward wearables, the AR browsers have not become as mainstream as it was predicted.

Technical reasons, also very important, include poor registration with the real world, not enough quantity and quality of the content, proprietary architectures and data formats, and bad user experiences [1].

In my opinion, AR has slowed down its evolution as the superstar of future technologies and is in a phase in which it is again searching for its core purpose. Google seems to rethink its plans with Google Glasses, turning them from ubiquitous devices to devices designed for very specific use cases (though Google Glasses were never a true AR device). Microsoft seems to have skipped this transition and is directly launching its AR-based devices (the HoloLens) with specific use cases in mind.

For the smaller - but AR focused and dedicated - players, the battle for the market has strengthened and it seems that the stronger and more resilient will take all (or most). Wikitude and Metaio[1] are continuing to invest heavily in developing a full range of AR software tools, which suits both publishers and consumers, while the third big traditional player, Layar[2], turned its focus on probably the most profitable business case of AR, the interactive printing one.

One of the shortcomings of today's AR browsers, mentioned earlier, is the content, which is rather static i.e. the browsing experience is a quite plain one, as the user can see Points of Interest around her, access them for some more information and click a link that sends them in another application (usually the mobile browser) for the full description. This aspect has attracted my attention the most, due to personal activities and interests detailed in the following paragraphs.

I am actively involved in the open data phenomenon, which is data that can be freely used, reused and redistributed by anyone - subject only, at most, to requirements that preserve provenance and openness [4]. Open data has been around since the beginning of user-generated data portals, but has grown significantly in popularity due to governments worldwide starting to publish non-personal, non-secret data as open government data.

Since May 2014, I am the Ambassador for Romania of the Open Knowledge Foundation, which is the biggest non-profit network in the world that promotes open data through advocacy, technology and trainings. The biggest projects undertaken by the Romanian group, which I lead, were related to managing the Open Data Census for Romania and the City Open Data Census for the cities in Romania, which aim to analyze, score and compare countries and cities based in their openness.

In addition, I have been actively involved in the coordination of the Smart City Association, which started as an informal community in the summer of 2013 and later became a formal organization. The association engages with universities, companies, public institutions and IT communities to develop applications for smart cities based on open data. The association won the prize, awarded by the Romanian government, for the NGO that best promoted transparency to the public administration during its activity in 2014.

I was involved in a smart city commitment [5] that was signed in 2014 by the Politehnica University of Timisoara, the Timisoara City Hall and the Smart City Association, in which the entities agreed to work together, in the long term, on a smart city concept for Timisoara, focused on open data, under the "European Innovation Partnership on Smart Cities and Communities". Other activities in the area include organizing several hackathons for open data and smart city applications and giving talks or trainings on these issues.

Given the focus and nature of my activity, I am a firm believer that exploiting such data, whether governmental or user-generated one, can significantly enhance applications that rely on good and extensive content, such as augmented reality apps. Integrating such data into AR browsers would lead to better, more dynamic and aware, an ultimately, useful AR browsers.

However, putting together such data is not a trivial task, due to various tools, methods and standards used to publish open data worldwide. Regardless of

[1] Metaio was acquired by Apple in May 2015 [2]

[2] Layar was acquired by Blippar in June 2014 [3]

the exploitation mode of the data, the most appropriate way for dealing with heterogeneous integration of data is using the Semantic Web principles.

The Semantic Web, through its more practical part, Linked Data, proposes tools and principles which have been proven as being mature enough to deal with this vast amount of data.

Although some projects have tackled the integration of linked open data in AR applications, I have found that there is a gap in terms of a clear, straightforward model for enhancing AR apps through these means.

As a result, in this thesis I plan to study such integration, by applying it to a tourism use case in my hometown, Timisoara. I exploit the fact that Timisoara is a candidate city for the European Capital of Culture in 2021; therefore, smart applications for tourists attract the attention of the public and benefit from help and feedback from citizens and authorities.

I plan to answer the following research questions:

1. What is the current status of AR browsers in terms of content for tourism?

2. How mature and appropriate are Linked Data principles and tools for enhancing AR applications?

3. What are the challenges in integrating linked data in mobile augmented reality applications?

4. Is there a model for this integration and what are its particularities?

5. How appropriate are the content and the structure of linked open data for augmented reality applications?

1.2. List of published articles

My research activity, in the field of the thesis or related areas, has been backed up by the following list of published articles, both as author or co-author:

Silviu Vert and Diana Andone, "A Review of Linked Open Educational Resources", Springer, 2015 (*to be published*).

Silviu Vert, Radu Vasiu, "Integrating Linked Open Data in Mobile Augmented Reality Applications - a Case Study", TEM JOURNAL - Technology, Education, Management, Informatics, ISSN 2217-8309, e-ISSN 2217-8333, vol. 4, no. 1, Feb. 2015, pp. 35-43.

Daniel Rusu, **Silviu Vert**, "City Alerts: Smart City Notification Platform Based on Public Open Data", Scientific Bulletin of the Politehnica University of Timisoara - Transactions on Electronics and Communications, ISSN 1583-3380, Vol. 59(73), No. 2, Dec. 2014, pp. 21-26.

Silviu Vert, Bogdan Dragulescu, Radu Vasiu, "LOD4AR: Exploring Linked Open Data with a Mobile Augmented Reality Web Application", in Proceedings of the 13th International Semantic Web Conference (ISWC 2014), Riva del Garda, Italy, 19-23 October 2014, ISSN 1613-0073, Vol-1272, pp. 185-188.

Silviu Vert, Radu Vasiu, "Integrating Linked Data in Mobile Augmented Reality Applications", in Proceedings of the 20th International Conference on Information and Software Technologies (ICIST 2014), 9-10 October 2014, Druskininkai, Lithuania, Eds. Springer International Publishing, ISBN 978-3-319-11957-1, ISSN 1865-0929, Vol. 465, pp. 324–333, WOS:000360092900026.

Silviu Vert, Radu Vasiu, "Relevant Aspects for the Integration of Linked Data in Mobile Augmented Reality Applications for Tourism", in Proceedings of the 20th International Conference on Information and Software Technologies (ICIST 2014), 9-10 October 2014, Druskininkai, Lithuania, Eds. Springer International Publishing, ISBN 978-3-319-11957-1, ISSN 1865-0929, Vol. 465, pp. 334–345, WOS:000360092900027.

Silviu Vert, "Linked Open Government Data for Smart City Applications", International Conference on Social Media in Academia: Research and Teaching, Timisoara, September 18-21, 2014, Medimond Publ. House, Bologna, Italy, ISBN 978-88-7587-712-5.

Silviu Vert, Diana Andone, "Open Educational Resources in the Context of the Linked Data Web", in Proceedings of the 10th International Scientific Conference eLearning and software for Education (eLSE 2014), Bucharest, Romania, 24-25 April 2014, ISSN 2066-026X, vol. 1, pp. 304–310, WOS:000357153000044.

Radu Vasiu, **Silviu Vert**, Sorin Nanu, "Sustenabilitate si inovare in societate [Sustainability and innovation in society]", U.T.Press, Cluj-Napoca, 2013, 248 pages, ISBN 978-973-662-896-2.

Diana Andone, Radu Vasiu, Iasmina Ermalai, Vlad Mihaescu, **Silviu Vert**, Bogdan Dragulescu, Daniel Ivanc, "Tehnologii Web 2.0 [Web 2.0 Technologies]", U.T.Press, Cluj-Napoca, 2012, 275 pages, ISBN 978-973-662-770-5.

Silviu Vert, Radu Vasiu, "School of the Future: Using Augmented Reality for Contextual Information and Navigation in Academic Buildings", in Proceedings of the IEEE 12th International Conference on Advanced Learning Technologies (ICALT 2012), Rome, Italy, 4-6 July 2012, ISBN 978-0-7695-4702-2 (print), 978-1-4673-1642-2 (CD), pp. 728–729, doi: 10.1109/ICALT.2012.156, INSPEC Accession Number: 12936316.

Silviu Vert, Radu Vasiu, "Mobile Applications for Smart Homes: A Case Study", in Proceedings of the IADIS International Conference Applied Computing 2011, Rio de Janeiro, Brazil, 6-8 November 2011, ISBN print 978-989-8533-06-7, 5 pages.

Vlad Mihaescu, **Silviu Vert**, "Learnability Testing: a Case Study", in Proceedings of the 6th International Conference on Virtual Learning, ICVL 2011, Cluj-Napoca, Romania, 28-29 October 2011, ISSN 1844-8933, pp. 135–140, WOS:000323685900017.

1.3. Structure of the PhD thesis

The thesis is organized in several chapters, described below.

Chapter 1 is an exposition of my motivation towards choosing the subject of this thesis. I put forward the arguments for my interest regarding the aspects researched in the thesis and I list the research questions that are the foundation of my work. The chapter includes a list of all the published papers in which I was an author or a co-author during the period in which I did the research for the thesis. The last subchapter is this short presentation of the structure of the thesis.

Chapter 2 gives an overview of the augmented reality field, starting with some general aspects and a short history of AR and continuing with a comprehensive description of AR software and hardware in use nowadays. The chapter focuses on AR applications for mobile devices (smartphones, tablets, smart glasses) and explores some challenges in implementing such mobile applications. In the last subchapter, I describe the implementation of such an application (Timisoara Street Art) on Android smartphones during an open culture hackathon.

Chapter 3 is a presentation of the Linked Data domain. It highlights the founding principles of Linked Data, it reviews the most significant datasets in the Linked Open Data cloud and it describes the standards of Linked Data i.e. RDF and SPARQL. A presentation of the principles and models for publishing and consuming Linked Data is included. The chapter ends with a focused critical review of the Open Government Data domain in the context of Smart City applications. I describe a case study of an application of such type, which I implemented for Timisoara (Street History).

Chapter 4 is a critical study of the aspects involved in the integration of linked data in mobile augmented reality applications for tourism. I present the current limitations of mobile AR applications regarding content, the advantages of integrating linked open data and some related work in the field. Next, I go into more details regarding the relevant aspects involved, such as geodata integration, trust, provenance and quality issues.

Chapter 5 presents a model for integrating linked open data in mobile augmented reality applications. I show in detail how such a model can be implemented for the touristic domain, in the form of a mobile web application that shows Points of Interest (POIs) around the tourist using the augmented reality medium, the POIs being fetched from various Linked Open Data platforms and Open Government Data portals and integrated into a single, consistent dataset. I point out challenges in this endeavor and solutions on how to overcome them.

Chapter 6 is an analysis of the appropriateness of linked open datasets for augmented reality scenarios. Based on literature review in the field of data profiling, I propose a methodology for profiling and assessing this appropriateness and I apply it to the datasets that I integrated in the developed prototype. Based on the resulting statistics, I draw some conclusions regarding the usefulness of such integration. The chapter also includes a comparison between the developed prototype and other similar projects, regarding the exploited data and other architectural approaches.

Chapter 7 is the concluding part of the thesis. I start with some general conclusions regarding the research that I have done, I continue listing the exact theoretical and practical contributions that this thesis brings in the scientific fields that it addresses and I finish by issuing some future research directions that might be useful to other interested researchers and even for my own postdoctoral research.

2. AUGMENTED REALITY

2.1. Introduction

Augmented reality is a representation of our surrounding space whose elements are augmented by text, image, sound and video generated by a computer. The technology is used to ameliorate the perception of the user on the reality.

Augmented reality is part of the concept of mixed reality, being placed between the real space and the virtual space, closer to the former one [6].

Figure 2-1 Milgram's Reality-Virtuality Continuum (image by Russell Freeman, licensed in the Public Domain [7])

The augmentation usually works in real-time and in registration with the elements of the real world, transforming it into an interactive and digitally manipulable world, by superimposing the virtual information over the real objects.

Although most of the applications enhance the visualization of the surrounding world by means of adding a visual layer of information, augmented reality is also about improving the other human senses, like hearing, smell, taste and touch. Augmented reality can even replace a sense, for example the sight, in the case of blind people.

There are also cases where augmented reality is used to alter the real world through subtraction. In this case, it is called *mediated reality* or *diminished reality* and it works by digitally removing objects from the representation of the real world. This is usually done by replacing them with the background of the surrounding area, thus leaving the impression that they do not exist. One typical example of using this technique is hiding aggressive billboards that are placed on the sides of the streets and which can distract the attention of the drivers (and replacing them with useful messages) [8].

For performing the visual augmentation of the surrounding reality, one can use multiple devices. The most popular at this moment are the smartphones, due to the fact that they are relatively affordable and omnipresent. However, they also come along with some disadvantages, such as low processing power and small displays. An alternative is represented by tablets, which solve the problem of small displays, but are cumbersome to manipulate in outdoor activities.

More approapriate devices for augmented reality, but less affordable and not widely used, are the so-called *head mounted displays* (HMD). These equipments are worn on the head, like a helmet, typically covering the eyes. Lately, researchers were able to reduce the size of these devices so they now resemble usual eyeglasses. Research is already on-going for miniaturized devices that fit even in contact lenses [9].

Augmented reality is used nowadays in many fields, such as:

- marketing (e.g. seeing the car in 3D on top of a printed ad for that car),
- production (e.g. simulating prototypes of what should be built next, placed in the context where they will be effectively used),
- games (e.g. transforming the surrounding area in a battle field with virtual monsters),
- education (e.g. learning the history of a place by exploring it with an augmented reality device),
- medicine (e.g. providing additional information to a doctor during a surgical procedure) or
- tourism (e.g. finding easily a restaurant which satisfies the tourists demands).

2.2. History

The father of augmented reality seems to have been Morton Heilig, which believed, in the 50s, that the cinema should be an activity that encompasses all our senses. This way, the person that is watching the movie would be completely absorbed by the story. He builds Sensorama in 1962 [10], a mechanical device described, at that time, as being the future of the cinema. Sensorama could reproduce stereoscopic 3D images (the technique of building 3D images from 2D images), was able to give the human body the perception of movement, to create stereo sounds, to create the perception of the existence of wind and to create smell aromas based on the movie scene. It was a device that today is called a multimodal device (multiple possibilities of in and out ways). Unfortunately, the project did not have commercial success during those times [11], so it remained until today just a technical curiosity.

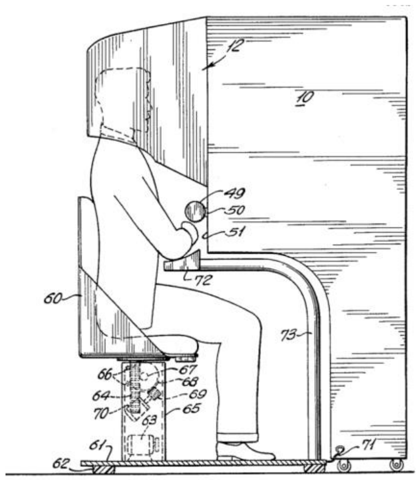


Figure 2-2 Sensorama (Figure 5 of U.S. Patent #3050870, Source: Wikimedia Commons, License: public domain) [12]

Starting with 1966, the scientist Ivan Sutherland does some first experiments on augmented reality, together with his colleagues. In 1968, he invents the first optic HMD (*head mounted display*) [13] which was nicknamed "The Sword of Damocles" [14]. In 1975, Myron Krueger creates Videoplace, a room that, for the first time, allowed users to interact with virtual objects [15]. Later, in 1990, Tom Caudell from Boeing coins for the first time the term "augmented reality", while helping his colleagues to repair some cables in the aircraft [16]. In 1992, L.B. Rosenberg develops, in an American military laboratory, the first augmented reality functional system, proving also that the system produces improvements in human performance [17].

In 1997, Ronald Azuma writes the first important paper on augmented reality, in which he gives a detailed definition of the field, suggesting application domains, limits and perspectives [18]. In 1999, Hirokazu Kato develops ARToolKit, one of the most used augmented reality software library [19], and in 2000, Bruce Thomas develops ARQuake, the first mobile augmented reality game for outdoor [20]. In 2008, Wikitude Travel Guide is launched, an application which uses augmented reality to help tourists to get a better sense of their surroundings. Since then, large varieties of applications have been launched, especially for mobile devices, which cover a range of domains, from touristic to medical. Starting from 2010, research worldwide focuses more and more on SLAM (simultaneous localization and mapping) techniques for augmented reality [21].

2.3. Extended definition

In 1997, Azuma defines [18] augmented reality as having these three main characteristics:

1. Superimposes virtual information over the real world
2. Is interactive and is in real-time
3. Is registered in 3D

In the case of the first characteristic, it is important to note the wording. The user stays in the real world and has direct interaction with it. Virtual information (visual, auditory) is only superimposed on the real wold, to help her achieve a task, perceive something hidden etc. Nothing makes the user think she is entirely in a virtual world, as it happens with virtual reality.

In case of the second characteristic, the user can interact in real-time with the virtual information (and the real world, of course), not merely watch it. She can touch, smell, taste etc. She can move it, change the color, change some features, improve it or simplify it.

In the case of the third characteristic, registration is crucial for a successful augmented reality experiece. If the virtual information "flows" on top of the real world, with no apparent connection to what is around, then the resulting image is not natural and convincing. Unfortunately, this happens often in common applications, as registration is hard to achieve. Registered in 3D means that inserting the virtual object into the real world must take into account shadows, ocluding objects etc.

Along with this spatial registration, temporal registration must also be produced, that is, when the user is changing her point of view (e.g. she is moving), there should be no lag in the virtual object keeping its position in the real world.

Augmented reality has some characteristics in common with other - some better known - media, technologies or ideas, such as virtual reality, telepresence, global positioning system (GPS), geographic information system (GIS), fabrication, cyberspace, mixed reality, virtuality or the metaverse [22]. However, these are different from augmented reality in various ways, even though people sometimes confuse one with the other, as they are inclined to simplify things and to work with concepts that they already know.

2.4. Augmented reality hardware

In the next subchapter, I describe the hardware components of the augmented reality technologies using the classification done by Alan Craig in 2013 [22, pp. 69–110].

2.4.1. Sensors

Their role is mainly to provide information on the position and the pose of the user [23] i.e. to track her. Types of tracking:

Optical tracking

A camera is usually used for doing optical tracking. The images taken by the camera are then processed by a computer to determine what the camera saw. This allows multiple objects to be tracked simultaneously.

The most common cameras are the ones integrated in smartphones and tablets. Due to high proliferation of quality cameras in such electronics, optical tracking tends to become inexpensive.

Special cameras are used for high complexity augmented reality experiences. In certain conditions, where optical tracking is not possible in visible light, one can use infrared or ultraviolet cameras.

The first experiments with optical tracking were based on artificial markers that were scattered in the environment and which were relatively easy to detect with computer vision algorithms. At the beginning of the 2000s, research started focusing on feature-based and model-based techniques for detecting objects, which did not require altering the environment with artificial markers [24].

Other tracking types, less used, but important depending on the use case, are:

Acoustical tracking means that microphones are used to determine the location of the object that is tracked, based on the properties of the sound(s) generated by the object. According to [25, pp. 54–56], acoustical tracking systems can be runtime trackers or phase shift trackers. The first type works by attaching multiple subsonic sources to the object being tracked and measuring the propagation time from each of the sources to a stationary receiver. Based on these values, the system can calculate the distances to the tracked object. The number of required emitters grows if one needs to know the orientation of the object, not only its position. The disadvantage of this technique is that the propagation of the sound in the air is affected by humidity, temperature or wind. The second type of acoustical trackers measures the distance to the tracked object by calculating the phase shift between two signals having the same frequency. One is generated by a fixed sound source and the other comes from the tracked object.

Acoustical tracking does not depend on lighting conditions, however, it has a limited range, cannot be used in noisy environments and each object being tracked must have its own sound source [22, pp. 75–76].

Electromagnetic tracking works by using a fixed transmitter with 3 orthogonal antennas which emit signals, sequentially through the antennas, to a similarly built receiver, attached to the object being tracked. The signal received back is processed, based on its level, to determine the 3D position and orientation of the receiver [26]. This type of tracking is very precise and accurate, being able to track in six degress of freedom. Its disadvantages include sensibility to metal in the environment, limitedness in range (usually a few meters) and cost of production (due to rare use) [22, pp. 76–77].

Mechanical tracking works by attaching linkages to the object one wishes to track and measuring with gears or potentiometers the rotations and movements of the object. The combination of all measured values will give the precise location and pose of the tracked object [27, pp. 16–17]. It is very fast, precise and accurate, but works only on very small areas, is visually obtrusive and is quite costly [22, pp. 77–78], being suited for very specific applications (e.g. augmented reality in surgery).

Hybrid tracking is a method that combines several sensors that individually would not be able to fully track an object [24]. They include:

- Accelerometers – to determine relative motion (they suffer from error propagation, if not corrected)
- Depth sensors (that work on technologies such as optical, acoustical, radar etc) – they measure the distance to an object
- GPS receivers – they detect absolute location, in terms of latitude and longitude, but with errors in the range of tens to hundreds of meters
- Compass – provides orientation based on the Earth’s magnetic field
- Gyroscope – provides values of rotation around a particular axis

Most modern smartphones and tablets already incorporate these sensors. Coupled with the bundling of a camera (optical tracking) and microphone (acoustical tracking), they become a ubiquitous and relatively inexpensible all-in-one powerful augmented reality system.

2.4.2. Processors

Processors have the task of getting the information from the sensors, processing it according to the case scenario and sending appropriate data to the display. They may be composed of CPU(s), for processing data, and/or GPU(s), for processing graphics.

According to [22, pp. 81–85], from an architectural point of view, the processor can act on a single machine or it can be distributed in a client-server model.

In the first case, the application can run on a truly mobile device (handheld), such as a smartphone or tablet, or on a more-or-less mobile device, such as a laptop or desktop computer. Handhelds are appropriate for scenarios where mobility is crucial. Today they are able to process a considerable quantity of information. However, they are yet unable to provide a realistic augmented reality experience which implies complex graphics. Laptops or desktops have a better performance in this respect, but mobility is quite hindered or inexistent. They are more appropriate in fixed-point scenarios (e.g. public kiosks, museums).

In the case of a client-server model, the handhelds or the laptop/desktop are connected via network to a server, which does all or most of the processing. This is very useful in case of processing big data or very complex graphics. However, the network implies latency, which can hinder greatly the augmented reality experience, so a decision must be made taking into account all the requirements of the use case. The network, in the most popular applications, is the Internet, and the client uses a browser to deliver the experience. Access to such applications is almost guaranteed, however, latency and technical browser limits can hinder the experience [28].

Several characterists have to be taken into account when choosing a processing system [22, pp. 88–90]:

- Number of processors and processor speed: both increase with the complexity of the processed images or data [15].
- Available memory and storage: memory in the processing unit is useful to store the data in it for lower latency; disk storage is dependent on how big the virtual objects are.
- Graphics accelerators: GPUs are very useful for high image processing tasks.
- Network bandwidth and latency: bandwidth is important for transferring large quantities of virtual objects, if the scenario requires it. Latency is another aspect, unrelated to bandwidth: a low latency network is crucial to a natural augmented reality experience, but it is generally hard to have such a network.

2.4.3. Displays

Displays are used to provide various sensors to the user, such as signals to the eyes, to the ears, to the mouth, to the skin or to the nose. They can be categorized as such [22, pp. 91–110]:

a) Visual displays

They are the most familiar type of displays. The computer monitor or the TV are well-known displays of this type.

They fall into several categories:

Stationary visual displays typically have a fixed position and can be of two types: kiosk or projection-based. Kiosks can vary from our personal computer monitor at home to computers monitors that are used in public places, such as shops or public institutions, used for sales or information purposes, and which use augmented reality to deliver the user experience.

Projection-based displays are different in that light does not come from the display itself, but is rather projected on a surface. It can be a projection screen, which has the advantage that the illumination and position are known and can be easily manipulated. Otherwise, it can be the real world itself, upon which virtual information is projected. This type of augmented reality is well suited for large groups of users and collaborative tasks [29, p. 11]. However, in very bright environment, it is extremely challenging to be able to project in such a way that the virtual objects are visible. Moreover, the working space is limited. When projection must be done on real objects that are moving in the environment, things become even trickier. According to [30, pp. 87–90], other disadvantages of projection-based displays include the fact that conventional projectors only focus on a single plane placed at a constant distance or, if multiple projectors are required, the increased difficulty to align the images and to calibrate the colors. On the other side, advantages include better ergonomics, unrestricted field of view and easier eye accommodation.

Head mounted displays (HMDs) are visual displays that move with the user's head. They can be:

Head worn, such as helmets or glasses (see Figure 2-3). They are optical see-through devices or video see-through devices. The optical ones allow the user to see directly the real world, on which virtual images are superimposed afterwards. The electronic ones provide the combination between the image captured by the video camera and the virtual objects generated by the computer. According to [29, p. 10], the advantage of the first type is that the resulting image has the natural resolution of the real world, its disadvantage being that the virtual objects are not registered perfectly with the real world and the resulting image looks unauthentic. The advantage of the second type is that the real image is combined with the virtual objects before being provided to the human eye, so they are better aligned, the disadvantage being that the overall resolution is limited by the resolution of the video camera.

Figure 2-3 Head worn device for augmented reality (image from Sensics Inc. via Wikipedia licensed as CC BY-SA 3.0 [31])

They may display to only one eye or to both eyes, in which case they display the same image to both eyes or they display distinct images, to achieve stereoscopic 3D viewing.

Eye worn. Research is already done to integrate augmented reality in contact lenses [9], which is very appropriate for those that do not want to use bulky equipment or those that need to hide it (spies, military, actors etc.).

Stationary. The overall system is fixed, but the component that holds the display can be attached to the head and is moveable (similar to stationary binoculars for viewing scenic landmarks) [22, p. 100].

Handhelds are visual displays that move with the user's hand (Figure 2-4). These are the most common ones, due to the proliferation of personal mobile devices such as smartphones and tablets. The display can be either the screen of the device or, more recently, any surface that the mobile device can project on. Mini projectors have become easier to afford lately. The experience, however, is sensitive to ambient conditions, if we take into account the small amount of light that is projected in this case.

Because mobile devices are so important (the latter two categories), I list here some advantages and disadvantages of them [22, pp. 212–218]:

Advantages:

- Augmented reality can happen anywhere the user can bring her mobile device at (and that means almost everywhere!)
- Can support a wide range of applications (education, entertainment, marketing, games, tourism etc.)
- Are low cost compared with stationary, specific use case systems, considering that they have most of the typical augmented reality components already integrated
- People already own them massively (in case of smartphones - and a lot of marketing is done for glasses also nowadays)

Figure 2-4 Augmented reality map on a smartphone (image from Wikimedia Commons licensed as CC BY-SA 3.0 [32])

Disadvantages:

- Limited in terms of memory, processing capacity, display area etc. – this does not mean that successful applications cannot be built using them, it is just that wise preparation and good optimizations should be made to assure good functioning
- For devices using vision augmented reality, ambient light is usually a problem
- For client-server architectures, there are situations when no network is available or the latency of the network is very high
- In harsh environment conditions, these devices are not able to operate
- It may not be clear for users where they can experience augmented reality

Mobile augmented reality has seen significant growth in the past few years, due to the proliferation of smart mobile devices. Mobile augmented reality applications have been proved to work successfully in domains ranging from education to gaming and tourism [33]. The most popular mobile augmented reality applications are those that allow the user to explore her surroundings, most of these applications being what we call augmented reality browsers.

b) Other types of displays

Audio displays. These produce "virtual sounds [which] are perceived as an extension to the natural ones" [34]. Similar to visual displays, they can be [22, pp. 103–107]:

- Stationary audio displays: the most common are the home stereo speakers; they work best for multiple users in the same place, but they suffer from echo problems, noise pollution, interference with other people's activities etc.

- Audio displays that move with the user's head: these can be headphones or earbuds. They provide the much-needed privacy to the user.
- Audio displays that move with the user's hand: smartphones, tablets etc.

Haptics. These displays target our kinesthetic sense (force, motion) and/or our tactile sense (texture, temperature) [35]. A well-known example of this kind of device is the robotic arm that gives force feedback.

Smell. Administered as a mist from an atomizer controlled by a computer or via tubes to a single person. It is hard to disperse it rapidly in large areas and even harder to eliminate it rapidly [22, p. 110].

2.5. Augmented reality software

In the folowing subchapter, I describe the software components of the augmented reality technologies using again the classification done by Alan Craig in 2013 [22, pp. 125–149].

Software used by the AR application

Environmental acquisition. This software is very specific to various sensors, such as the camera, the GPS or the accelerometer, and its role is to access the data that they capture.

Sensor integration software. The role of this software is to process the sensor signals and to transform them into useful data. The most common example is processing the image from multiple cameras, with the final goal of better position and orientation tracking.

Application engine. It is the core of the augmented reality application and mediates between the sensor integration component and the rendering component of the displays. It is responsible for the flow of the application, the "tricks" of the augmented reality experience and the content management (loading and unloading the virtual objects). The developer usually writes from scratch the application engine, as the tool is very specific to the use case. However, the programming language is heavily influenced by the language the augmented reality library is written in and by the platform that the application will be working on (Windows, Android, iOS etc.).

Rendering software. It is software used to display data on the chosen display, which can be visual, audio, haptic etc. One of the well-known libraries for visual augmented reality is OpenGL[3], which is an open graphics platform present in many hardware accelerators optimized for it.

Usually, this software (all of the components presented above) is available as augmented reality libraries. Some of the established players in this field include Metaio, Wikitude, Vuforia, Layar and D'Fusion. Metaio, most probably the leading company in this area, was bought by Apple in May 2015 [2], which shows a great interest towards augmented reality from the big technology companies.

A detailed comparison of existing augmented reality libraries is shown in [36].

[3] https://www.opengl.org/

Software for creating the AR application

This software basically refers to integrated toolsets for developing augmented reality applications. For example, Eclipse, Netbeans or Android Studio for developing applications in Java for Android, Apple's Xcode for developing in Objective-C or Swift for iOS etc.

Software for creating the AR content

Ussually, there is a specialized software for each type of content to be produced. If we are considering only visuals and sounds, there are three main categories [22, pp. 136–147]:

Software for creating 3D graphics. 3D graphics can either be created from scratch, or they can be "imported" by scanning a real-object or by analyzing photos that portray the object from multiple angles. In the first case (from scratch), the graphics are built and stored somewhere, from where they will be loaded in the application later, or they can be dynamically built during application runtime. It depends on the graphics complexity and on the speed and space available.

For scanning purposes, nowadays companies offer software that runs on the smartphone and can be used by anyone to scan everyday objects that one wants to use in the augmented reality application. A 3D model can also be obtained from another source and imported.

The most common 3D formats are .obj, .max, .skp, .blend, .wrl (or the newer .x3d) and .dae.

Software for creating and editing 2D graphics. This software is of two types: vector graphics software (for images created from scratch) and raster graphics software (for editing pixel-based images). The most common 2D graphics formats are .gvg and .cgm (for vector graphics) and .tiff, .jpg, .gif, .png and .bmp (for raster graphics).

Software for creating sound. Sounds can be created "on the fly" in the augmented reality application, can be recorded in the real world, synthetized or just edited and then served to the augmented reality application. Common audio file formats include .wav, .aiff and .mp3.

2.6. Challenges in developing augmented reality browsers for tourism

The focus of this thesis is represented by augmented reality applications that work as a browser, which turists use to discover Points of Interest around them.

A typical problem in augmented reality browsers is correctly projecting, on the screen of the mobile device, the real world position of and distance to the POIs.

A mathematical model for achieving this is proposed by Zander et al. [37]. In this paper, the researchers present a computational model for the replication and integration of Linked Data in mobile augmented reality applications, demonstrating the approach for a tourist who is wandering in a mountain area.

The mathematical model described in [37] is briefly illustrated below. It uses the Vincenty's Inverse Formula [38], which is more accurate than the better known haversine formula [39]: the former formula considers the shape of the earth to be ellipsoidal, while the later one considers the shape to be spherical.

The Vincenty's Inverse Formula is used, in the beginning, to compute the following functions:

$$v_{\text{bearing}}: P \times C \rightarrow [0, 2\pi) \qquad (2.1)$$

$$v_{\text{distance}}: P \times C \rightarrow [0, r) \qquad (2.2)$$

$$f_{\text{incl}}(p, c) = \arctan\left(\frac{\psi_{poi} - \psi_{cam}}{v_{distance(p,c)}}\right) \text{ with } p \in P, c \in C \qquad (2.3)$$

where:

- $v_{bearing}$ calculates the bearing angle β from a camera point $c \in C$ to a geographic location $p \in P$ using the Vincenty's Inverse Formula
- $v_{distance}$ calculates the distance from c to p using the Vincenty's Inverse Formula
- f_{incl} calculates the inclination angle θ from a camera point c to a POI p
- P is a set of POI geographic locations $p = (\lambda_{\text{poi}}, \varphi_{\text{poi}}, \psi_{\text{poi}})$ (longitude, latitude and altitude of the POI)
- C is a set of camera parameters $c = (a, \rho, H_{aov}, V_{aov}, \lambda_{poi}, \varphi_{poi}, \psi_{poi})$ which correspond to: azimuth of device, pitch of device, horizontal angle of view, vertical angle of view, longitude / latitude / altitude of camera

Based on the azimuth and pitch of the camera, retrieved from its orientation sensors, as well as the horizontal and vertical angles of view from the camera's field of view, two horizontal and vertical angle-of-view border functions are defined as:

$$l_{border}(c) = (a - H_{aov}/2) \bmod 2\pi \qquad (2.4)$$

$$r_{border}(c) = (a + H_{aov}/2) \bmod 2\pi \qquad (2.5)$$

$$t_{border}(c) = (\rho + V_{aov}/2) \qquad (2.6)$$

$$b_{border}(c) = (\rho - V_{aov}/2) \qquad (2.7)$$

which define the bearing angles of the left and right border of the camera's horizontal field of view, respectively, the angles of the top and bottom border of the camera's vertical field of view.

Next, the authors [37] define a function f_{aov}: $P \times C \rightarrow S$ which takes POI locations and camera position and yields angle-of-view coordinates for the POIs:

$$f_{aov}(p,c) = \begin{pmatrix} t_x(v_{bearing}(p,c),c) \\ t_y(f_{incl}(p,c),c) \\ v_{distance}(p,c) \end{pmatrix} \tag{2.8}$$

where

- S is a vector field of POI positions within the camera's field of view $s = (x_{deg}, y_{deg}, z)$ where $x_{deg} \in [0, H_{aov}]$ is the horizontal angle inside the camera's field of view, $y_{deg} \in [0, V_{aov}]$ is the vertical angle inside the camera's field of view and $z \in [0, r]$ is the distance from the camera to the POI
- t_x is a horizontal transformation function t_x: $[0,2\pi) \times C \rightarrow [0,1]$ defined as:

$$t_x(\beta,c) = \begin{cases} \beta - l_{border}(c) & \text{if } \beta \geq l_{border}(c), \\ \beta - (2\pi - l_{border}(c)) & \text{otherwise} \end{cases} \tag{2.9}$$

- t_y is a vertical transformation function t_y: $[-\pi, \pi) \times C \rightarrow [0,1]$ defined as:

$$t_y(\theta,c) = t_{border}(c) - \theta \tag{2.10}$$

t_x and t_y result in a coordinate $x_{deg} \in [0, H_{aov}]$ and $y_{deg} \in [0, V_{aov}]$ which define the position of the POI within the field of view of the camera.

The final coordinates for each POI on the camera of the mobile device are obtained as such:

$$x_{screen} = \frac{x_{deg}}{H_{aov}} w \tag{2.11}$$

$$y_{screen} = \frac{y_{deg}}{V_{aov}} h \tag{2.12}$$

The value of z is used to show the distance to the POI. An illustration of the calculated values is shown in [37, Fig. 2].

2.7. Timisoara City Art – a prototype for a mobile augmented reality application

The following subchapter describes the design, features and implementation of an augmented reality application for tourists, which was coordinated by the author of this thesis.

Timisoara City Art is a project developed during the Timisoara Open Culture Hackathon organized in Timisoara in 18-19 April 2015. The hackathon was the first event in Romania to facilitate the reuse of open cultural data. Developers, graphic designers, artists and representatives of national cultural institutions were invited to develop applications and design platforms that can reinvent the relationship between the public and the cultural works. Cultural data refers to collections, art works, books and other types of publications, audio visual materials, photographs, archived documents, monuments (content), as well as to descriptive information about these (metadata), such as title, creator, year, dimension, technology used etc. The event was organized by the Romanian Coalition for Open Data, the Department for Online Services and Design - Romanian Government, the Multimedia Center of the Politehnica University Timisoara, West University Timisoara, Smart City Association, Kosson Community, Open Knowledge Romania and the Timisoara City Hall.

Timisoara City Art was proposed as a project for the hackathon by the Multimedia Center of the Politehnica University Timisoara and was developed by a team of five students: Silviu Vert (coordinator and augmented reality expert), Victor Holotescu and Aurel Chiper (Android developers), Sorin Voina (graphic designer) and Daniela Imbrescu (copy editing). The open cultural data was provided by the Triade Foundation, a well-kwown organization acting in the cultural landscape of Timisoara, and consisted of 15 monuments or groups of monuments that belong to Timisoara public street art. Each monument featured information such as title, creator, biography of creator, year of creation, material that it is made of, description and photo. The projects developed at the hackathon, including this one, contribute to the efforts of the city in the race towards the title of European Capital of Culture in 2021.

The application is built for the Android system and works on versions of the operating system equal or greater than 5.0 Lollipop. It boosts a fresh and intuitive design that respects the guidelines of the Android operating system. The information is saved in a local database on the smartphone, so the application is available for tourists without a mobile data plan. The mobile application has five big sections, which are reflected in the design of the menu: home, map, augmented reality, list and favorites.

The home screen features a big version of the logo of the application and a short description of it. The logo was chosen in such a way that it resembles modern abstract art, so it cannot be considered biased towards one of the many cultural currents that exist in Timisoara. On the home screen we also placed photos of some of the monuments in the application, so the user can quickly jump to the details page of one of them and see what the application is about. The monuments that appear on the home page can be scrolled through by sliding the finger on the screen from left to right and from right to left (a carrousel type of component). Initially we wanted to place on the home page only 3 monuments, the most representative ones, but it was suggested to us by experts in culture that were attending the hackathon that this is a very subjective way of choosing and it leaves space to critiques. As such, we picked 7 monuments in a random way to be shown on the home page, which is shown in Figure 2-5.

Figure 2-5 Screenshot of the home page of the Timisoara City Art mobile application

The map section of the application is designed in such a way that the user can see all the monuments in Timisoara overlayed on the map, so she can get a bigger picture of what she is able to visit in the city. The map is based on the API provided by Google Maps. On clicking a pinpoint on the map, a popup opens with the name of the monument, and the user can click on the popup to go to the details page. The user can also click the "go to" icon at the bottom of the map, which shows her driving/walking directions towards the point of interest which she desires to visit. This section is shown in Figure 2-6.

The augmented reality section helps the tourist to easily and intuitively discover monuments around her, by looking at the surroundings through her smartphone. She is able to see big pinpoints overlayed in space on the monuments around her and the name and picture of the monument, from the database, so she can recognize the real one. On clicking the photo, the application redirects her to the details page. The augmented reality view is implemented using the Metaio SDK platform. We chose the Metaio platform as it provides a complete solution for free (with watermark) and as such it is appropriate for current and future developments of the application. One drawback of it is its poor documentation (lack of a complete example for a location-based implementation), at least for the free version, which we were able to use in the hackathon. For the moment, the implemented prototype uses only location-based augmented reality techniques. The augmented reality view is shown in Figure 2-7.

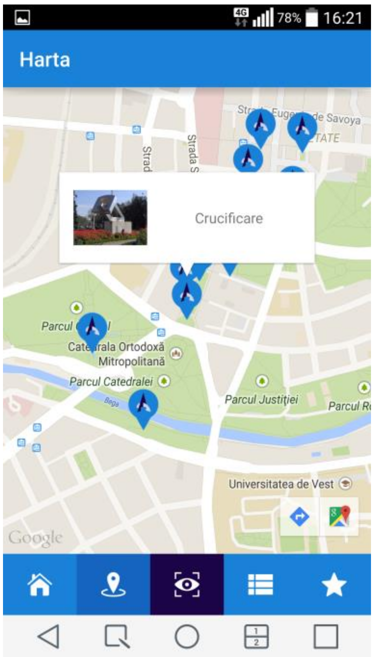


Figure 2-6 Screenshot of the map section of the Timisoara City Art mobile application

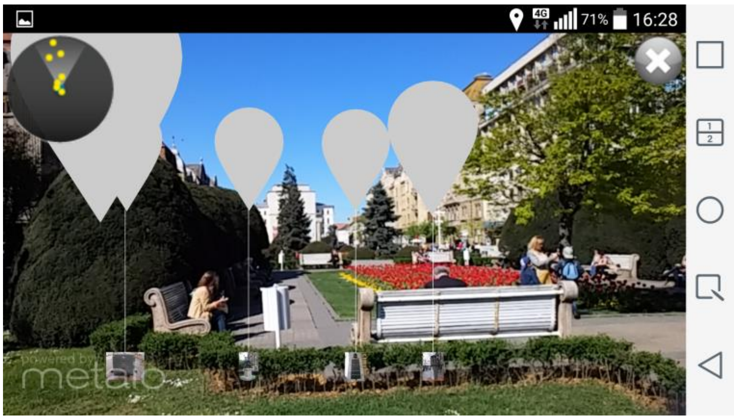


Figure 2-7 Screenshot of the augmented reality section of the Timisoara City Art mobile application

The list section contains a list of the monuments in the application, with name, picture and a “mark as favourite” icon. Clicking on one of the monuments redirects to the details page. This section is shown in Figure 2-8.

Figure 2-8 Screenshot of the list section of the Timisoara City Art mobile application

The favourites view is similar to the list view: it only features the monuments that were marked as favourite.

The details page consists of the title of the monument, its photo and some details that were already mentioned at the beginning of this subchapter. This section is shown in Figure 2-9.

The application is in its alpha stage, having been developed and designed in 15 hours along the 2 days of the hackathon, but is fully functional with the data that it had available. Future work, regarding the data, consists of integrating the whole dataset of monuments that exist in Timisoara, which will be provided gradually as open data by the Triade Foundation. Such a dataset is expected to contain around 100 monuments and detailed information about them. On the technical side, we plan to integrate more advanced augmented reality views in the application, which rely on 3D models for the monuments that have this information.

The source code of the application, according to the hackathon rules, will be shortly available online with an open license.

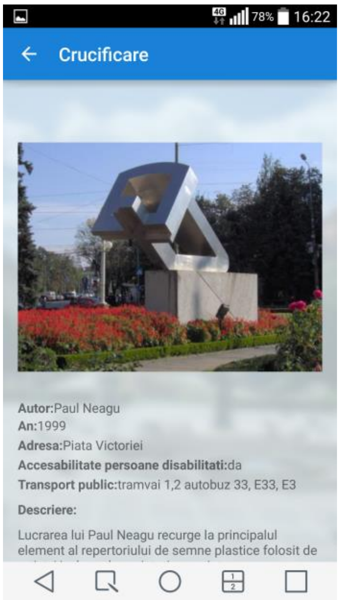


Figure 2-9 Screenshot of the details section of the Timisoara City Art mobile application

2.8. Conclusion

This chapter is dedicated to an overview of the augmented reality domain and sets the foundation of the present thesis. In it, I review the major components of augmented reality, both hardware and software, to get a clearer image of the current status of these technologies.

I start by giving both a narrower and a broader definition of augmented reality and I show that it is placed on a reality-virtuality continuum, closer to full reality and farther from full virtuality. I present a few milestones in its history, so the reader can better understand how these technologies evolved: from bulky equipments and unrealistic experiences to pervasive, ubiquitous technology which fits in wearable devices.

Next, I show that augmented reality hardware is composed, roughly, of sensors, processors and displays, with the trend being to incorporate all the three components in one device. Although augmented reality can address all five human senses, the most popular applications are the visual ones.

Great augmented reality experiences require considerable processing power and memory. This is possible to achieve with professional equipments on the market. However, these are bulky and expensive. The trend in this field is for augmented reality equipments to become smaller and smaller, to the point where they become indistinguishable from everyday objects or clothing. The most popular equipments nowadays are the smartphones, which are omnipresent and which have the advantage that they incorporate most of the sensors needed for a good augmented reality experience. Given that smartphones tend to become a universal device, I choose to focus on such type of mobile equipments in the present thesis.

The research literature points out some challenges encountered in creating mobile augmented reality applications. One of them, in the case of location-based applications, is how to project Points of Interest from the surroundings on the display of the mobile phone. In this respect, I describe an approach from the literature, which uses a mathematical model based on Vincenty's Inverse Formula to calculate the position of a Point of Interest on the smartphone display, and the distance to it, in a timely manner.

In the end of the chapter, I describe my experience in developing, with a team of students, a mobile augmented reality application for tourists, called Timisoara Street Art. I explain the context, the choice of features and the usage of augmented reality software. This way, I aim to prove the viability of developing augmented reality applications on mobile devices.

In conclusion, based on the findings presented in this chapter, which constitute a theoretical contribution to the present thesis, I consider that augmented reality is a mature field, with proven usefulness in various domains, and that mobile devices are the appropriate medium to bring augmented reality to everyone, everywhere. The practical contribution of this chapter consists in developing the Timisoara Street Art application. Some results presented in this chapter have been published in [40].

The next chapter will focus on content exposed as linked data, as a way to further enhance such mobile augmented reality applications.

3. LINKED DATA

3.1. Introduction

We live in an age where information is all around us – and it often feels like it overwhelms us. The timetable of the trains, the active substances in a headache drug, the list of touristic attractions in a popular city, the spending figures in our City Hall, the winners of the most recent Olympic Games, our business appointments next week – just to name a few, are data that we can find at a few clicks distance, in our computer or on the Internet. Due to all the novelties brought by the information engineers, we are able to find, sort, view and understand them. However, they are too many to cope with them all, and we wish there would be an intelligent robot, sort of a personal assistant, that would help us filter the most important information, make useful connections between independent data and suggest action to us – or better, do it itself.

This was the dream of Tim Berners-Lee, the inventor of the World Wide Web, who in the '90s proposed the idea of a Semantic Web [41] that, in the end – or, better, at the beginning – would make the popular dentist appointment for us. The problem with the initial WWW, he explained, was that all the data was published in documents – HTML – which would be linked between them, so one could jump from one website to the other, in search for more information. The well-designed web pages were perfect for people to understand them, and information was little and comprehendable by all. However, machines were not able to understand the information and get the data out: pages were full of unstructured data (which to people looked structured only because of the templating – not understood by machines) and linking between them, for a machine it only meant: "look, there's more information there".

Tim Berners-Lee proposed a stack for the Semantic Web, which implied the use of ontologies, trust, cryptographic keys etc. (see Figure 3-1).

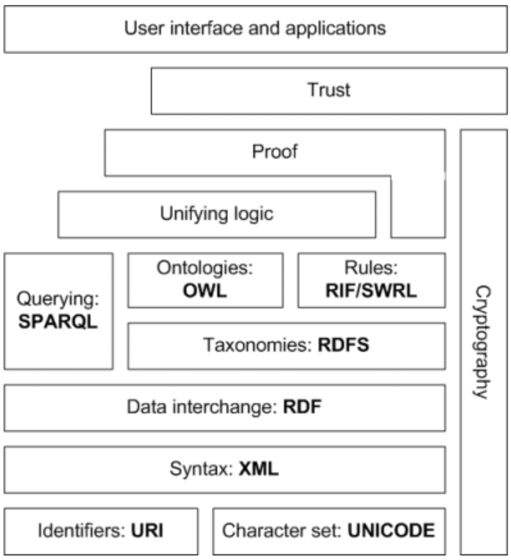


Figure 3-1 Semantic web stack (image from Wikimedia Commons, licensed as CC0) [42]

Far from saying that the Semantic Web vision was a failure, it is useful and sincere to say that the whole thing did not take the envisioned direction.

On a shallow level, but with great adoption in the real Web, the concept of structured data materialized in microformats, which were easier for developers to understand and to embed in their web pages [43].

More closer to the original proposal, standards and principles started to emerge from the universities and the research groups and were adopted, slowly but surely, by the big public and the big players.

The data model, RDF, and the quering language, SPARQL, stand at the base of the growth. RDF, with its simple model subject-predicate-object, offers an easy way to structure even the most complicated dataset, of which there are plenty in the wild. The two objects and the link between them have meaning – they can be described by simple vocabularies or more complex ontologies (for which OWL was created). These links can be made not only inside the dataset, but also between data from different datasets – and this is a major breakthrough, as it allows meaningfully linking isolated silos of information, and by traversing the data, discovering new information and inferring new knowledge. SPARQL helps to query all this heterogeneous data at once – something considered impossible in the world of relational databases – and thus bridge the gap between datasets and create easy access to the desired information.

Even though the full range of semantics has not been adopted yet into the mainstream [44], due to its complexity, a simpler set of principles has been implemented with visible and useful results, based on the standards already mentioned.

Creating names for data, using the URI model, publishing it in various formats that give atomic access to data, linking the datasets and using RDF for modelling gave rise to the concept of Linked Data, which we can call the more practical and successful part of the Semantic Web.

Linked Data [45] is, on one side, a set of principles for publishing structured data on the Web, and, on the other side, the way in which the data itself is called.

The biggest visible result of it is considered to be the Linking Open Data project, which is a network of interconnected datasets from all domains of knowledge, published as linked open data, and which had an astonishing growth: from 12 datasets in 2007 to 570 datasets in 2014 [46].

Alongside the universities, governments, user-generated community platforms and other individual players that published linked data in this form, there is clear adoption of linked data by big players such as IBM, Oracle, Google or Facebook [47], so Linked Data has set itself on a successful journey to a bright future.

3.2. Principles of Linked Data

In a W3C note [48] written in 2006, Tim-Berners-Lee proposed a set of guidelines for publishing and working with Linked Data, that have become the founding norm of it. They are:

1. Use URIs as names for things
2. Use HTTP URIs so people can look up those names
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
4. Include links to other URIs

The guidelines work with concepts that have already made WWW successful, such as URIs, HTTP and hyperlinks. As such, Linked Data is not something designed from scratch, but an addition to the WWW that enables the publishing of structured data, making it easy to grasp and implement.

The principles are described in detail in the following subchapters.

3.2.1. Principle 1: Use URIs as names for things

There are many things of interest in the world, starting from concrete ones, like persons or places, to more abstract ones like love or peace. To be able to point to them, which is mandatory in the envisioned world of structured data, one should have a system of identification and use it accordingly. The best-known system on the Web is the URIs (Uniform Resource Identifiers) system. By using URIs, a city would have the identifier

```
http://dbpedia.org/page/Timisoara
```

and the concept of love would have the identifier

```
http://dbpedia.org/page/Love
```

DBpedia is the Linked Data version of Wikipedia. More details about DBpedia can be found in subchapter 3.3.1.

3.2.2. Principle 2: Use HTTP URIs so people can look up those names

URIs are not only URLs, as in the example above, but also other types of identifiers, for example DOIs (Digital Object Identifiers). While a doi:10.4018/jswis.2009081901 is a valid identifier for a thing of interest, a regular person would not know what to do with it / how to find what it means and where it points to. However, an HTTP URI, such as

```
http://www.igi-global.com/article/linked-data-story-far/37496
```

which is what the above DOI is about, would make more sense for the average person, as she would be able to paste it into the browser and get a webpage with detailed information. Again, Linked Data is shown to be working on the foundations of the WWW.

Now, the HTTP URIs should be dereferenceable, meaning that upon requesting them, the Web server should have knowledge about the resource and return some information, either in human format or in machine-readable format, depending on who requests it.

In the Web of Data, it is essential that URIs can point to real-world things, such as a person or an abstract concept. This is something else from the web document that describes the thing of interest, although they are clearly linked.

When a URI is called, the server uses a HTTP mechanism named content negotiation [49] to decide if it sends back HTML for humans or RDF for machines.

To make URIs - that describe real-world entities - derefereanceble, one can choose from two strategies: 303 URIs and hash URIs, These strategies, with their pros and cons, are well described in a W3C Interest Group note called “Cool URIs for the Semantic Web” [50].

3.2.3. Principle 3: When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)

Dereferencing an URI should lead to useful information on the topic that the URI is pointing at. To be able to understand the descriptions across all topics from all domains, these descriptions should use the RDF data model. As such, RDF would act as the *lingua franca* for all the data on the web [51, p. 15].

The RDF data model and the RDF serialization formats, which are important in this context, will be described later in the thesis.

3.2.4. Principle 4: Include links to other URIs

Being able to go from one object to another is fundamental to the Web of Data, and the name Linked Data says it all about the importance of links.

When these links are external, one can deduce relationships between data in different datasets and surf the Web to find more data or different views on the same data.

RDF links are of three types [51, pp. 20–25]:

Relationship links, which show how a concept relates to another. For example, *foaf:knows* between two persons that know each other or *geo:near* between places that are nearby.

Identity links, which express that concepts from different sources are basically the same thing (for example, the popular *owl:sameAs*). They should be treated in a more relaxed way than the technical definition requires, meaning that concepts connected through identity links should be treated as diferrent views on the same subject, not as perfectly identical concepts.

Vocabulary links, which point from data to the definition of the vocabulary term that describes the data and further to other related vocabulary terms. This enables the very important self-descriptive aspect of the Web of Data.

3.3. The Linking Open Data Project

The most visible result of the linked data efforts, so far, is the Linking Open Data project (LOD). It is an initiative started in 2007, which aims to crawl, analyze and publish statistics on the quantity, type and interconnectedness of all the datasets published as Linked Data on the Web, which have at least some links between them and which have an open license. It is a community activity bootstrapped by the W3C SWEO (Semantic Web Education and Outreach) Interest Group [52, p. 14]. The cumulation of the datasets analyzed in the project is called the LOD cloud.

There was an explosion in the size and number of datasets from May 2007 to August 2014. The number of datasets grew from 12 to 570 in this period, so by a factor of 4750%. There are a total of 188 million triples in the 2014 version (depicted in Figure 3-2), belonging to eight big knowledge domains. The latest statistics are shown in Table 3-1.

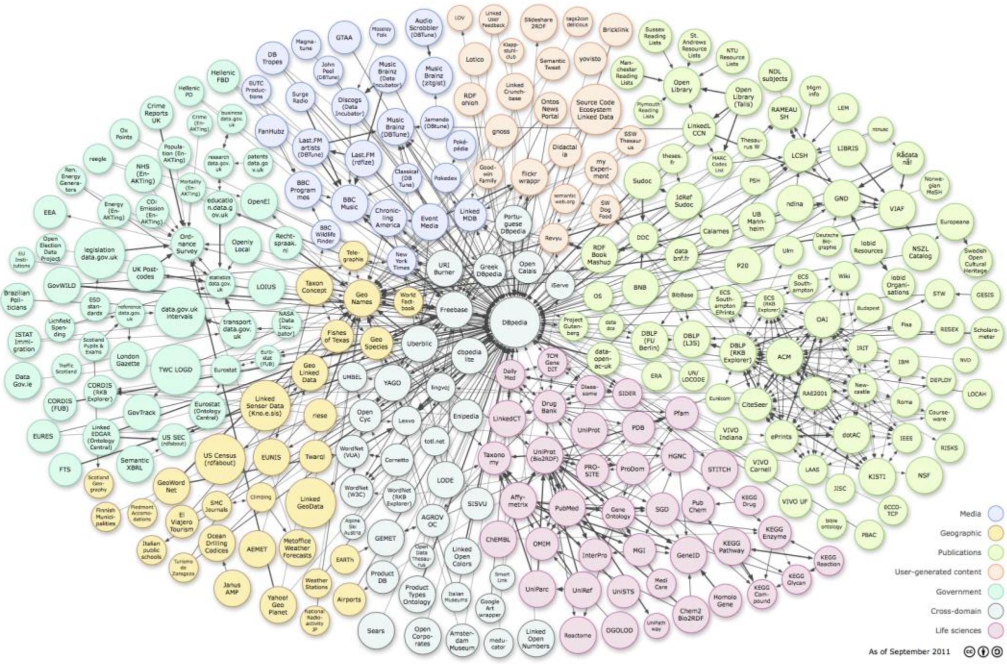


Figure 3-2. The Linking Open Data cloud diagram from August 2014 [46] (licensed as CC BY-SA 3.0)

Table 3-1 State of the LOD cloud in 2014 [53]

Domain	Number of datasets	Percentage of LOD
Media	24	2%
Government	199	18%
Publications	138	13%
Geographic	27	2%
Life Sciences	85	8%
Cross-domain	47	4%
User-generated Content	51	5%
Social Networking	520	48%

From 2007 until 2011, the datasets to be crawled and analyzed were identified based on metadata posted on the Linking Open Data Cloud group[4] which belongs to the Data Hub. In the 2014 State of the LOD, the researchers used a more comprehensive strategy, which meant also including the datasets that could be crawled starting from the existing ones [53].

Some of the most important datasets in the LOD cloud, along with the knowledge domain that they belong to, are described in the following subchapter.

[4] http://datahub.io/group/lodcloud

3.3.1. Prominent datasets

Cross-domain datasets are excellent hubs of information linking the other datasets with them and between them.

DBpedia[5] is the Semantic Web version of Wikipedia, the popular user-created encyclopaedia on the World Wide Web [54]. The English version of DBpedia consists of 4 million objects, most of them being classified using a consistent ontology named the DBpedia Ontology. In time, due to the fact that DBpedia, like Wikipedia, defines millions of concepts from various domains, other Linked Data providers have started to link to DBpedia, making it the most important data hub in the Linked Open Data cloud [55].

Freebase[6] is a community-curated database of people, places and things. It can be accessed as Linked Data either through the Freebase RDF API or through RDF data dumps. It contains 1.9 billion triples [56].

BBC[7] (The British Broadcasting Corporation) is one the largest public service broadcasters in the world. It offers a huge quantity of information from various domains of interest. It massively used Linked Data technologies for the 2012 Olympics website[8] and now is extending them to more domains, such as news, learning and music. Their music portal is interlinked with the MusicBrainz platform and the weather site with Geonames [57].

Geographic datasets also act as important linkages between data that is georeferenced on the web.

Geonames[9] is an open geographical database (released under a Creative Commons Attribution license) which contains data from public sources and which is curated by users through a wiki-style interface. It features more than 10 million geographical names and stores information such as latitude, longitude, elevation, population, administrative area, postal codes and so on.

The data is available as a daily export, or via a search interface, through a number of Web Services (APIs) and as Linked Open Data. Permanent URIs are available for each Geonames resource. It uses a Geoname ontology built using OWL [58]. Feature classes and feature codes are described using SKOS. It is linked to DBpedia articles.

LinkedGeoData[10] is an effort to publish data from the OpenStreetMap[11] project in a Linked Data-style. OpenStreetMap is a popular open database for spatial data, where users added a huge number of items consisting mainly of nodes, ways and relations. The LinkedGeoData uses a lightweight OWL ontology derived from the structure of the data in OpenStreetMap and features links to DBpedia, Geonames and other datasets. It offers a REST interface, a static SPARQL endpoint, a live SPARQL endpoint (with recent changesets from OpenStreetMap), Linked Data via content negotiation and RDF dumps of the data [59]. The data consists of approximately 20 billion nodes.

---

[5] http://dbpedia.org/
[6] https://www.freebase.com/
[7] http://www.bbc.co.uk/
[8] http://www.bbc.co.uk/sport/0/olympics/2012/
[9] http://www.geonames.org/
[10] http://linkedgeodata.org/
[11] http://www.openstreetmap.org/

Due to the fact that researchers were the first to adopt the linked data paradigm, there are some prominent LOD datasets that belong to the research, the publications and the libraries domain.

PubMed is a well-known search portal on life sciences and biomedical topics. It mainly accesses the MEDLINE database of references and abstracts and also includes very old references from the print version of Index Medicus and full-text books. It contains links to full-text articles, some of which are free. As of February 2014, PubMed's database holds over 23 million records. Although its search tools are very powerful, many alternative interfaces were built for PubMed, including a Linked Data interface [60].

ACM (Association for Computing Machinery) is one of the largest scientific and educational computing societies in the world. It counts more than 100.000 members as of 2011 and is headquartered in the New York City. Its Digital Library is composed of the organization's journals, conference proceedings and magazines published since 1950. It holds metadata such as abstracts, references and statistics, which are all open. There are some free full-text articles also. The linked data repository contains publications and details of the authors [61].

DBLP (meaning today "The DBLP Computer Science Bibliography") is a bibliography website, hosted at Universität Trier in Germany, which tracks important journals and proceedings papers in the computer science field. It lists over 2.3 million articles. The Faceted DBLP[12] is a faceted search interface which provides category search based on year, author, venues and so on. It also provides RDF dumps of DBLP data, using the D2R tools for exporting SQL to RDF [62].

Nature Publishing Group is a high impact publisher of scientific and medical information through various journals, online databases and services, Nature[13] and Scientific American[14] being two of the well-known publications of the group. Through its Linked Data platform, it provides access to its datasets in a Linked Data-style. It supports SPARQL 1.0 standard on two query interfaces: a non-streaming browser-based interactive query form (which allows full text searching via an extension) and a streaming service endpoint for remote queries. The SPARQL queries return results in various formats, such as JSON, XML, CSV, N-Triples, RDF/XML and Turtle. The architecture is based on a 5store[15] storage instance with an interface built upon Apache Jena. The datasets include data about articles (published since 1845), product and subject ontologies. These datasets are organized into RDF graphs which are described using the VoID RDF schema vocabulary [63]. The platform also supports the OAI-PMH protocol and can generate dumps of the datasets [64].

The library of Congress is the largest library in the world and the authoritative source of information in the world of libraries. Through the linked open data service [65], it offers its trusted authorities and vocabularies as linked open data, creating a reliable node in the LOD cloud. Users can either search through the linked data service portal or download the available datasets (15 currently). The information is interconnected with other datasets from the LOD cloud, making it very useful.

Bibliothèque nationale de France (BNF) is the largest cultural heritage organization in France and has a strong activity in the linked data world since 2011. BNF provides URIs for resources (through Archival Resource Key Identifiers), open

---

[12] http://dblp.l3s.de/
[13] http://www.nature.com/nature/
[14] http://www.scientificamerican.com/
[15] http://4store.org/trac/wiki/5store

licensing (with attribution) and RDF output / data dumps. Its semantic web portal allows extracting the data in these syntaxes: RDF-XML, RDF-N3 and RDF-NT. It aligns its datasets with other institutions (Library of Congress, German National Library), schemas and ontologies (Dublin Core Metadata Initiative, FOAF) or portals (DBpedia, Geonames). Some of its ontologies and vocabularies are specific to the institution (e.g. bnf-onto) but it generally reuses known vocabularies to provide interoperability [66].

The New York Times started publishing its linked datasets in 2009. Its core data is the New York Times Index, published since 1913, which contains a cross-referenced index of all the names, articles and items that appear in the newspaper. In January 2010 it already published approximately 10,000 subject headings as linked open data, distributed in categories such as people, organizations, locations and descriptions [67]. The dataset constitutes a trusted research resource for students, scholars and librarians throughout the USA and not only.

EUscreen is a European platform that offers multilingual and multicultural access to television heritage. It currently runs a linked open data pilot [68]. EUscreen harvests data from registered affiliated organizations and is itself aggregated by the Europeana Portal. There is a SPARQL endpoint available for querying the information as well as a 4store repository interface (web based). The dataset respects linked data principles (it uses RDF standards, it is open, machine readable, linked to external projects) so it can be rated five stars according to Tim Berners-Lee's schema.

Europeana is Europe's main aggregator of digitized cultural heritage elements. Europeana does not store all the information; rather, it aggregates metadata and, eventually, small sized images about the item. When the user wants to find further information it is redirected to the website of the museum or library that owns the item. The Europeana Linked Open Data Pilot provides metadata and image thumbnails totalizing 20 million objects from over 1500 providers. The pilot offers fully open metadata, as it is licensed as CC0 [69].

OpenLibrary is an open library catalog which anyone can edit. It aims to create a web page for every book in the world. As of May 2014, it holds over 20 million records. It offers a RESTful API, which developers can use to get data in JSON or RDF/XML. The RDF format is available for books and for authors. Data dumps are only available in JSON format [70].

3.3.2. The 5-Star Data Scheme

While the Linking Open Data project is currently the best-known effort in publishing Linked Data, there are many more datasets in the wild that only partially fulfil the envisioned world of Linked Data.

Tim Berners-Lee has proposed a rating system for published data, in which data receives an additional star for each improvement, up to a total of 5 stars [48] (which is how the datasets in the LOD cloud are rated).

Several costs and benefits, for consumers and publishers, of publishing 5 star data, are detailed in [71].

★	The data is available on the web with an open license (whatever format).
★★	The data is openly licensed and is available as structured data (e.g. excel instead of scanned pdf with tables).
★★★	The data is openly licensed, structured and in a non-proprietary format (e.g. csv instead of excel).
★★★★	All the above plus: one uses URIs to identify things, so that people can point at the data.
★★★★★	All the above plus: one links the data to other people's data to provide context.

3.4. The RDF data model in the Linked Data context

3.4.1. The data model

Linked Data works on the foundations of the RDF [72] data model, which allows for complex information to be expressed in simple structures, combined from heterogeneous sources, serialized in known formats, processed unitarily and over the Web architecture.

The basic structure of RDF (and the only one!) is called a triple (or RDF statement), which describes a simple structure made of a subject, a predicate and an object, as below.

Mary	**has**	**dogs.**
subject	predicate	object

An RDF triple can be defined formally as [73]:

$$(s,p,o) \in (U \cup B) \times U \times (U \cup B \cup L) \quad (3.1)$$

where U is a set of URIs, B is a set of blank nodes and L is a set of literals (as in Figure 3-3).

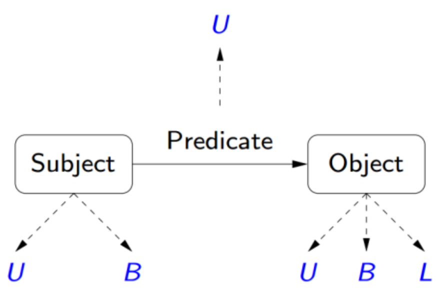


Figure 3-3 RDF formal model [73] (© 2013 Marcelo Arenas. Reproduced by permission.)

The subject should be an URI and the predicate has to be an URI, while the object can be either an URI or a literal (string, number or date). The literal can be of two types: typed (a string having a datatype URI [74]) or plain (a string having an optional language tag) [51, pp. 15–16]. Predicate URIs are described in vocabularies, which are defined for each domain or subdomain of knowledge and which will be explained more in the next subchapter. Sometimes, subjects and objects can be blank nodes, meaning they do not have an URI assigned, a strategy used for simplifying work with data in the interior of a dataset.

Table 3-2 summarizes what the components of a triple can be.

Table 3-2 Summary of triple structure

Triple component / Type	URI	Literal	Blank node
Subject	should be	cannot be	can be
Predicate	has to be	cannot be	cannot be
Object	can be	can be	can be

Several triples form an RDF graph, where the subjects and objects are nodes and the predicates are directed arcs. Linking external RDF triples between them leads to the giant global graph envisioned by Tim Berners-Lee [75].

3.4.2. Remarks on RDF and Linked Data

Developers working with Linked Data should strive to use URIs instead of literals in objects (especially when publishing the data in the LOD cloud). This way they can extend the information by linking it to internal or external resources [52, p. 28] (corresponds to principle P1 and P4 from subchapter 3.2).

In addition, it is better not to use blank nodes, as these cannot be referred back, from inside or outside, given the lack of a unique identifier[51, p. 17], [52, p. 35].

Even though RDF allows for multiple types of URIs to be used (e.g. HTTP, FTP, ISBN, DOI), HTTP URIs should be used instead of other types of URIs, for these URIs to be resolvable [52, p. 12] (corresponds to principle P2). Also, RDF URIs should lead to information (corresponds to principle P3).

RDF allows for merging data from different sources, data that is described by different schemata and that is structured with different complexities. This way, it fulfills the vision of Linked Data for a global graph of information.

3.4.3. RDF vocabularies

While the RDF data model specifies a structure for the information, it does not define domain-specific terms that describe the components of a triple. This is done using vocabularies and ontologies, which can be defined using languages such as RDFS and OWL. The best part is that these schemata are described using the same concepts – RDF – and the same principles of Linked Data, so they are easy to follow and to be read by machines.

RDFS is a basic ontology for RDF, consisting, simply put, in classes and properties that describe what type the resources are and what properties they have [76].

The most used RDFS terms are:

- rdfs:label – provides human-readable name for a resource
- rdfs:comment – provides human-readable description of a resource
- rdfs:type – denotes the class a resource belongs to
- rdfs:subClassOf – states that all the instances of a class also belong to another class
- rdfs:subPropertyOf – states that all the resources having a property also have another property
- rdfs:domain – states that resources which have a property are an instance of a certain class
- rdfs:range – states that all values of a property are instances of a certain class
- rdfs:seeAlso – specifies that a resource provides more information about another resource

OWL is a more expressive ontology as it allows to define more complex relations between resources [77]. It is not used to its full extent in the Linked Data world [44], as it can put a burden on the processing resources for the data. However, some terms are highly popular in the Linked Data world, such as owl:sameAs, which is used as a social contract that says one resource is similar to another, so it can be used to find more data in a follow-your-nose fashion, or owl:inverseOf, which states that one property is the inverse of another, a fact which can be used to crawl information more easily.

Vocabularies have evolved organically, as there is no central authority that manages them. According to [52, p. 38], some have become so popular that they are now considered core vocabularies. Others, not so much used, but important for their knowledge domain, are authoritative vocabularies.

Table 3-3 lists some of these vocabularies and what concepts they address.

Table 3-3 Some core and authoritative vocabularies [52, Fig. 2.11, 2.12]

Type	Name (prefix)	Describes
Core	Friend-of-a-Friend (foaf:)	People and organizations
Core	vCard (vcard:)	Contacts, business cards
Core	Description of a project (doap:)	Projects
Core	Dublin Core (dc:)	Publications
Core	SKOS (skos:)	Vocabularies
Authoritative	SIOC (sioc:)	Online communities
Authoritative	Geo (geo:)	Geography
Authoritative	Good Relations (gr:)	Business products
Authoritative	BIBO (bibo:)	Bibliographies
Authoritative	Creative Commons (cc:)	Licenses

Good resources for finding vocabularies to use are Linked Open Vocabularies[16] or LOD Stats[17].

There are many situations where one does not find a vocabulary that suits the targeted use case. In this case, one should try to extend a known vocabulary, by using terms like owl:sameAs. If this is still not feasible, one can try and built her own vocabulary. A good resource on this issue is "Semantic Web for the Working Ontologist" [78]. A vocabulary should only express what it is intended for and not more than that, in order to keep itself simple and agile (suitable for inference engines) [51, p. 63].

3.4.4. RDF serialization formats for Linked Data

The RDF data model is only a specification of how the data is structured and not a syntax for writing down the information. For this, several formats have emerged over time, each having various advantages and disadvantages, depending on the context they are used in. The five RDF serialization formats described in this chapter are:

- RDF/XML
- Turtle
- N-Triples
- RDFa
- JSON-LD

They can be mixed together (various sources can use different formats) given that one uses a software to convert data from one format to another.

RDF/XML

It is the oldest format standardized by W3C and is still widely used [79]. Standard XML processors already know how to handle it, so it is a good choice in enterprises. However, it is hard to follow with the human eye, due to the complexity of its syntax.

```
<?xml version="1.0" encoding="utf-8" ?>
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" >
  <rdf:Description rdf:about="http://dbpedia.org/resource/Timi%C8%99oara">
    <rdf:type rdf:resource="http://dbpedia.org/class/yago/CitiesInRomania" />
    <rdfs:label>Timişoara</rdfs:label>
  </rdf:Description>
</rdf:RDF>
```

[16] http://lov.okfn.org/dataset/lov/
[17] http://stats.lod2.eu/vocabularies

Turtle

Its name comes from Terse RDF Triple Language and was specifically designed to be read and written easily by humans. Subject, predicate and object follow each other on the same line, which ends with a full stop. Some other syntax shortcuts are used for multiple triples with the same subject or the same object. URI prefixes are allowed to further simplify the code [80].

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

<http://dbpedia.org/resource/Timi%C8%99oara>
  rdf:type <http://dbpedia.org/class/yago/CitiesInRomania> ;
  rdfs:label "Timişoara" .
```

N-Triples

N-Triples is similar to Turtle, minus the shortcuts in the syntax. Because of this feature (more exactly, the lack of it), the format can be read line by line, so it is very approapriate for large dumps of RDF data. The disadvantage is that it has a very large size [81].

```
<http://dbpedia.org/resource/Timi%C8%99oara>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://dbpedia.org/class/yago/CitiesInRomania> .
<http://dbpedia.org/resource/Timi%C8%99oara>
<http://www.w3.org/2000/01/rdf-schema#label> "Timi\u00C8\u0099oara" .
```

RDFa

It was designed to be embedded in HTML files, so it is easier for developers to adopt and exploit [82]. Technically, it does not require additional server configurations (as opposed to serving RDF/XML or other formats). It was standardized by W3C and it is heavily used for search engine optimization along with the schema.org scheme.

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN"
"http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd">
<html xmlns="http://www.w3.org/1999/xhtml"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">

  <head>
    <meta http-equiv="Content-Type" content="application/xhtml+xml;
charset=UTF-8" />
    <title>Information about Timisoara</title>
  </head>

```

```
  <body>
    <div about="http://dbpedia.org/resource/Timi%C8%99oara"
typeof="http://dbpedia.org/class/yago/CitiesInRomania">
      <span property="rdfs:label">Timişoara</span>
    </div>
  </body>

</html>
```

JSON-LD

It was designed for developers that are accustomed to the JSON format. Its advantage is that almost all of the platforms have libraries that can parse JSON. It is very lightweight in size. It is a W3C recommendation as of 16 January 2014 [83].

```
{
  "@context": {
    "dbpedia": "http://dbpedia.org/resource/",
    "rdf": "http://www.w3.org/1999/02/22-rdf-syntax-ns#",
    "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
    "xsd": "http://www.w3.org/2001/XMLSchema#"
  },
  "@id": "dbpedia:Timi%C8%99oara",
  "@type": "http://dbpedia.org/class/yago/CitiesInRomania",
  "rdfs:label": "Timişoara"
}
```

3.5. SPARQL for Linked Data

Every data model needs a query language. In the case of Linked Data, the query language is SPARQL [84]. SPARQL is a recommendation by the W3C and is similar in nature to SQL, so it is fairly easy to grasp by developers accustomed to relational databases. SPARQL allows querying, simultaneously, several different datasets on the web, whether they are RDF files or SPARQL endpoints. While the initial release allowed only reading the data, the later SPARQL 1.1 Update allows also writing data in a file or triple store.

A typical SPARQL query is organized as follows:

```
#set prefixes for URIs
PREFIX
PREFIX
…
#specify the information that is wanted
SELECT / DESCRIBE / …
#state which RDF graphs are being queried
FROM …
#specify what to query for
WHERE {
  …
```

```
}
#add constraints to the query: ordering, slicing etc
ORDER BY / LIMIT / …
```

3.5.1. Types of SPARQL queries

SELECT

The SELECT query returns some information, which it selects from the dataset based on some constraints.

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>
SELECT (SAMPLE(?name) AS ?famous_name)
FROM <http://dbpedia.org>
WHERE {
  ?person dbpedia-owl:birthPlace
<http://dbpedia.org/resource/Timi%C8%99oara> .
  ?person foaf:name ?name .
}
GROUP BY ?person
LIMIT 10
```

This query returns the first 10 names of famous people that were born in Timisoara, as it can be found on DBpedia.

ASK

The ASK query is used to test if a query would return some results or not. Its response is a true or false value. It is mainly used to test if certain conditions would return results, before actually requesting those results with SELECT, as it is much faster.

```
PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>
ASK
FROM <http://dbpedia.org>
WHERE {
  <http://dbpedia.org/resource/Timi%C8%99oara> dbpedia-owl:abstract
?abstract .
}
```

This query returns true if the city of Timisoara has a description (an abstract) on DBpedia.

DESCRIBE

The DESCRIBE query returns all the triples for which the inputted URI is a subject, thus describing that resource. Such a query is often executed as a prestep for getting the information that is ultimately desired.

```
DESCRIBE <http://dbpedia.org/resource/Timi%C8%99oara>
FROM <http://dbpedia.org>
```

This query returns all the triples describing the city of Timisoara on DBpedia.

CONSTRUCT

The CONSTRUCT query returns an RDF graph that is personalized by the developer using constraints in the WHERE clause.

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>
PREFIX vcard: <http://www.w3.org/2001/vcard-rdf/3.0#>
CONSTRUCT {
  ?person vcard:FN ?name
}
FROM <http://dbpedia.org>
WHERE {
  ?person dbpedia-owl:birthPlace <http://dbpedia.org/resource/Timi%C8%99oara> .
  ?person foaf:name ?name .
}
```

This query returns an RDF graph with vCard information of all the famous people that were born in Timisoara, as stated on DBpedia.

The SPARQL 1.1 Update queries

The SPARQL Update language is used for specifying and executing updates to RDF graphs. It may be used only on SPARQL endpoints that one has permission to write to (where standard access and authentication techniques are in use [85]). The most commonly used is the INSERT query, which adds triples to a database.

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
INSERT DATA
{
  <http://www.upt.ro/ > foaf:name "Politehnica University of Timisoara" ;
     foaf:based_near <http://dbpedia.org/resource/Timi%C8%99oara> ;
     foaf:birthday "1920" .
}
```

This query inserts three RDF triples into an RDF graph.

Other "write" queries are CREATE, DROP, LOAD, DELETE or CLEAR.

3.5.2. SPARQL result formats

SPARQL returns results in various formats. Some of the most used are XML, JSON, CSV and TSV [86].

Below is an example of how a result in XML looks like:

```
<?xml version="1.0"?>
<sparql xmlns="http://www.w3.org/2005/sparql-results#">
<head>
  <variable name="person"/>
  <variable name="famous_name"/>
</head>
<results distinct="false" ordered="true">
  <result>
    <binding name="person"><uri>http://dbpedia.org/resource/Vasile_Deheleanu</uri></binding>
    <binding name="famous_name"><literal xml:lang="en">Vasile Deheleanu</literal></binding>
  </result>
</results>
</sparql>
```

3.6. Publishing Linked Data

Publishing Linked Data should follow some guidelines [51, pp. 41–84], [52, pp. 233–238], in order for the process to respect the Linked Data principles and to be useful to an application that consumes Linked Data. The steps are detailed in the following paragraphs.

1) Minting HTTP URIs

"Minting" is a colloquial term for designing URIs. "Cool" URIs [50] should:

- be unique;
- hide implementation details (for example, no .php or .aspx at the end of the URI);
- be human-readable (avoiding IDs and using natural keys for the terms);
- be persistent and created in a namespace that is owned by the creator;
- and should be different for the real resource and for the document describing it.

The following URI pattern is common [52, p. 235]:

```
[authority] / [container] / [key for the item]

Example: http://dbpedia.org/resource/Timisoara
```

For designing government URIs, a useful resource can be found at [87].

2) Choosing vocabularies

Various vocabularies for different domains have been presented in subchapter 3.4.3. If none of them is usable for a certain use case, one should consider extending an existing one, or, finally, creating one from scratch.

Good resources for customizing or creating vocabularies can be found in [88] and [78]. Some great software tools that ease the development process are TopBraid Composer Standard Edition[18] and Protégé[19].

3) Including RDF data

According to [51, p. 45], the RDF description for a particular resource should include:

1. triples that describe the resource with literals
2. triples that describe the resource through links to other resources
3. triples that act as incoming links
4. triples describing related resources
5. triples describing the description itself
6. triples describing the dataset that the resource belongs to

The first two kinds of triples are the basis of Linked Data. The third type is useful, in order to help increase the knowledge derived from crawling the Web of Data. However, care must be taken, in order not to put a heavy burden on the data, due to its significantly increase in size. This is also the disadvantage of having too many triples belonging to the fourth category. The fifth and sixth categories are often skipped, being deemed unnecessary, although they are very important for data discoverability and quality assessment.

4) Publishing metadata

It is important for the data to be self-described, a fact that enhances discoverability and allows for checking of licensing terms and other information. According to [51, p. 48], there are two main approaches: using Semantic Sitemaps [89], which are similar to the classic Sitemaps protocol that Web developers use to enhance their SEO, and voiD descriptions (the Vocabulary of Interlinked Datasets) which is a vocabulary specifically tailored to describing Linked Data sets [90].

Useful metadata include the date of creation, the date of the most recent update, the creator, the publisher or the usage license.

[18] http://www.topquadrant.com/tools/modeling-topbraid-composer-standard-edition/
[19] http://protege.stanford.edu/

5) Interlinking the data with other data(sets)

This is the foundation of the Linked Data vision. The interlinking process can be manual, for smaller datasets, or (semi-)automatic, for larger datasets. There are some RDF predicates that are very useful for interlinking scattered resources, such as foaf:knows, foaf:based_near, foaf:topic_interest, owl:sameAs, owl:equivalentTo or rdfs:seeAlso. To find similar resources, developers can make use of platforms such as sameas.org, which provides sets of URIs that apparently refer to the same resource. In addition, tools like Silk [91] and LIMES [92] are heavily used for streamlining the interlinking process.

6) Publishing the data – patterns

According to [51, pp. 69–71], publishing Linked Data depends greatly on the type of input data. From this point of view, we have:

1. From queryable structured data (RDBMS, API) to Linked Data
2. From Static Structured Data (CVS, XML) to Linked Data
3. From Text Documents (with tools such as Open Calais[20], OntosMiner[21] or DBpedia Spotlight[22]) to Linked Data

Common Linked Data publishing workflows are shown in Figure 3-4.

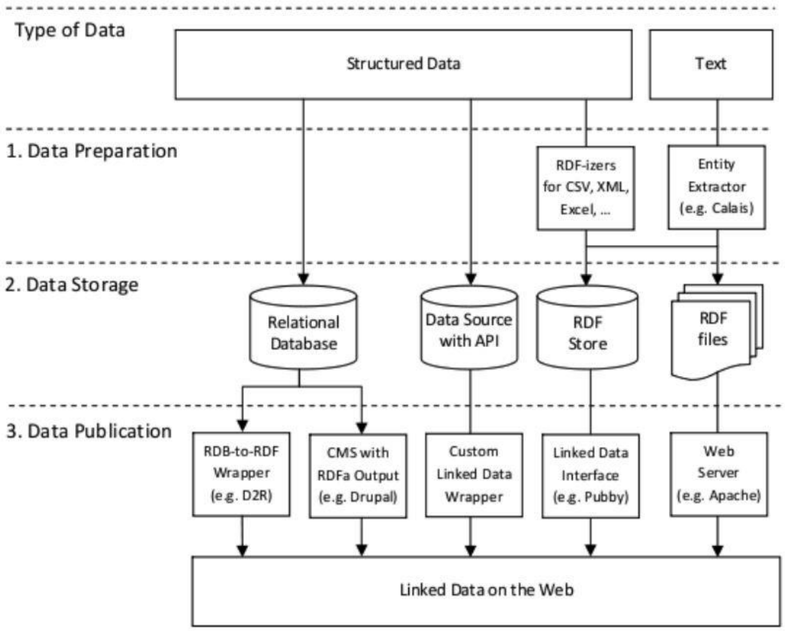


Figure 3-4 Linked Data publishing workflows [51] (© 2011 Morgan & Claypool Publishers. Reproduced by permission.)

[20] http://new.opencalais.com/
[21] http://www.ontos.com/products/ontosminer/
[22] http://spotlight.dbpedia.org/

3.7. Consuming Linked Data

The researchers in [51, pp. 97–98] propose an architecture for Linked Data applications and three patterns:

1. the Crawling Pattern (sources are crawled and cached in a single datastore),
2. the On-The-Fly Dereferencing Pattern (URIs are dereferenced on the spot) and
3. the Query Federation Pattern (a query is issued to a fixed known set of sources).

The architectural patterns are depicted in Figure 3-5.

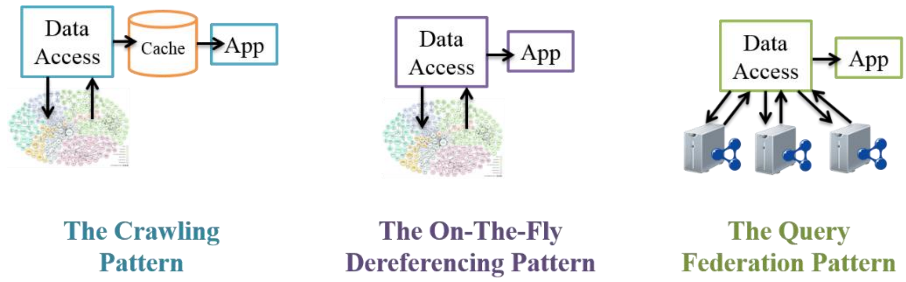


Figure 3-5. Linked Data architectural patterns [93] (Published by Euclid Project, licensed as CC-BY 3.0, based on [51, pp. 97–98])

The steps involved [51, pp. 98–105] are detailed below.

1) Accessing the data

Usually, it is done by dereferencing URIs into RDF descriptions and traversing the discovered RDF links towards other data. Additionally, accessing new data can be made by quering SPARQL endpoints, downloading RDF dumps, extracting RDFa info or accessing APIs of Semantic search engines.

2) Vocabulary mapping

Since data comes from different datasets with various schemata, the vocabularies must be mapped to a single target schema. One can use various techniques, such as relying on terms like owl:equivalentClass or rdfs:subClassOf. Tools like R2R [94] can help in this endeavor.

3) Identity resolution

This means identifying that the same resource is present in different datasets. It is done by analyzing links (owl:sameAs) and comparing various parameters (name, geographical position etc.). One of the well-known tools that can help with identity resolution is Silk [91].

4) Provenance tracking

Data provenance is useful for deriving data quality indicators and for tracking the original source of the data.

5) Data quality assessment

Data on the Web is "dirty" and the same stands true for Linked Data. There are several approaches to data quality detection: content-based heuristic, context-based heuristics or rating-based heuristics. After the data has been assessed, one can rank the data, filter the data or fuse the data.

6) Caching web data locally

To store RDF data locally, developers use triple stores. Analyses of triple store performances are done with benchmark systems such as BSBM [95].

7) Exploiting the data

Data can be accessed via SPARQL or RDF API [96] and exploited in various applications.

3.8. Linked Open Government Data in the context of Smart City Applications

3.8.1. Introduction and Definition of Terms

We are witnessing a prolific activity during the recent years regarding the publication and exploitation of Open Government Data and, alongside it, a great debate on the advantages of undertaking this task and the challenges towards accomplishing it.

To better understand this field, I will first of all define some terms that are discussed in this chapter.

The term Open Government, often used interchangeably with the term Gov2.0, has a long history of usage and debate and is often perceived as being for the traditional government what Web2.0 is for the classic web. Tim O'Reilly defines Gov2.0 as being "the use of technology - especially the collaborative technologies at the heart of Web 2.0 - to better solve collective problems at a city, state, national, and international level" [97].

The most widely analysed and discussed aspect of these issues, the Open Data, is defined by the Open Knowledge Foundation since 2006. It was revised several times, the most recent version being the Open Definition from 2014:

"Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)." [4].

As such, Open Government Data would be the fuel that would propel the collaboration and transparency for a better government for the people.

As the greatest amount of Open Government Data exists on the Web, a new flavour of Open Government Data has been coined, one which benefits from a Web-style approach, namely Linked Open Government Data.

The authors in [98] define Linked Open Government Data as such:

"Linked Open Government Data are all stored data of the public sector connected by the World Wide Web which could be made accessible in a public interest without any restrictions for usage and distribution."

Open Government Data can be found today in many open data portals of countries from all over the world. While in 2011 there were 21 countries that published Open Data, with 1/3 of them publishing also Linked Data [99], today the number has grown drastically. From 77 countries studied in 2013, over 55% of them have seen a smaller or greater implementation of Open Government Data policies [100]. In the most recent crawling of the Linking Open Data cloud, Government Data has been found to occupy 18% of the total number of linked datasets in the wild [53].

This data, either raw or linked, has been put for good use in the form of mash-ups, platforms, visualizations, web apps and mobile apps [101], [102] .

A prominent category of these apps is composed of Smart City applications that have the role of fostering a better life and engagement of citizens in urban spaces.

Smart City is defined as such: "A city can be defined as 'smart' when investments in human and social capital and traditional (transport) and modern (ICT) communication infrastructure fuel sustainable economic development and a high quality of life, with a wise management of natural resources, through participatory action and engagement" [103].

3.8.2. Approaches in publishing Open Government Data

Open Government Data can come in multiple formats, each one with its own advantages and disadvantages.

First of all, there are the raw data files, which are the most published type of data, because the government employees work directly with them. The authors in [98] have classified them in different flavours. For example, CSV, XML or RDF are machine readable, have an open format and the specifications are available. Although PDF is an open format with specifications, it is not machine readable. XLS is machine readable, but is not an open format [98].

Raw files are plagued with issues such as the lack of context, as there is rarely metadata attached to the file to explain what the data inside it means, and the lack of unique identifiers, which leads to difficulties in understanding the data. As such, the data cannot be linked with other files for a better understanding of it and for creating better applications [102].

Another type of data is the one stored in Relational Databases. Although they are stable and well-researched in closed fields, on the Web, they have the disadvantage that they are not flexible, are difficult to share and the developer must have known the schema before using the Relational Database [102].

A third approach consists of APIs, which are heavily used nowadays for providing data and services on the Web. However, they are opaque (the internal structure and data is hidden to the developer), they are not standards based and they only provide answers for what they were designed to answer for [102].

Due to all these approaches being mixed together, the available Open Government Data is hard to process, visualize and integrate.

Linked Data, as part of the Semantic Web vision [45], has been researched as a suitable approach for overcoming these issues. Linked Data draws its force from using a representation standard (RDF) and a query mechanism (SPARQL) for enhancing the interoperability of the data. Based on the principles issued by Tim Berners-Lee [48], Linked Data has been established as de facto language for powerful publication and exploitation of Open Data. However, Linked Data does not come without its challenges, such as frequent changes and instability of the data, or data quality and provenance issues [104].

3.8.3. Linked Open Government Data lifecycle stages and initiatives

In order to publish Linked Open Government Data in a complete and useful manner, one has to pass through each stage of the Linked Open Government Data lifecycle. Such a lifecycle was proposed by [105] and involves specification, modelling, generation, publication and exploitation of the data.

The specification stage consists of identifying and analysing the government data sources, designing the URI schema and defining the license to be attached to the data.

The modelling stage requires choosing the right ontology to describe the data. Best practices involve reusing an existing ontology, if possible, extending an existing one or, lastly, creating a new ontology that fits the purpose.

The generation stage consists of transforming the data usually from raw data files to RDF, cleaning the data and linking it to other datasets.

The publication stage consists of publishing the dataset itself, publishing the metadata and enabling discovery (e.g. including it in the Linking Open Data cloud).

In the final step, exploitation, the data is integrated, processed, explored or visualized in various mash-ups and applications.

In Spain, within the Ciudad2020 project, the researchers applied the Linked Data Lifecycle in 4 vertical domains of the Smart City: Transport, Environment, Energy and the City [106]. In the specification phase, the researchers deal with static data sources, such as information on museums and libraries or energy certificates of buildings, on which they use an ETC (Extract, Transform and Load) process to generate RDF formatted data [107], and also with Streaming Data sources, such as information on available bikes and slots in various bike stations or information from weather stations, on which they use the morph-stream technology [108]. For each major domain, a use case is presented from the exploitation phase.

The biggest experiment so far is the TWC LOGD portal that supports the deployment of Linked Open Government Data from the US data platform to enable greater consumption and reutilization of it [109]. It consists of an open source tool called csv2rdf4lod, which does heavy automatic, plus some manual, transformation of the data from raw to semantic. The authors identify 3 stages for data publication: the catalog stage, where each dataset is identified and categorised for a complete inventory; the retrieval stage, where for each dataset a snapshot is created, including the data itself, the metadata, the point in time etc; and the conversion stage, in which data is converted to RDF using various levels of conversion, which helps overcome some of the usual challenges in this process.

UK's open data portal, data.gov.uk, is another big contributor to the Linked Data Web. The authors in [110] identify 4 important research challenges in exporting open government data to the linked data web i.e. discovering the datasets and migrating them to RDF; integrating open government data into the Web of Linked Data, by means of popular ontologies, and linking the data; identifying the points of reference that can link datasets; and building apps to explore and use the data. The authors describe the approach taken to tackle these challenges inside the EnAkTing project.

In [111], the authors propose a roadmap for Linked Open Government Data, which consists of: the open stage, where the government publishes open datasets on centralized platforms; the link stage, in which both government and citizens clean, convert, enhance and link the data; and the reuse stage, where citizens and companies create value-added apps with the data. The authors envision 3 grand challenges associated with these stages i.e. a million catalogued datasets, a million linked datasets and a million Linked Open Government Data applications.

In [112], the authors adopt the lifecycle of Linked Open Government Data from [113] and describe a vision for an ecosystem of Linked Open Data visualizations composed of three layers: the Linked Open Data cloud, at the bottom; the Choreography layer, in the middle, that analyses the datasets and can suggest suitable visualizations; and the Linked Open Data exploration layer, consisting of tools for various types of visualizations: domain specific, spatial faceted, faceted browsing etc. Two case studies are presented, namely the Digital Scoreboard of the European Commission and the Financial Transparency System of the European Commission.

The authors in [106] do a review of Linked Data initiatives in Smart City contexts, classifying them by domain, type of data used and target users. The review shows various combinations of open data, linked data and streaming data being used in multiple domains for various purposes, but points out that there is still a lack of approaches for multi-domain, multi-user and multi-nature data.

An interesting application presented in [106] is Zaragoza Bizi, a map browser that displays static data and real time data about bikes and cycle paths in Zaragoza, Spain. The citizens can use the browser to see available bikes, points of interest near the bike stations, routes between stations and so on. Behind the browser there is RDF data and a SPARQL endpoint that uses open government data.

3.8.4. Case study - Timisoara Street History

In the following subchapter, I present a case study of a Smart City application built using Linked Open Government Data.

Timisoara Street History is a map-based web application that displays the current and past street names in Timisoara using an intuitive, usable map that displays well on desktops and smartphones. The code is adapted for Timisoara, Romania, with the written consent of Noah Veltman, the author of the original project done for San Francisco [114].

The project aimed to prove the usefulness of a dataset published recently on the Government's Open Data Portal by the Timisoara City Hall. The dataset is a comprehensive list of the current and past street names in the city. As such, citizens are now able to browse through the streets in Timisoara and learn how they were called and how the name changed during the previous century.

The specification stage consisted of choosing the above dataset which has a proper license (the Romanian Open Government License).

In the modelling and generation stage, the dataset, in XLS format, was transformed to RDF using the Open Refine tool.

A documented set of steps [115] was used to get the list of streets and their coordinates from OpenStreetMap which has the license Open Data Commons Open Database License (ODbL). This dataset was needed to be able to show the streets from the government dataset on the map.

The generation stage also involved linking the two datasets using the Silk tool [91], based on the Levenshtein distance for the current names of the streets.

The exploitation stage consisted of developing the app itself which is online at http://apps.mysmartcity.ro/Street-History/.

Figure 3-6 displays a screenshot of the application.

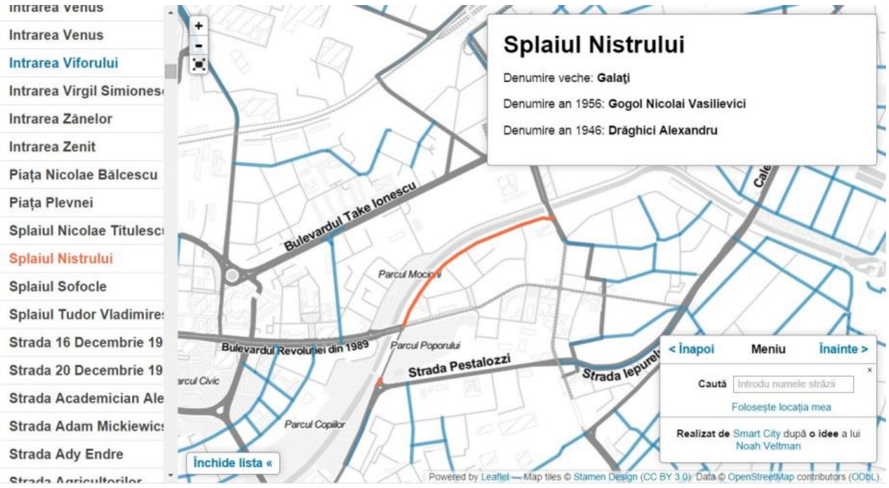


Figure 3-6. Screenshot of the application – highlighting a street on the map to view its past names

3.9. Conclusion

The Web of Documents, that seem to be the working base for augmented reality now, has evolved into a Web of Data, due to the adoption of a set of principles and technologies belonging to the Semantic Web. Although the full vision of the Semantic Web has not emerged yet, a more practical part of it, named Linked Data, has grown in popularity and adoption. Linked Data works on standards of the Semantic Web (a representation standard, RDF, and a query mechanism, SPARQL) and is focused on uniquely identifying things in the world, making their description accessible on the Web and interlinking them to provide more context and information. The Linked Open Data cloud of linked datasets has emerged as a proof of the adoption of such principles.

This chapter is a review of the current landscape in which Linked Data is evolving. I present the four fundamental principles of Linked Data, along with the Linking Open Data project, the proof of the materialization of these principles in the real world. Next, I explore the RDF data model in a Linked Data context and I go over the major RDF serialization formats (RDF/XML, Turtle, N-Triples, RDFa, JSON-LD) and the SPARQL types of query (select, ask, describe, construct). Additionally, I review some methods for publishing and consuming Linked Data, showing that several tools have been developed which can be used successfully in a production environment, given that the developers comply with good practices in this field: minting “cool” URIs, choosing the right vocabularies, publishing metadata and interlinking the datasets – to name just a few.

These theoretical issues are meant to help the reader in understanding the major principles and tools of Linked Data. One of the conclusions that can be derived from this critical review is the advantage that Linked Data brings in exploiting more and better information, including in augmented reality scenarios.

The chapter also touches on a subject important to the present thesis, Open Government Data, which has seen a strong building momentum lately. Linked Data can help with the publication and dissemination of it so it can be better processed and understood by applications and, ultimately, citizens. There is a lifecycle in publishing Linked Open Government Data and the steps that it involves have been used in many initiatives around the world.

I present some of these initiatives and a case study, the Timisoara Street History application, in which the Linked Open Government Data lifecycle was used. The application uses Open Government Data recently published on the Romanian Government Open Data Portal. I also use other linked open datasets, not from governmental sources, but from popular user-generated data platforms. The combination has proven to be a powerful one, as the datasets complement each other. Overall, the case study presented proves the usefulness of the Linked Open Government Data lifecycle in maximizing the value of the data.

The theoretical contribution of this chapter to the present thesis consists in an overview of linked data and a critical study of tools and architectures for publishing and consuming linked data, while the practical contribution consists in the implementation and description of a Smart City application based on Linked Open Government Data. Some results presented in this chapter have been published in [116] and [117].

4. INTEGRATION OF LINKED DATA IN MOBILE AUGMENTED REALITY APPLICATIONS FOR TOURISM

4.1. Introduction

Mobile augmented reality has emerged as the most popular and convenient form of augmented reality, mainly due to the proliferation of mobile devices and ubiquitous computing. Several fields of use have proven to be proper for deployment of mobile augmented reality technologies. In this chapter, I argue for the natural fit of mobile augmented reality applications in the field of tourism and I present some research and commercial projects in this area. Next, I address the issue of content sources for mobile augmented reality applications. The usual approach of developers and content publishers in this respect is to use isolated databases for content, which limits the information depth of surroundings exploration for tourists. I identify the benefits of exploiting Linked Data principles and technologies for enriching content in mobile augmented reality applications for tourists and explore some projects that tackle this approach.

Several aspects have been identified in the literature as being relevant to Linked Data integration. I place these aspects in the context of mobile augmented reality applications for tourism and present current efforts in research in these areas. The most important aspect is geodata integration i.e. ontology matching and entity disambiguation during integration of several heterogeneous and overlapping datasets from the Linked Open Data cloud, with the purpose of providing a consolidated, enriched and more relevant unique dataset for tourism purposes.

Other discussed aspects are the assessment of the quality of the data, usage of provenance information and inference of trust.

4.2. Related Work and Arguments for Linked Data integration

4.2.1. Mobile Augmented Reality Applications for Tourism

Mobile augmented reality has seen tremendous growth in the recent years. The set of technologies commonly called mobile augmented reality is deployed successfully in a number of fields, such as tourism and navigation, entertainment and advertisement, training and education, assembly and maintenance and so on [33].

Due to its inherent strong alignment with the real world [18], mobile augmented reality is well suited for enabling location-aware applications and services. Also, mobile augmented reality applications provide the ground for innovative services within the ecosystems of smart cities [118].

Mobile augmented reality is a natural fit for tourism applications and services, being regarded as having a significant impact in this area [119], due to its ability of enhancing the surroundings of the tourist [120]. A recent online survey [121] showed that location based services, augmented reality browsers and tourism and travel applications are the most used types of augmented reality applications, after games and entertainment applications based on these technologies.

Additional benefits of augmented reality in urban heritage tourism are presented in [122].

One of the first experiments in exploring the surrounding urban landscape with a mobile augmented reality system was the Touring Machine [123], followed by the MARS system [124].

Researchers propose in [125] a system called Wikireality which consumes text information and images from Wikipedia and overlays them on nearby Points of Interest. A more complex project is pursued in [126], where the authors describe a large-scale mobile augmented reality system that overlays 3D footprints of buildings and their name on top of urban buildings.

Several similar applications have also been developed in the commercial area. An overview [127] of mobile augmented reality applications for tourism classifies them as augmented reality browsers (such as Layar, Junaio and Wikitude), dedicated augmented reality applications (such as Acrossair, Augmented Reality UK and WhereMark) and augmented reality view-enabled applications (such as mTrip, TripWolf and Yelp).

A popular category is composed of mobile augmented reality applications that display historic images on top of the current landscape. Well known projects are PhillyHistory in Philadelphia [128], StreetMuseum in London [129] and "Paris, then and now" [130].

Cultural tourism is a subdomain that is also proper for deployment of augmented reality technologies. One of the first projects in this area is PRISMA [120], an interactive visualization system that is a combination of tourist binoculars and augmented reality. In [131], the authors describe a tracking framework that uses different tracking flows for more efficient identification of targets, to be used in augmented reality cultural heritage tours. The LIMES project is a recent augmented reality platform developed to raise awareness and present in an innovative way the ancient Roman Frontier called the Roman Limes [132]. The Points of Interest are added using the backend of the platform and are made available to users through popular augmented reality browsers.

4.2.2. Content-Related Limitations of Mobile Augmented Reality Applications

The plethora of research projects and commercial applications leveraging augmented reality technologies in the mobile tourism field has shown the benefits of implementing this set of technologies, but has also highlighted some challenges that need to be further tackled.

Content is one of them and is "a critical aspect for acceptance" [133], revealed a usability evaluation of the MobiAR project. The study pointed out that, when selecting a Point of Interest, people would like to be shown how to get there (route navigation) and would like to see additional information (like parking and opening hours).

This type of functionality is limited by the nature of content sources that augmented reality applications for tourism (and not only) usually use today, and for which projects like MobiAR [134] and LIMES [132] are representative: isolated silos of information, typically relational database management systems, which store only a few attributes about the Points of Interest. Due to this simple form of storing content, more complex functionalities of the augmented reality application are hard to develop (e.g. reasoning). The user can simply see the Points of Interest around her and select one to see some more details (he cannot ask, filter etc.).

To provide the users with more information, the Points of Interest reference a link that typically opens another application on the mobile device (e.g. the mobile browser). Thus, the immersive experience in the augmented reality application is discontinued.

Moreover, popular augmented reality browsers (Junaio, Wikitude) consume Point of Interest information (they basically issue queries to an augmented reality server, which in turn queries a server, that stores Points of Interest, using a radius filter) using different standards [135], which highlights the "lack of interoperability across mobile platforms" is this respect [136]. Integrations and reusability are massively hindered, due to the heterogeneous landscape of standards (or lack of) for augmented reality applications [137].

More flexible strategies are required for managing data sources and for making them consumable by different applications [33].

4.2.3. Related Work on Linked Data in Mobile Augmented Reality Applications for Tourism

There is a trend to develop augmented reality applications that integrate heterogeneous sources of content, which have various degrees of data quality (ranging from user-generated data to open government data). [138]

This type of openness and integration can be achieved by leveraging Linked Data principles and technologies. These can help by dynamically selecting and integrating data from various sources, thus providing enriched content, by means of enabling the exploitation of more contextual information, not only location, and by creating a Web-like browsing experience for the augmented reality application user [139].

The authors describe in [140] a framework in which an augmented reality application for urban navigation features content that is organized according to Semantic Web principles. Reasoning with OWL takes place to help the user navigate in the urban space.

Some of the early efforts in integrating Semantic Web in mobile applications are mSpace Mobile and DBpedia Mobile, although these applications do not leverage augmented reality technologies. mSpace Mobile is an exploring application that keeps track of location and context, while integrating various resources from the Web in a Semantic Web style [141]. The prototype was tested in London, making use of such sources as Open Guide to London, IMDb, BBC and others.

DBpedia Mobile is a client application that makes use of DBpedia content (an effort to extract Linked Data from Wikipedia) to offer a map-based interface for exploring the surroundings of a user [142]. The user can follow links to other related content and can contribute to the Linked Data content by publishing photos or reviews of nearby points of interest.

Augmented reality-based mobile applications that integrate Linked Data principles have started to emerge more clearly in 2010.

In [139], the authors describe an envisioned future of mobile augmented reality and discuss the current limitations in the landscape of mobile augmented reality applications that hinder this desirable future. Linked Data principles, with a focus on the Linked Open Data cloud, are proposed as an appropriate mechanism to override these limitations. Some concerns regarding the implementation of Linked Data, such as the perceived complexity of RDF and SPARQL technologies, along with trust issues, are presented.

In 2011, the authors highlight in a position paper [143] the personalization aspect of Point of Interest recommendations that can be achieved through the use of Linked Data principles in a heritage-based augmented reality application. The authors argue that Points of Interest described as Linked Data can benefit from enhanced associated metadata, through the processing of resources that are linked to the Point of Interest, thus enabling recommender systems to personalize and contextualize better. Also, Linked Data-based Points of Interest enable a browsing-like experience for the user, helping her to easily find more information.

The SmartReality project, a nationally funded project in Austria, started at the end of 2010 with the purpose of investigating the combination of augmented reality with Semantic technologies and Web services for a smarter information delivery [144]. The authors propose a general workflow for the envisioned SmartReality platform. According to this workflow, the platform is responsible for filtering, ranking and formatting Semantic Web retrieved content, based on the TOIs (Things of Interest) recognized by the augmented reality mobile client. The TOI is a modified version of the more common POI (Point of Interest) and features only a name, an identifier (unique URI) and a category (also a unique URI). The project aims to reuse current standards for augmented reality and the other technologies involved.

A demonstration of the SmartReality concept is showcased in [145], where an augmented reality client enhances common posters found on the streets with music-related content (e.g. information about the artist, booking a ticket at the concert). An annotation tool is implemented for the poster, to help content creators to link posters or parts of the posters with Semantic-based content or services. A more detailed workflow of the SmartReality platform is presented, along with tools and processes used for retrieving and processing the information. The proof of concept is limited to augmenting street posters with content from just one Linked Data source (play.fm).

In [37], the researchers take a novel approach by arguing for and implementing a model that replicates, aggregates and consolidates Linked Data graphs directly on the mobile device, thus eliminating the need for a separate processing server and an always-on Internet connection. The sensor-based augmented reality mobile application is deployed in a mountain area and uses data sources such as LinkedGeoData, DBpedia and Geonames for exploring the surroundings. The approach is currently limited to a reduced number of possible replicated RDF triples and to a sensor-based tracking, which is sufficient only in some scenarios.

ARCAMA-3D is an augmented reality location-based mobile application that facilitates surroundings discovery by overlaying 3D models of buildings on the real world as seen by the urban user [146]. The 3D models are interconnected with the Linked Open Data cloud, thus extending the information offered to the user but also extending the Linked Open Data cloud itself. The authors propose an ontology called arcama-owl to describe, in space and time, the OiIs (Objects of Interest). In a follow-up to the project [147], a web application is implemented that allows for uploading 3D models and linking them to the Linked Open Data cloud. Also, a mediator ontology is employed for linking information such as the roles of buildings with similar information from DBpedia.

The authors describe in [148] a framework and an implementation, the WantEat application, which provides an augmented reality browsing experience for accessing ontological described knowledge mainly in the field of gastronomic tourism.

Several augmented reality projects have been developed also in the field of cultural tourism with the help of Linked Data principles.

The application described in [149] enables the user to search and browse cultural heritage information in Amsterdam with a location-aware mobile device which displays an "enriched local map" of Points of Interest. The authors present a detailed approach for integrating the various Linked Data sources (from the Linked Open Data cloud and from specialized knowledge sources in the cultural heritage domain of the city), while highlighting some common challenges, such as harvesting, merging and aligning the information. It is concluded that Linked Open Data sources deliver limited information if processed in isolation, but integrating them, together with some other more specialized repositories, can yield a very informative location-based service.

In [150], the authors implement an augmented reality application for exploring cultural heritage sites such as popular cemeteries. They use content sources from scraped websites and DBpedia.

Integrating Linked Open Data in location-based mobile applications for touristic purposes (albeit not augmented reality enabled) is pursued also in [151] and [152]. The latter describes Telemaco, a client-server system in which the server integrates various Linked Open Data sources of content and information from social networks to offer personalized recommendations to users.

Some projects highlight that further research should focus more on the semantics of the Linked Data, i.e., the augmented reality application should also be able to suggest touristic paths, not only to allow browsing links from the Points of Interest.

Previous projects show successful attempts at integrating Linked Data principles in mobile augmented reality applications, while highlighting common challenges in pursuing this approach. Further research should focus on overcoming such challenges on a wider scale (significant number of Linked Open Data sources integrated) and in more general use cases.

4.2.4. Advantages of Linked Data Integration

Linked Data principles are well-suited for organizing content for mobile augmented reality applications. In [45], the authors present four characteristics of the Web of Data: data is separated from format and presentation; data is self-describing (vocabularies that describe the data can be found via URI dereferencing); data uses a standardized access mechanism (HTTP) and a standardized data model (RDF); and it is open (new data sources are constantly added and can be dynamically integrated).

In his popular W3C design note on Linked Data [48], Tim Berners-Lee issues four simple rules for publishing data: use URIs as names for things; use HTTP URIs so people can look up the names; provide useful information using the standards (RDF, SPARQL) when people look up a URI; include links to other URIs, to enable discovery of more things.

These simple yet effective rules have enabled the constant growth of the so-called Linked Open Data cloud: a significant number of datasets, published by various organizations or individuals, in various domains, with an open license, linked between them. The Linked Open Data cloud has evolved from 12 datasets in May 2007 to 570 datasets in August 2014 [46]. Some datasets have become hubs in the Linked Open Data cloud because they store information for very common concepts (e.g. DBpedia) or location-based information (e.g. Geonames).

The authors identify in [139] three ways through which Linked Data in general and the Linked Open Data cloud in particular can enhance augmented reality: through dynamic selection and integration of data sources, through enabling the utilization of a wide range of contextual data and by offering the user a Web-like browsing experience in augmented reality applications.

Previous projects have successfully demonstrated the integration of various datasets from the Linked Open Data cloud in an augmented reality application [37], [146], [149]. This integration is possible due to common encoding and standards (RDF, RDFS, OWL and SPARQL).

Moreover, easy discovery of new datasets to be integrated is enabled through various means, such as semantic search engines (e.g. Sindice), follow-your-nose principle and consulting catalogs for dataset metadata (e.g. CKAN) [153].

Previous projects have also demonstrated the possibility for the user to find additional information by following the links that are attached to the Points of Interest. Due to the way that Linked Data is working, the augmented reality application is able to show bits of additional data inside the view of the application, without requiring the user to leave the application and to open a new application (usually the mobile browser).

The following scenario reveals the advantages of integrating Linked Open Data in mobile augmented reality applications for the end user:

Maria is for the first time in Palermo, Italy, and she is interested in local cultural touristic attractions. Luckily, her augmented reality-enabled glasses can help her to explore the surroundings in every way she desires. She starts her journey in front of the cathedral. While admiring the outstanding building, she scrolls through the history of the cathedral as documented on Wikipedia. She would like to visit it, but unfortunately, it is closed. She easily checks the timetable, as it *was published on the open data portal of the City Hall, and finds out that she can visit it the next day in the morning. Maria is curios which other buildings nearby have so many architectural styles as the cathedral (Norman, Gothic, Baroque and Neoclassical). The augmented reality application suggests some nearby buildings that fit the criteria. Maria chooses one of them and the application highlights on her surroundings the shortest route to get there, using information from OpenStreetMap. As she walks towards this destination, she is able to see photos of how the streets of Palermo looked like in the past, superimposed on the actual view of the city.*

Not only augmented reality applications benefit from the Linked Open Data cloud, but the vice versa is also true. There is an increasing amount of content created for augmented reality applications. If this content is linked into the Linked Open Data cloud, then the Linked Open Data cloud grows in size and diversity and this benefits Linked Data application developers.

4.2.5. Some Concerns and Discussion Points

While applying Linked Data principles to augmented reality can yield great benefits, there are some issues that need to be taken into consideration.

One of the biggest advantages of the Linked Open Data cloud, its significant size, might easily become a disadvantage, if content is used in a mobile application without specific filtering. This is because a query in the Linked Open Data cloud might yield a great number of Points of Interest around the user, which would overwhelm an experience typically provided on a small screen device. Filtering which takes into account the context (e.g. what the user is doing, what preferences she has set in her social network) is certainly necessary.

Many issues regard the inherent nature of Linked Open Data. Datasets are heterogeneous in terms of vocabularies used and have overlapping information. Common Linked Data publishing workflows include vocabulary mapping, interlinking datasets and cleansing data for an integrated dataset to be obtained. Entity resolution issues (deciding if two entities from different datasets are referring to the same thing or not) are worsened by ambiguity, which can be of two types: name ambiguities (due to typos, different languages and homonyms used) and structural ambiguities (inconsistent relationships to other entities). These can be resolved using ontology matching techniques [154]. Content itself can differ between datasets that have overlapping information, and integrating them involves taking a decision whether to use one source or the other.

Not only integrating various datasets can raise issues, but also the datasets in isolation, because "authors that work with user generated Linked Open Data have to deal with duplication, misclassification, mismatching and data enrichment issues" [155].

In [149], the researchers report on having to deal with multiple Semantic Web challenges while integrating sources from the Linked Open Data cloud for the cultural heritage augmented reality browser: different schemas, different labeling conventions, different geodata, errors in geodata and conflicts in typing. For the schema matching, the authors had to create around 200 mapping rules by hand. The paper reports that it typically found discrepancies of 20m, in some cases even hundreds of meters, in coordinates of the same location, as reported by various sources in the Linked Open Data cloud. Inaccuracy of the geodata required a range of at least 35 meters in order to find all possible candidates for a Point of Interest.

These results indicate that a more sophisticated algorithm for integration of spatial information is required.

In [37], the researchers also report on differences in content retrieved from similar-domain Linked Open Data sources, which required aggregation and consolidation of the data.

Other Semantic Web specific issues are trust, provenance, quality, relevance, privacy and licensing. For dealing with the provenance aspect, developers can use the PROV ontology proposed by the W3C [156].

Another significant discussion point is the architectural pattern for integrating Linked Data in the augmented reality application. Three architectural patterns are presented in [51, pp. 97–98]: the Crawling Pattern (sources are crawled and cached in a single datastore), the On-The-Fly Dereferencing Pattern (URIs are dereferenced on the spot) and the Query Federation Pattern (a query is issued to a fixed known set of sources).

In augmented reality situations, where registration with the real world needs to happen in real-time, the first pattern should be the preferred one, although it has the disadvantage that the information is not always up-to-date. Depending on the use case, this might be a problem or not. In [149], the authors report on retrieving the RDF statements on-the-spot, although in some cases it takes even 50 seconds. They believe this might not be a problem if the content is preloaded as the application tries to guess in advance which Points of Interest the user is approaching.

4.3. Relevant Aspects for Linked Data Integration

The most important type of data to be integrated in mobile augmented reality tourism applications is geodata, which is currently the backbone of the Linked Open Data cloud [46]. Also, due to the nature of open data in general, of considerable importance is the quality of the data, along with provenance, both of which trigger trust. These key concepts are depicted in Figure 4-1.

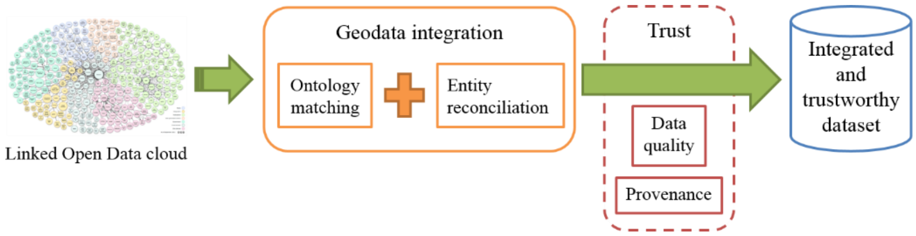


Figure 4-1. Highlighting of relevant aspects for Linked Data integration in mobile augmented reality applications for tourism

4.3.1. Integrating Geodata

Geographic information is at the center of the biggest category of information on the Web nowadays. A popular quote says that "80% of all information is geographically referenced". Due to their inherent nature, augmented reality applications treat geographic information aspects as first class citizens of the Web.

There is a trend to integrate more and more the fields of Geographic Information Systems and Augmented Reality, as this helps to reduce the gap between the real world and the virtual world by means of projecting the virtual world in space and time [157].

There are many popular sources of geographic information in the Linked Open Data cloud, such as Geonames, DBpedia and LinkedGeoData. A developer of mobile augmented reality applications, wishing to integrate several of these resources so they can complement each other, is facing a challenge due to the sources using different vocabularies to describe the data and due to duplicate records that appear after the data integration.

The first issue, of aligning different vocabularies, is known in literature as ontology matching and applies to all fields, not only to geographic information [154].

In [158], the authors enumerate 15 translation/mapping patterns that are useful in aligning ontologies. Some tools can assist in semi-automatically matching ontologies, such as the COMA++ tool [159]. Integrating multiple distributed ontologies is likely to yield inconsistencies. One example of a tool that can aid with this type of issues is RaDON [160].

The second issue is called in literature entity reconciliation and aims to identify and merge data that refers to the same real world feature.

The problem of integrating several location-based services has been studied in the literature before the proliferation of the Linked Open Data. Researchers present in [161] and [162] uncertainty issues in fusing information from various location-based services and categorize integration issues as: geographic integration (differences in geographic representation – points/line/polyline/volume – and in the reported position), place name integration (differences in spelling of location names) and semantic integration (differences in describing the data/metadata).

The author proposes in [163] a supervised machine learning approach for duplicate detection and data consolidation over gazetteer records. Features that help link the records between them are categorized by: place name similarity, geospatial footprint similarity, place type similarity, semantic relationships similarity and temporal footprint similarity.

The expansion of the Linked Open Data cloud led to early work on aligning geographic datasets, such as on interconnecting LinkedGeoData and Geonames based on type information, spatial distance and name similarity, which allow single access to an integrated dataset and validating the information, one against the other [164].

A more complex solution for aligning geographic datasets is proposed in [165]–[167]. It is based on constructing restriction classes using owl:sameAs links. It is well suited also for enhancing a poorly described ontology with the help of a richly described one.

In [168], the researchers publish two geospatial datasets, GADM (Global Administrative Areas) and NUTS (Nomenclature of territorial units for statistics), using a proposed NeoGeo vocabulary and integrate them with other datasets from the Linked Open Data cloud. They describe an algorithm for finding equivalent geometric shapes across multiple datasets.

The creation of such vocabularies as NeoGeo was determined by the lack of a standardized RDF vocabulary for managing geographic information in the Linked Data space. In the meantime, a W3C working group standardized GeoSPARQL, which is comprised of a small ontology for representing features and geometries and SPARQL query predicates and functions [169].

However, the implementation of GeoSPARQL in current triple stores is still in its infancy. A recent benchmark [170] on geospatial capabilities of current triple stores reveals the need for huge optimization in several aspects, such as performance of spatial indexing, query optimization and GeoSPARQL compliance.

A survey of the current geodata providers in the Linked Open Data cloud and their data modelling approach is presented in [171]. Furthermore, the authors propose an alignment process using the GeOnto vocabulary, focusing on interlinking French geodata with the Linked Open Data cloud.

The researchers describe in [172] the process of integrating several Linked Open Data and non-Linked Open Data sources, with the potential of using the integrated dataset in an augmented reality touristic application.

Further research is needed to improve precision and performance for successful integration of geographic datasets on a large scale. In the case of a mobile augmented reality tourism application, the data needs to be integrated and cached in a datastore. The time cost for integrating the data in real-time would be too high for augmented reality applications, which require registration with the real world with almost no delays for a proper experience. Even with this approach, the development of such an application is still hindered by the poor implementation of the GeoSPARQL standard in current triple stores.

4.3.2. Data Quality

Exploitation of Linked Data sources should take into account the quality of data. This is even more important in such areas as mobile augmented reality applications for tourism, as the tourist relies on the data for real-time exploration of the surrounding environment.

Linked Data assumes an open-world philosophy according to which anyone can say anything about anything [173]. This leads to issues such as inconsistencies in data provided for the same entity by different sources, concerns of timeliness, completeness and accuracy, just to name a few. A comprehensive list of quality criteria for Linked Data sources, grouped by content, representation, usage and system, is proposed in [174].

Corroborated with limitations induced by the shallow expressivity of the published knowledge and the heterogeneity of the describing schemas, Linked Data runs the risk to become "merely more data" [175], in the absence of efforts to overcome these challenges.

Various approaches have been identified to tackle these issues. For example, in [176], [177], the authors describe a framework to identify data quality problems, such as missing literal values, false literal values and functional dependency violations, using generic SPARQL queries. Also, researchers have proposed vocabularies for data quality management, such as in [178].

Attempts to quantify the data quality of user-generated content have been pursued in the literature. Of interest to mobile tourism applications is, for example, the comparison between proprietary geodata and Volunteered Geographic Information (a term coined by [179]), such as OpenStreetMap (with its Linked Data equivalent LinkedGeoData). Several papers [180], [181] have reported on the high reliability of Volunteered Geographical Information in urban centers, a finding that is confirmed by the fact that the quality of data increases with the number of contributors [182].

Data quality cannot be always assessed in absolute terms. A pragmatic approach is to evaluate the data quality based on the *fitness for use* principle, which takes into account the specific task that is to be achieved [183]. This principle should be further investigated for mobile augmented reality applications for tourism.

4.3.3. Provenance and Trust

Along with general data quality factors, provenance is one of the main triggers for trust in Web content [184]. Although some tools for tracking provenance are hardwired in Linked Data (such as dereferenceable HTTP URIs), the ever increasing integration of various heterogeneous datasets makes it harder to keep track of detailed provenance metadata. This is the case also for mobile augmented reality applications in the field of tourism.

Usually, provenance refers to keeping track of the workflow that led to the creation of the data. In today's Web of Data, it is important to also keep track of data access information, meaning metadata about the providers of the data and the way they deliver it [185].

The Open Provenance Model [186], a vocabulary that describes provenance using the terms "artifact", "process" and "agent", is used in applications such as El Viajero, a platform for managing Linked Data in the travelling domain, which integrates several heterogeneous datasets [187].

The W3C Provenance Working Group recently proposed the PROV standard for dealing with provenance information. In [156], the authors discuss the PROV standard and propose an implementation of Tim Berners-Lee's "Oh, yeah?" button [188]. Also, the application of the PROV standard in modelling uncertain provenance is discussed in [189].

4.4. Conclusion

I have shown in this chapter that mobile augmented reality is a promising set of technologies for the tourism field. However, current approaches in content delivery for mobile augmented reality applications have some limitations due to content being stored in isolated silos of information.

More exactly, content has always been obtained from fairly simple databases which act as isolated silos of information. This is typically found in augmented reality-based touristic guides, where a user, to find more information about a Point of Interest, has to click a link that opens another application (usually the mobile browser). The content is usually handpicked and very specific for the kind of application that it is made for and cannot be reused.

Linked Data principles and technologies can aid in this respect. First of all, integration of data from several sources is possible due to the use of common encoding and standards, such as RDF and SPARQL. Secondly, the Web of Data offers a large variety of contextual data, due to the interlinking of datasets from multiple domains. As such, the tourist can jump from information to information in search of what is interesting for her. This can happen without leaving the augmented reality application.

However, in the process of integrating such data, the researchers need to address some concerns that are related either to the inherent nature of Linked (Open) Data or directly to the integration of Linked Data in augmented reality applications.

Because of its open nature, Linked Data suffers from aspects such as wrong information, duplicate information or incomplete information. Several sources might report different things about the same subject. Additionally, data is heterogeneous in terms of vocabularies used, a fact which hinders alignment and integration. The projects described in this chapter confirm these findings.

To better understand what this exploitation means, I explored several aspects that are relevant, namely integration of geodata, data quality, provenance and trust.

Integration of geodata seems to be the most important one, as most of the information nowadays is geotagged. Such integration consists, first of all, in ontology matching, to align different vocabularies, and secondly, in entity reconciliation, to identify and merge data that refers to the same geographic feature. Currently, geodata integration is hard to achieve in a timely manner on a large scale and existing triple stores lack proper geodata support.

Data quality is important for a successful augmented reality touristic experience but this aspect has not been tackled yet in the literature. Provenance is also significant, and both contribute to the overall trust of the tourist in the data and the application.

In conclusion, I have presented in this chapter the current limitations of mobile augmented reality applications regarding the content used, along with the Linked Data principles, mainly openness and standardization, which can help overcome the current issues. I highlighted some discussion points regarding the interconnection of these two fields and I pointed out some concerns, like trust, quality and integration, which should be addressed in this endeavor. Also, I have reviewed several projects that seem to confirm these concerns.

These analyses and findings constitute a solid theoretical contribution to the present thesis. Some results presented in this chapter have been published in [190] and [104].

5. THE PROPOSED MODEL FOR THE INTEGRATION AND THE IMPLEMENTED PROTOTYPE

5.1. Introduction

As already discussed in previous chapters, augmented reality is an intuitive interface for people doing specific activities in multiple fields, one of them being the tourism domain [33], which is also the focus of my research. Tourism is a significant industry, even in times of crisis [191], that benefits lately more and more from the boom in gadgets and technology. Mobile augmented reality is a suitable approach for delivering a better touristic experience, both of them being based on visual discovery. This benefic combination has been speculated in many mobile augmented reality applications that have been developed with tourism cases in mind [127], already presented in chapter 2. However, these applications are rather static, due to the nature of the content that they process and store. They either use closed databases of information (that is gathered and processed for that application only) or open (but single and disconnected) databases called channels (such as Wikipedia). This is perceived as a limitation by the tourists [1], who are more and more accustomed to dynamic, context-sensitive and adaptive information which they also expect to see in such applications.

Integrating linked open sources of information has been proposed as a suitable solution to overcome such limitations of current mobile augmented reality applications. Linked data, being based on standards for representing and accessing data that are flexible enough to allow adding and removing sources of data, navigating from one source to another and processing it according to its meaning, can break the barriers of closed and disconnected information in augmented reality applications. However, this integration is not straightforward and developers have to deal with known linked data integration issues and issues due to its open world assumption, as presented in chapter 4, such as geodata integration, data quality assessment, provenance and trust issues.

A benefit can be also seen the other way around. There is an intense research in finding suitable ways for visualizing and understanding the increasing amount of linked open data published on the web. One of the targets for this endeavor are lay users, which, as opposed to more experienced ones, need simpler, more intuitive visualization-based browsers for linked open data [192]. It is non-trivial to be able to show intuitive, easy to comprehend and action-ready linked open data and, at the same time, to be able to show large quantities of data from various fields.

While integration issues have been studied in projects such as those presented in chapter 4, there is still a lack of research in terms of a model for developing mobile augmented reality applications that integrate multiple sources of linked open data information, with a focus on combining open user-generated information with open governmental data. There is also missing an analysis of how the quality and quantity of available information changes when switching from isolated silos of information to open and interlinked sources of content.

The objectives of this chapter are to propose a model and a set of guidelines for implementing a mobile augmented reality touristic application which integrates linked open data, and to develop a prototype of such an application. It also aims to analyze the usefulness of integrating multiple sources of linked data information, to highlight obstacles and problems that show up in processing datasets "in the wild" and to propose solutions and ignite discussions based on the findings.

The model I proposed and the prototype I implemented show that there are clear steps in integrating linked open data in mobile augmented reality applications and that this process yields benefits for the tourist in terms of information access.

5.2. Major components and steps

The model of developing an application that exploits linked data in augmented reality mobile applications resembles closely the architecture of developing a generic linked data application, as described in [51], with some adjustments specific to augmented reality applications. I consider open government data to be a strong pillar in linked open data for augmented reality applications, but, as most of the government data is still only published as raw data, I need to also apply to my model the general Linked Open Government Data lifecycle, as described in [105].

As such, to develop an application that fulfils the objectives set in the introduction, a developer should be aware of the steps and components described in Figure 5-1 (updated version of the workflow that I published in [193]).

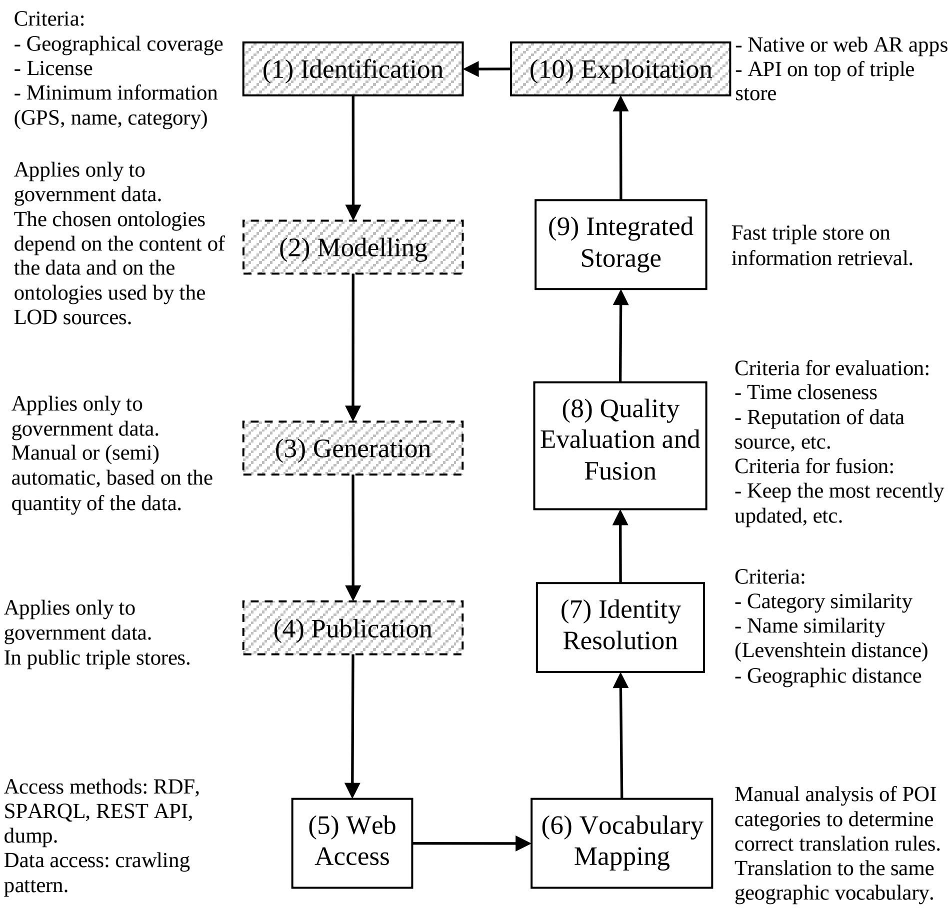


Figure 5-1. Major steps of the application model

The prototype I developed integrates data from user-generated information sources, like Wikipedia and OpenStreetMap, and from governmental ones, like the Romanian Government Open Data Portal[23]. The former two sources are already published as linked data under the name of DBpedia[24] and, respectively, LinkedGeoData[25]. The governmental source was processed by me to become linked data. I integrate these sources using a linked data integration framework and store the consolidated dataset in a triple store. The geographic area I target is the whole country of Romania, with a focus on the city of Timisoara. I built an augmented

[23] http://data.gov.ro/
[24] http://dbpedia.org/
[25] http://linkedgeodata.org/

reality mobile application that works directly in the browser of the mobile device and which queries the triple store from an Application Programming Interface (API) and shows the information to the user in a typical augmented reality experience.

The steps in the process are detailed as follows.

5.2.1. Identification

The first and foremost important step is identifying the most appropriate datasets to be used for the application's purpose. The selection is based on the quantity of information in the dataset (does it cover the geographic area that I want the augmented reality experience to take place in and does it have a significant number of POIs?), the quality of the information (how complete and accurate is the description of the POIs?) and the attached license (am I allowed to use the information?).

A minimum set of descriptors are required for the POIs to be used in an augmented reality application: label, GPS coordinates and category. While label and GPS coordinates are certainly required to be able to show the POI in space and say what it is, the requirement considering the presence of a category is debatable. I believe that some form of categorizing the POI is strongly needed for proper filtering of the displayed information. Otherwise, there is the danger of showing a cluttered (and, therefore, unusable) display to the user. Other descriptors, not required, but useful to have in an augmented reality environment, are: description (for further information), email/telephone numbers (for the user to be able to email/call the contact person of the POI immediately from her smartphone), opening hours/closing hours (to be able to filter the POIs around her based on the availability at that certain time), picture of the POI (to be able to recognize it easier), links to similar POIs (for the user to be able to further navigate to the other POIs based on similar architecture, functionality etc).

In this thesis, I am focusing on open datasets (the ones that are compatible with the Open Definition [4]), from governmental sources and from user-generated ones. To check if a license is truly open, developers can use the list of licenses that are conformant [194] with the specifications written in the Open Definition.

Big linked open datasets can be found easily by using the LOD diagram [46]. The diagram shows what major type each dataset is and the most connected datasets in the world. This visual search is quick and intuitive and works well for identifying big, well-known datasets. Smaller datasets are harder to find this way. One can use tools like DataHub[26] or PublicData[27]. A comprehensive worldwide list of open data catalogs can be found on DataCatalogs[28]. The most recent crawling of the Linked Open Data space [53] shows that geographic datasets occupy 2% of it. Other domains with potential data on POIs are the governmental one, with 18% of the LOD space (the category with, by far, the biggest growth since the last report from 2011) and cross-domain, 4% of the LOD space.

For our prototype, I aimed to identify datasets to be used by tourists visiting Romania. The testing phase would be done in the city of Timisoara, which is currently a candidate for the title of European Capital of Culture in 2021, so it is a place with high touristic potential. Using the above criteria, I identified DBpedia and LinkedGeoData as being suitable candidates. DBpedia covers around 4000 POIs in

[26] http://datahub.io/

[27] http://publicdata.eu/

[28] http://datacatalogs.org/

Romania and LinkedGeoData around 20,000 POIs (the numbers are approximate due to the fact that, for straightforwardness, the crawling of the datasets was done for a circular area centered on the geographical midpoint of the country, so small parts of the neighbouring countries were also included). These datasets have the required descriptors and a few other optional ones as well. Their licenses, Creative Commons Attribution-ShareAlike version 3.0 (CC BY-SA 3.0) and GNU Free Documentation License (GNU FDL) in case of DBpedia and Open Data Commons Open Database License (ODbL) in case of LinkedGeoData, are compatible with the Open Definition.

I could not find appropriate linked open governmental data sources for Romania, so I selected a raw government dataset that was published recently on the Romanian Government Open Data Portal and that lists all the official museums in Romania. The 951 POIs are well described and very appropriate for being used in an augmented reality environment. The dataset[29] was published by the National Institute of Heritage who at the end of 2014 won the Open Government Partnership prize from the Romanian Government for transparency in administration. The portal's license, the Open Government License[30] (OGL ROU 1.0), is compatible with the Open Definition. I believe that the extra work for transforming this raw dataset into a linked dataset is worthwhile. Various valuable raw government data are published currently in many places in the world and the process of transforming it to linked data should be taken into account as a default step by developers.

Table 5-1. Comparison between various aspects of the integrated sources of data

Name	License	No. of POIs for Romania	Ontology
DBpedia	CC BY-SA 3.0 and GNU FDL	~ 4000	dbo
LinkedGeoData	ODbL	~ 20,000	lgdo
Romanian Open Data Portal	OGL ROU 1.0	951	---

Comparing the identified sources (Table 5-1 and Table 5-2) reveals a great variation in licenses, ontologies or form of access, which makes the integration a non-trivial task.

5.2.2. Modelling

The modelling stage concerns directly the governmental data source. This step consists of analysing the information and deciding on an ontology to be used in modelling that information. If the other sources of information to be mixed are already identified, it is helpful to use for the governmental source the same ontology as the one which is the most relevant in the other sources. However, in many cases, each source of information has its own ontology, so a suitable approach is to use the ontology of the most important of them.

29 http://data.gov.ro/dataset/ghidul-muzeelor-din-romania

30 http://data.gov.ro/base/images/logoinst/OGL-ROU-1.0.pdf

In our use case, the user-generated data sources use quite different approaches in modelling. LinkedGeoData uses a lightweight OWL ontology, called the LinkedGeoData Ontology (lgdo), which maps pretty directly on the key-value pairs, known as tags, that are specific to POIs in OpenStreetMap [59]. DBpedia uses a shallow, multi-domain ontology, called the DBpedia Ontology (dbo), which maps Wikipedia infoboxes through hand-made rules [54].

For the government data source, I considered a combination of vocabularies (such as FOAF or Basic Geo) but also a small part of the DBpedia ontology, for easier alignment.

5.2.3. Generation

The generation step involves transforming the raw data into linked data and it applies to the government data source. For this purpose, I chose the OpenRefine tool, as it has some very useful features for dealing with data: exploring, cleaning, transforming, reconciling and matching [195]. Because the effort was for generating linked data for just one dataset, I chose a manual approach. In case of a significant number of datasets, a (semi-)automatic approach is more suitable, such as the one employed by the csv2rdf4lod tool used for the TWC LOGD portal [109].

5.2.4. Publication

This step refers to the publication of the generated linked dataset. This is not required if the linked data is only to be used in the application. However, the data might be useful to other developers as well, so it is a nice-to-have feature. I will publish the linked government dataset after switching from the development triple store to the production triple store.

5.2.5. Integration (Access + Integration + Storage)

For the integration of the datasets, I have chosen the popular Linked Data Integration Framework (LDIF) [196]. This framework covers four modules from the generic schema of linked data applications: the Web Data Access Module, the Vocabulary Mapping Module, the Identity Resolution Module and the Quality Evaluation and Fusion Module. The tool is very useful in such cases as it streamlines the operations of accessing the data, aligning it, merging it and checking its quality. LDIF has configuration files where one can modify parameters to change its runtime behaviour. A similar tool that developers might use is Apache Marmotta[31].

Integration – Access. The data sources can be accessed in various ways (see Table 5-2).

[31] http://marmotta.apache.org/

Table 5-2. Comparison between accessing methods available for the integrated sources

Name / Accessing method	RDF	SPARQL	REST API	Dump
DBpedia	x	x		x
LinkedGeoData	x	x	x	x
Romanian Open Data Portal			x	x

Linked data applications present three types of data access patterns [51, pp. 97–98]: the crawling pattern, in which several sources are crawled, the data is fetched and stored in a single datastore, from where it is accessed later; the on-the-fly dereferencing pattern, according to which the data sources are crawled by dereferencing the URIs on the spot and as such going from one to the other in real time; and the query federation pattern, in which a query is sent to a fixed set of data sources that are known beforehand.

The latter two patterns provide the most up to date data; however, they might not provide the data fast enough. This can be an issue in augmented reality environments, where the availability of the data for the registration to happen in real time is a critical aspect of a successful experience. The first pattern might have the disadvantage of not providing the latest data (though a frequent refresh of the data can make this issue a minor one) but the resulting data is available immediately.

In my use case, I favoured the crawling pattern, which is also the one used by the LDIF platform. I specified in the LDIF configuration file how each data source should be fetched, depending on the available data access methods presented in Table 2. DBpedia data was downloaded both through SPARQL queries requesting categories specific to Romanian POIs and through data dumps, while LinkedGeoData and the governmental source were inputted only as data dumps. For the ease of processing, I crawled those POIs from DBpedia and LinkedGeoData that were in a certain radius around the center of the country, meaning that I also got some POIs outside the country (as it does not have a circular geographic area).

Integration – Vocabulary Mapping. LDIF uses internally the R2R framework [94] to achieve the translation of the ontology in the original source to the desired resulting ontology. Looking into the data of the selected sources, we see various categories that the POIs belong to. The dbo ontology uses straightforward categories, such as *SportFacility* or *EducationalInstitution*. However, DBpedia also uses the yago ontology, which has categories that are hard to translate automatically (such as *PlacesOfWorshipInTimisoara* or *SynagoguesInRomania*) due to the way they are built (extending Wikipedia categories were extended with WordNet [197]), which are hard to manually process. Because of this issue, the categories that should be translated must be picked up manually from the data with the desired purpose in mind. LinkedGeoData has some more straightforward categories, such as *FuelStation*, *Parking*, *Supermarket,* etc.

Filtering by category is important for the user, but displaying all these niche categories on a small screen might puzzle the user. Moreover, displaying such niche categories would mean that the user might not find anything close to her in several categories and would be disappointed by the augmented reality experience.

As such, I empirically devised 7 supercategories that would encompass all the smaller categories: Tourist (churches, museums), Utility (fuel station, exchange offices), Food&Drinks (restaurants, pubs), Shopping (supermarkets, shopping malls), Emergency (pharmacies, hospitals), Entertainment (sports, cinema) and Institution (townhall, university).

The translation rules were specified by indicating that the source's category is *rdfs:subClassOf* our own supercategory, which belongs to our ontology named tomldifo. Below are snippets of code showing examples of translation rules for DBpedia and LinkedGeoData categories.

```
yago:PlacesOfWorshipInTimişoara rdfs:subClassOf tomldifo:tourist .
lgdo:Monument rdfs:subClassOf tomldifo:tourist .
…
dbo:ShoppingMall rdfs:subClassOf tomldifo:shopping .
lgdo:Shop rdfs:subClassOf tomldifo:shopping .
…
dbo:University rdfs:subClassOf tomldifo:institution .
lgdo:University rdfs:subClassOf tomldifo:institution .
```

The museums in the governmental data source were translated to the tourist supercategory.

Generally, the translation step should also involve aligning all the sources to the same geographic vocabulary. It was not the case for our prototype, as all the sources are using the WGS84 system. However, I had to create the *geo:geometry* attribute, from the latitude and longitude attributes, for the governmental source (this step was not done in OpenRefine) and for data that was missing this attribute in the other sources. This attribute is needed internally by Silk in the next module, Identity Resolution. Below is a snippet of code showing the creation of the *geo:geometry* attribute.

```
1 mp:toGeometry
2 a r2r:Mapping ;
3 r2r:prefixDefinitions "geo:
<http://www.w3.org/2003/01/geo/wgs84_pos#> . type:
<http://www.openlinksw.com/schemas/virtrdf#> ." ;
4 r2r:sourcePattern "?SUBJ geo:lat ?lat . ?SUBJ geo:long ?long" ;
5 r2r:targetPattern "?SUBJ geo:geometry ?'geometry'^^type:Geometry" ;
6 r2r:transformation "?geometry = concat('POINT(',?long, ' ',
?lat,')')" .
```

Integration – Identity resolution. For the identity resolution, LDIF uses the Silk tool [91] to be able to match and merge identical POIs that appear in multiple sources. Uncertainty in fusing information might yield three main integration issues: geographic integration, place name integration and semantic integration [162]. Correspondingly, I used as criteria for merging: the geographic distance between the POIs, the similarity of the names of POIs (using the Levenshtein distance algorithm) and the supercategory that they belong to.

These were by no means the only criteria one could use for integration purposes. The GPS coordinates of the boundaries of a POI are a piece of very useful criteria, for example, but this kind of information is virtually impossible to find in real user-generated data.

A clear condition I set for two POIs to be considered for integration is that they belong to the same supercategory. For the geographic distance between the POIs and the Levenshtein distance on names, I used empirically determined thresholds. To determine these thresholds, I imported the information of the POIs in Google Fusion Tables [198] and displayed them on the map of Romania, then zoomed in on the area of Timisoara, which I know well. I colour coded the pinpoints of the POIs based on the source of the data, so I was able to spot quite rapidly if certain POIs are susceptible of being integrated based on the geographic distance (Figure 5-2). I then looked over their other information (category and name) to see how much they differ, in the real world of user-generated data, even if they are the same POIs actually. I favoured the more restrictive thresholds, as I believe it is better for the user to see two POIs that are basically the same then miss one totally because it was merged wrongly with another.

Figure 5-2. Screenshot of the Google Fusion map with overlaying POIs (Screenshot © 2015 Google - Map data © 2015 Google, used with permission)

Below is a short snippet of code, showing the rules for merging POIs belonging to the "institution" supercategory. The threshold for Levenshtein distance is set to 1 (meaning the name can differ by maximum 1 character) and for the geographic distance is set to 100 (meaning it can differ by maximum 100m). The merging is done by creating sameAs links, as in the code snippet below.

```
<Interlink id="institution">
  <LinkType>owl:sameAs</LinkType>

  <SourceDataset dataSource="SOURCE" var="a">
    <RestrictTo>?a rdf:type tomldifo:institution .</RestrictTo>
  </SourceDataset>

  <TargetDataset dataSource="TARGET" var="b">
```

```
    <RestrictTo>?b rdf:type tomldifo:institution .</RestrictTo>
  </TargetDataset>

  <LinkageRule>
    <Aggregate type="average">
      <Compare metric="levenshteinDistance" threshold="1"
required="true">
        <TransformInput function="lowerCase">
          <Input path="?a/rdfs:label" />
   </TransformInput>
   <TransformInput function="lowerCase">
     <Input path="?b/rdfs:label" />
   </TransformInput>
      </Compare>
  <Compare metric="wgs84" threshold="100" >
    <Input path="?a/geo:geometry" />
    <Input path="?b/geo:geometry" />
    <Param name="unit" value="m" />
  </Compare>
</Aggregate>
  </LinkageRule>

  <Filter />
</Interlink>
```

Integration – Quality Evaluation and Fusion. For evaluating the quality of the integrated data and resolving potential conflicts in data, LDIF uses the Sieve module [199]. This module allows, in the first phase, to assign quality indicators, based on various scoring functions, to the data. The scoring functions can be related to how recent the data was crawled or updated, the reputation of the data (where it comes from) or they can compute indicators based on some numeric properties of the data.

For assessing the quality of the data in our prototype, I used criteria such as time closeness and reputation. I favoured the information that was updated more recently and I favoured the English Wikipedia over the other localized versions of it.

In the second phase, the Sieve module performs the fusion of the data, based on rules set by the developer and which take into account the quality indicators. For our prototype, I chose to keep the GPS coordinates of the most recently updated POI from several POIs that were identified as being the same.

Integration – Storage. For storage purposes, I use an OpenRDF Sesame[32] triple store version 2.7.9 on a Linux 2.6.32 server. There are a total of 1237760 statements currently stored on the server, belonging to POIs in the areas of Trento, Italy (crawled for demo purposes [200]) and the whole Romania.

5.2.6. Exploitation

The exploitation step involves building some sort of application or mashup that exploits the integrated linked data. Our implemented prototype is a mobile augmented reality application that works in the browser of the smartphone or

[32] http://rdf4j.org/

tablet. Therefore, the tourist is not required to download a separate application for her purpose. This way, we come closer to the concept of ubiquitous augmented reality.

Because the capability of mobile phones to interact directly with SPARQL endpoints is still in its infancy, even more in case of using just code in the browser, I have built an API on top of the triple store that allows the web application to get the information in JSON format. The web application can make two kinds of requests on the API:

a) get the list of POIs (with name and GPS coordinates) based on the position of the tourist, the desired radius (which is actually an almost-square rectangular area, due to poor geospatial support in the triple store [170]) and the desired type of POIs (supercategory)

b) get the complete information about a certain POI, i.e. name, description, picture, original source and link to further information

The application works like this (see also Figure 5-3):

1) the tourist opens the web application in her mobile browser and can see pinpoints overlaid on real POIs around her. The pinpoints have different icons, depending on the category the POIs belong to, so it is easy for the tourist to understand at a glance what is around her. There is a menu available all the time on the top right corner of the screen, from where the user can choose to see all or just one of the supercategories and can adjust the radius for displaying the POIs.

2) the tourist taps on a pinpoint and a small window opens on the bottom of the screen, showing the name of the POI, the distance to it and a link to complete information.

3) if the tourist presses that link, a bigger window opens, covering almost all the screen, and displaying complete information on the POI (name, description, picture, source, link to webpage).

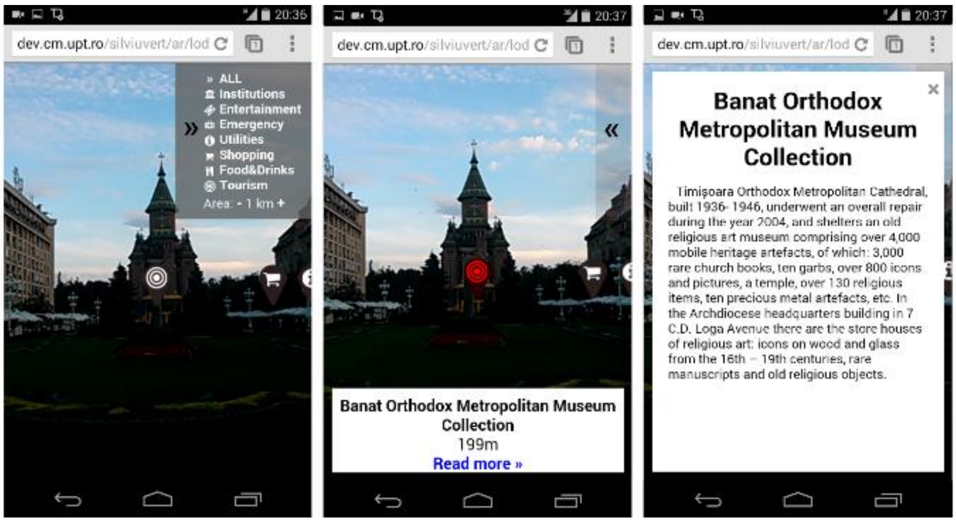


Figure 5-3. Screenshots of the web application

The web application is built on top of awe.js, an augmented reality web library[33] that uses technologies like HTML5, CSS and Javascript, combined with access to modern capabilities of mobile devices, such as geolocation, photo/video camera, WebRTC or WebGL, to show 2D or 3D augmented reality objects on top of the surroundings. The awe.js library works on top of three.js, a library that allows developers to design 3D experiences in the browser. According to the developers of awe.js, this library should work on the latest versions of Chrome and Firefox on Android, as well as on devices such as Leap Motion, Oculus Rift and Google Glasses. I successfully tested the application on various Android smartphones and tablets.

To be able to correctly draw the POIs on the screen, I needed to make several geometric calculations and transformations, as the awe.js library does not work out of the box with absolute geographic locations.

Projecting the awe.js camera (point of view) at the position of the tourist, as suggested in [201], yielded some undesired effects in the application. To avoid these issues, I kept the camera in the origin position (0, 0, 0) and updated the position of the POI icons with the difference between the absolute geographic position of the POIs and the absolute geographic position of the tourist.

I determined empirically that latitude corresponds to the z-axis, longitude corresponds to the x-axis and altitude corresponds to the y-axis, where the axes are represented as in Figure 5-4. I ignored the y-axis, because the altitude is very rarely present in the data.

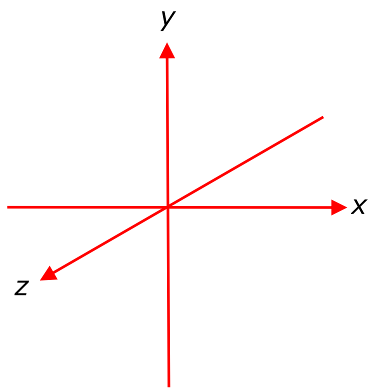


Figure 5-4 Three.js right-hand coordinate system

One of the issues that I encountered was the fact that the icons representing the POIs did not fully face the camera (the tourist's view). This happened because awe.js projected the icons by default to face north. Consequently, I had to rotate the icon on its y-axis with an angle that would "cancel" the difference between the tourist's direction of view and the north direction of the earth. Looking at the icon of the POI from above the earth, from

[33] https://github.com/buildar/awe.js

where it can be seen as a line in space, this would mean a rotation in 2D around the center of the line (as depicted in Figure 5-5).

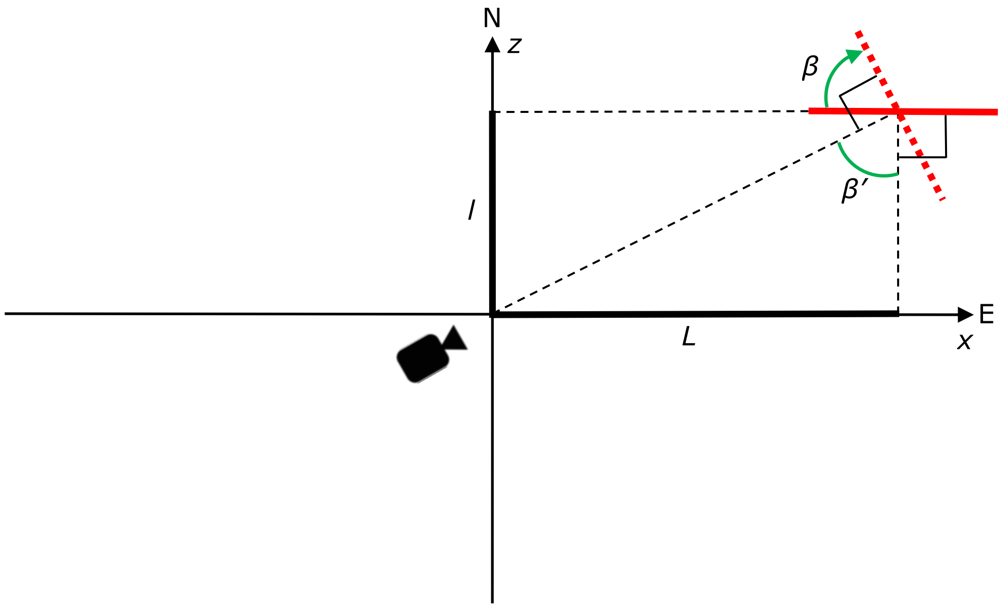


Figure 5-5 Calculation of the rotation angle for the POI icon in the NE quadrant ("Video Camera" icon by John Chapman from the Noun Project)

The icon of the POI (red solid segment) faces north and needs to be rotated in such a way (red dotted segment) that it completely faces the tourist (the camera of the awe.js library). As such, to correctly rotate the POI icon, we need to find the β angle.

The original position of the POI icon creates a right angle with the projection of the *l* segment and the correctly rotated POI icon creates a right angle with the hypotenuse of the right triangle created by the *L* segment and projection of the *l* segment.

Because "two angles are equal if their arms are pairwise perpendicular to each other" [202, p. 73], then the angles β and β' are equal.

To find the β' angle, we employ the following formula for calculating one of the interior angles of a right triangle:

$$\beta' = \arcsin\left(\frac{-L}{\sqrt{L*L+l*l}}\right)\times\frac{180}{\pi} \qquad (5.1)$$

where:

- L is negated because three.js, the 3D graphics library that awe.js is based on, is using a right-hand coordinate system

- The result in degrees is multiplied by 180/π to obtain the value of the angle in radians, due to the way the three.js library works

For the other quadrants, the correct rotation angle was empirically determined by adding or substracting radians from the above calculated values. A formula was hard to determine due to poor documentation of the awe.js library.

To calculate the distance between the user and a Point of Interest, I used Vincenty's Inverse Formula, as described in subchapter 2.6. The Javascript code that I used, written by Chris Veness [39], is shown below.

```
var L = λ2 - λ1;
var tanU1 = (1-f) * Math.tan(φ1), cosU1 = 1 / Math.sqrt((1 + tanU1*tanU1)), sinU1 = tanU1 * cosU1;
var tanU2 = (1-f) * Math.tan(φ2), cosU2 = 1 / Math.sqrt((1 + tanU2*tanU2)), sinU2 = tanU2 * cosU2;

var λ = L, λ′, iterationLimit = 100;
do {
    var sinλ = Math.sin(λ), cosλ = Math.cos(λ);
    var sinSqσ = (cosU2*sinλ) * (cosU2*sinλ) + (cosU1*sinU2-sinU1*cosU2*cosλ) * (cosU1*sinU2-sinU1*cosU2*cosλ);
    var sinσ = Math.sqrt(sinSqσ);
    if (sinσ==0) return 0;  // co-incident points
    var cosσ = sinU1*sinU2 + cosU1*cosU2*cosλ;
    var σ = Math.atan2(sinσ, cosσ);
    var sinα = cosU1 * cosU2 * sinλ / sinσ;
    var cosSqα = 1 - sinα*sinα;
    var cos2σM = cosσ - 2*sinU1*sinU2/cosSqα;
    if (isNaN(cos2σM)) cos2σM = 0;  // equatorial line: cosSqα=0 (§6)
    var C = f/16*cosSqα*(4+f*(4-3*cosSqα));
    λ′ = λ;
    λ = L + (1-C) * f * sinα * (σ + C*sinσ*(cos2σM+C*cosσ*(-1+2*cos2σM*cos2σM)));
} while (Math.abs(λ-λ′) > 1e-12 && --iterationLimit>0);
if (iterationLimit==0) throw new Error('Formula failed to converge');

var uSq = cosSqα * (a*a - b*b) / (b*b);
var A = 1 + uSq/16384*(4096+uSq*(-768+uSq*(320-175*uSq)));
var B = uSq/1024 * (256+uSq*(-128+uSq*(74-47*uSq)));
var Δσ = B*sinσ*(cos2σM+B/4*(cosσ*(-1+2*cos2σM*cos2σM)-
    B/6*cos2σM*(-3+4*sinσ*sinσ)*(-3+4*cos2σM*cos2σM)));

var s = b*A*(σ-Δσ);

var fwdAz = Math.atan2(cosU2*sinλ,  cosU1*sinU2-sinU1*cosU2*cosλ);
var revAz = Math.atan2(cosU1*sinλ, -sinU1*cosU2+cosU1*sinU2*cosλ);
```

5.3. Discussion and further work

The prototype that I implemented highlighted some well-known linked data integration issues, but also issues specific to our use case.

Sources differ greatly in terms of vocabulary used, form of access, license and so on, so proper identification and analysis is needed to prepare them for further processing. Tools and guidelines for this step are available and some have been presented here.

A significant manual effort is needed for translating the data. For example, some categories are titled in such a way (e.g. *TypeOfBuilding*In*City*) that it is impossible to know them beforehand. As such, developers have to look thoroughly into the data and pick up that information manually as best as they can.

Popular user-generated sources of information have a large quantity of information available for almost every geographic area. They are also great in linking datasets between them. However, the POIs that are not in the top of the interestingness ladder, which is most of them, are usually described shallowly. Governmental sources can be much more detailed – usually, if one element in the list is well-described, so are all the others. Their downside is that, with rare exceptions, the data is neither translated nor linked.

Geographic support is still poor in current triple stores, despite the fact that the GeoSPARQL standard has been launched in 2012. I could not request the POIs that are in a circular geographic area centered on the position of the tourist, due to the fact that the OpenRDF Sesame triple store does not implement this feature. I instead requested the POIs that are inside an almost-square rectangular area centered on the position of the tourist. This approach barely hinders the experience of the tourist; however, it is certainly more expensive in terms of the request's response time. It is advisable that the production server uses a triple store that implements SPARQL requests on a circular geographic area, such as Virtuoso[34].

The data is accessed in real-time, which is fine for those that have a data connection. This happens in the case of most citizens and tourists from the same country. Foreign tourists are less probable to have a 3G connection, so a solution must be searched in this case (e.g. downloading a consolidated dataset for one city at a time, before visiting it).

Names and geographic position vary greatly between POIs in different information sources that are actually the same, this being a classical issue of user-generated data. I observed that it is almost impossible to set thresholds for detection algorithms that keep both the number of missed unmerged similar POIs and the number of false positives low. For example, there are identical POIs that differ by hundreds of meters in geographic distance and by several words in the title (e.g. "Timisoara Zoological Garden" in DBpedia vs "Padurea Verde Zoological Garden" in the governmental dataset, separated by 213m, or "Stadion CFR" in LinkedGeoData vs "Stadionul CFR (Timisoara)" in DBpedia, reported as being 101m away one from the other). Setting such large thresholds, that would enable proper merging of identical POIs, would undoubtedly trigger false positives. Of course, I also encountered situations where identity resolution is easy to do, such as for the POIs "Museum of Banat" in DBpedia and "Museum of Banat" in the governmental dataset, with a reported distance between them of 25m. The algorithms are hampered by the fact that DBpedia usually has labels in English and LinkedGeoData in Romanian, for the geographical area of Romania.

[34] http://virtuoso.openlinksw.com/

These settings for integration can be improved by including additional datasets in the process, extending the targeted geographic area and analysing the output. The application can benefit from an automatic discovery and inclusion of governmental datasets, which is a field where research is taking place nowadays. Further analysis is useful in the quality evaluation step, as the information that guides a tourist in real-time in an unknown place is regarded as being crucial for a successful experience.

Regarding the augmented reality web application, I demonstrated the possibility of showing this integrated data as augmented reality directly in the browser, thanks to newer capabilities of mobile devices. However, the user interface can be improved by, for example, showing subcategories inside supercategories, for better filtering, or showing links to similar POIs (regarding architecture, functionality and so on) for a more complete experience.

5.4. Conclusion

In this chapter, I proposed a model for building an augmented reality application that integrates linked open data. I also developed and showcased a prototype that implements this model. I showed how each step of the process, regarding both linked data and augmented reality, can be handled, in a sufficient generic manner, which can be replicated by other researchers. I pointed out that integrating linked data in mobile augmented reality applications has the potential of leading to more and better data for the tourist. I have discussed what type of problems can show up, how to overcome them and what are some next steps in improving this kind of applications. Next, I present in detail the conclusions that belong to this chapter.

The focus of the present thesis is the integration of linked open data, especially open governmental data, in mobile augmented reality applications. I have studied some projects that tackled this issue and have found a lack of a clear model for obtaining the desired results. As such, I proposed a model that, naturally, combines the architecture of developing a generic linked data application with the general linked open government data lifecycle and that highlights the particularities of the augmented reality scenarios.

To prove the usefulness of this model, I showed how a prototype of a mobile augmented reality application was developed, which follows closely the proposed model. The prototype is a mobile application for tourists visiting Romania.

In the identification phase, I showed some criteria for selecting proper data for the desired integration. The selection process takes into account the quantity of data available for a certain geographic area, the quality of data and the usage license. I pointed out that there is a minimum set of descriptors that the data should have for augmented reality exploitation and what the optional – but recommended – descriptors are. I also suggested some places on the Internet where developers should look up for data sources. For demonstration purposes, I proposed the usage of DBpedia, LinkedGeoData and the Romanian Open Data Portal.

The modelling phase concerns directly the open governmental data. The ontologies to be used depend greatly on the type of the content in the dataset. For easier aligment later, it is advisable to also try to use parts of the ontology that is the most relevant in the other sources, if there is one as such. I used vocabularies such as FOAF and Basic Geo and a small part of the DBpedia ontology.

The generation phase can be manual or (semi-)automatic, depending on the quantity of data, and concerns only the open governmental data. Because I worked with a relatively small dataset, the list of museums from the Romanian Open Data Portal, I used the well-known tool called OpenRefine to transform the raw data into linked data.

I skipped the publication phase for the moment, as it did not serve our purpose for the prototype.

The integration phase has three components: access, proper integration and storage. To integrate the datasets for the prototype, I used the LDIF platform and explained the reasons for choosing it.

Access methods vary from dataset to dataset (RDF, SPARQL, REST API or dump). As data access pattern, I favoured the crawling pattern, as it offers access to data in a timely manner for augmented reality scenarios. The proper integration consists of vocabulary matching, entity resolution and quality evaluation and fusion. For the former two, the data required manual analysis to identify the best thresholds for aligment and identification. I proposed seven big categories for filtering the Points of Interest, in which I grouped the smaller categories present in the datasets. These big categories, along with geographic distance and name similarity (based on the Levenshtein distance algorithm), were used as criteria for merging the POIs. For assessing the quality, I used criteria such as time closeness and reputation of the localized version of DBpedia.

For storage purposes, I used an OpenRDF Sesame triple store and gave some statistics about the size of the data.

In the final phase, exploitation, I showed how the mobile augmented reality application was built and how it interacts with the triple store, which holds the integrated dataset. I preferred an augmented reality approach based just on the browser of the mobile device as this enables the application to be used by tourists anytime and anywhere, without the need of downloading a specific application. The prototype is based on an augmented reality web library called awe.js, which I had to customize in order to be able to display the Points of Interest in a proper fashion. I explained the mathematical calculations that needed to be done in order to achieve the desired effects.

Towards the end of the chapter, I discussed the issues encountered in the process described above, noting that user-generated data is usually messy, which complicates the integration process, while geographic support in current triple store is still poor, along with other considerations on the topic.

The theoretical contribution of this chapter to the thesis consists in the proposal of a model for implementing linked open data in mobile augmented reality applications for tourists, a model that resembles known tools and architectures and that it is straightforward to implement.

As a practical contribution to the thesis, I showed how the model proposed could be implemented in a real-case scenario, a mobile augmented reality web application for tourists visiting Romania. The fact that the model has also been applied succesfully for the Italian region of Trento (because the application was showcased at the International Semantic Web Conference 2014 in Trentino) proves the general and reproducible character of the model.

Some results presented in this chapter have been published in [193] and [200].

6. DATA PROFILING AND APPLICATION COMPARISON

6.1. Introduction

It is not easy to assess the precise structure and content of the data in the Web of Data and its appropriateness in specific fields of interest. Also, it is still unclear if and how it complements and adds further value to the user-generated data and if, together, they provide a more appropriate source of information for the targeted domain.

This was also the challenge that I faced in my research on integrating these types of data in the mobile application for tourists, which is based on augmented reality technologies.

In this chapter, I present an overview of the current approaches in the research literature for profiling the Web of Data, I introduce my approach and I show how to process the data to get comparable datasets and detailed statistics on the structure and the content of the data in the analyzed datasets. In the end, I compare similar applications and discuss the statistics presented here.

6.2. Related work on profiling the Web of Data

Our desired analysis is part of the challenge of profiling the Web of Data that has been approached at some extent already in the research literature. Profiling data is one of the important prerequisites for data integration, along others such as query optimization, data cleansing, scientific data management and data analytics [203]. Profiling data is a much-needed task, because of the open nature of the web, which allows anyone to say anything online. Consequently, usually, there is no certainty about the content and the structure that one will find in the data.

Researchers are studying some typical challenges that are encountered in profiling the Web of Data [204]. This type of profiling differs in many aspects from profiling the classic relational data, for which well-established tools and methods already exist. Commercial tools for relational data include IBM InfoSphere Information Analyzer[35], Microsoft Integration Services[36] or Informatica Data Explorer[37]. In the Web of Data world, the usual means for describing a dataset are the voiD [90] vocabulary and the Semantic Sitemaps [89], but they do not describe the information extensively. What is missing is a more detailed insight into these datasets.

An early project dealing with profiling the Web of Data is RDFStats [205], which is a framework for generating statistics both on RDF documents and on SPARQL endpoints. The tool was published as an open source project and was built to be extensible in the future. Although it was primarily developed for the SemWIQ federator and optimizer [206], it can be integrated in other platforms as well.

Another early project is ExpLOD [207], which allows the user to explore summaries of RDF data in a dataset along with the interlinking of the dataset with others from the Linked Open Data cloud.

A project that tackles the bigger picture of the Web of Data is LODStats [208], a tool written in Python which calculates 32 different statistical criteria and was developed to be specifically interlinked with the Data Hub, so as to provide a big picture of the Web of Data. It covers quality analysis, coverage analysis, privacy analysis and link target identification.

The first web-based tool for profiling – and not only – the Web of Data is ProLOD++ [209], which is a successor of the ProLOD tool [210]. ProLOD++ is able to perform tasks related to profiling, mining and cleansing arbitrary datasets provided by the user of the tool. One of the basic operations is calculating the number of occurrences of distinct predicates, along with their values. A demo is available online[38] but does not seem to work at the moment with the preloaded datasets, and does not allow uploading of new datasets by the user.

A recent project was the winner of the Open Track at the 2014 Semantic Web Challenge[39] and is called RapidMiner Linked Open Data extension [211], as it is a software module built specifically for the RapidMiner platform[40], which is a powerful predictive analytics platform for data analysis and not only. The extension allows importing of RDF data into the platform (via RDF dumps uploading or SPARQL endpoints querying) and using the wide range of RapidMiner operators to analyze the data in various ways, including generating statistics. The extension also allows one to extend the knowledge on local data by searching and linking to data in the Linked Open Data cloud.

In [212], the researchers propose an automated approach for generating structured profiles that describe the topics covered by linked datasets. These

---

35 http://www-03.ibm.com/software/products/en/ibminfoinfoanal
36 https://msdn.microsoft.com/en-us/ms141026.aspx
37 https://www.informatica.com/content/dam/informatica-com/global/amer/us/collateral/data-sheet/data-explorer_data-sheet_7011.pdf
38 https://www.hpi.uni-potsdam.de/naumann/sites/prolod++/app.html
39 http://challenge.semanticweb.org/2014/winners.html
40 https://rapidminer.com/

profiles are exposed in a format based on the Vocabulary of Interlinked Datasets (VoID[41]) and Vocabulary of Links (VoL[42]).

In the next chapter, I describe my approach in profiling the datasets exploited by the augmented reality application.

6.3. Profiling the datasets exploited in LOD4AR

This subchapter starts by presenting a general methodology for data profiling in the industry. Next, I propose an approach to apply this general methodology for profiling and assessing appropriatness of linked datasets for augmented reality scenarios. The remaining part of the subchapter is a report on the detailed application of the methodology on the targeted datasets and the results that were obtained.

6.3.1. Methodology for data profiling and assessment in augmented reality scenarios

The general methodology for data profiling specifies the following big steps [213]:

1. Prepare for the project
2. Prepare for the analysis
3. Extract and format the data
4. Sampling
5. Analysis

Applying this methodology for profiling and assessing the appropriatness of linked datasets for augmented reality scenarios demands some particularities to be taken into account.

Generally, the first two steps involve deciding on the scope of the activity, training the team, software setups etc. For our particular case, the team is composed of just one developer, whose scope is to asses the usefulness of some datasets for exploitation in augmented reality scenarios.

The third step, which consists in extracting and formatting the data, is crucial for correctly analyzing the data afterwards. The targeted linked datasets might be accessible in various ways, such as RDF, SPARQL, REST API or data dump. If the data profiling tool works offline, then it is necessary to create a dump of that data. In addition, the serialization format of the data can vary between RDF/XML, Turtle, N-Triples or JSON-LD. Because there are not so many tools for profiling directly the linked data, one might be constrained to extract only the necessary information from the dataset and save it in a common file format such as CSV. This way, one can use regular data profiling tools to analyze the data.

In augmented reality scenarios for tourism, it is advisable to do the data profiling on a certain geographic area, which is of interest for the tourist using the application. However, given the incompleteness issue of the Web of Data, it is not trivial to identify and download the POIs that are in a certain area.

41 http://www.w3.org/TR/void/
42 http://data.linkededucation.org/vol/

The fourth step, sampling the data, is required for one to get decent processing times during analysis. Of course, sampling is compulsory only when the dataset is very large. This depends, as stated in the previous paragraph, on the size of the targeted geographic area.

The fifth step is the actual analysis of the data. The targeted linked datasets can be assessed from various points of view, one of them being how well they cover all the necessary information for a touristical augmented reality application. In addition, because the application integrates data from multiple sources, it is important to measure how well the datasets complement each other.

Good content for an augmented reality application for tourists should cover the following information:

- name (official name, nickname etc.)
- description (short description – for a snippet, long description – for full page information etc.)
- picture (thumbnail, full picture, album of pictures etc)
- contact (website, email, phone etc.)
- address (geographic coordinates, address, city, street, number etc.)
- category (structural type, functional type etc.)
- provenance (contributor, source of information, last updated etc.)
- other (accessibility, parking, etc.)

There are many types of information – predicates in RDF – in the Web of Data, but only a few show up in the data consistently, so they can be deemed important. I ignored the predicates that appeared for less than 1% of the total number of POIs.

6.3.2. The process for extracting the necessary data

The third step of the methodology consists in extracting and formatting the data. In this subchapter, I show how the process was applied in the case of LOD4AR.

For building a consolidated dataset, I used information from DBpedia, LinkedGeoData and the Romanian National Open Data Portal. From the former two I chose the data about various POIs (Points of Interest) in Romania. From the latter one I chose a dataset that contains all the museums in Romania, as provided by the National Institute of Heritage.

A major problem that I encountered was deciding, for DBpedia and LinkedGeoData, which Points of Interest belong to Romania and which not. The initial download of the information was done by getting all the data that was geotagged with GPS coordinates placed on circular area centered on the geographical midpoint of the country. Given the shape of the country, it is obvious that in this way, many POIs were downloaded, POIs that belong to neighboring countries. It was not possible to download just the POIs that are inside the borders of Romania, on other criteria than the GPS position, as the information is user-generated and as such incompletely or wrongly tagged with a country name or code. In addition, I was not able to identify in the research literature a well-established method that allows downloading POIs from DBpedia or LinkedGeoData and that belong just to a single country.

To solve this issue, I used a three-step approach (just for the LinkedGeoData dataset, but the process is similar for DBpedia) that is detailed below.

1) Downloading the data from the server. The downloaded dataset from LinkedGeoData is stored on our university's Sesame server, which has a SPARQL interface. Using a query on this interface, I generated a CSV file from the server, which had three columns: URI of the POI, latitude and longitude. Next, using the OpenRefine[43] tool, I cleaned the CSV file, removing the datatype notations, to be able to process the data further in a straightforward manner.

Consequently, lines from the original downloaded file like this

```
<http://linkedgeodata.org/triplify/node2498136757>
       "45.4051"^^<http://www.w3.org/2001/XMLSchema#double>
       "25.4797"^^<http://www.w3.org/2001/XMLSchema#double>>
```

were turned into this

```
<http://linkedgeodata.org/triplify/node2498136757>,45.4051,25.4797
```

The SPARQL query that I used is showed below.

```
SELECT distinct ?poi ?lat ?long WHERE
{
?poi <http://www.w3.org/2003/01/geo/wgs84_pos#long> ?long.
?poi <http://www.w3.org/2003/01/geo/wgs84_pos#lat> ?lat.
}
LIMIT 100000
```

2) Processing the data to identify the correct country for each POI. To test that a pair of GPS coordinates describes a geographical point that is inside the borders of a country, I used an open source algorithm[44] published on GitHub. The code uses the World Borders dataset, which is available online[45] with a Creative Commons Attribution-Share Alike License.

This algorithm (showed below) runs through the previously cleaned CSV file and returns, for each pair of GPS coordinates, the country code that the corresponding geographical point belongs to, which it writes in another CSV file.

```
import csv
import urllib
import urllib2
import time
import json
import countries

f = open('linkedgeodata-ro-only-pois.csv')
```

43 http://openrefine.org/
44 https://github.com/che0/countries
45 http://thematicmapping.org/downloads/world_borders.php

```
csv_f = csv.reader(f)

stats = []
for row in csv_f:
  poi = row[0]
  lat = row[1]
  long = row[2]

  cc = countries.CountryChecker('TM_WORLD_BORDERS-0.3.shp')
  country = cc.getCountry(countries.Point(float(lat), float(long)))

  if country is not None:
    code = country.iso
  else:
    code = 'NONE'

  stats.append([poi,lat,long,code])


with open('stats.csv', 'wb') as csvfile:
  csvwriter = csv.writer(csvfile, delimiter=',')
  for row in stats:
    csvwriter.writerow(row)
```

A summary of the result of running the algorithm shows how many POIs were identified in each country: 4047 in Bulgaria, 3319 in Hungary, 3753 in Moldova, 143 in Poland, 23171 in Romania, 2225 in Serbia, 2690 in Slovakia, 4224 in Ukraine and 194 unidentified. The number of POIs that were retrieved but are in another country than Romania is quite high, so running this algorithm is clearly justified.

As it can be seen, the algorithm could not identify the country for 194 POIs. Because the number was small, I used a manual method to identify the country for these POIs. I overlaid these POIs on a map using Google Fusion Tables (see Figure 6-1). It turned out that most of the unidentified POIs are located in Romania, on the seashore, so it seems that the algorithm has problems in identifying POIs that are placed in the respective geographical area. The several (few) POIs that were located in the neighboring countries were tagged by me manually with the correct country code. I tagged the rest of them with the Romanian country code using proper tools from OpenRefine.

3) Generating the country triples for Romania. I exported from OpenRefine a CSV file only with the POIs belonging to Romania. Using the previous CSV file and OpenRefine, I generated RDF triples for each POI, through which I stated that the POI has the country code of Romania (using the LinkedGeoData ontology). Below is an example of a line from the generated CSV file.

```
<http://linkedgeodata.org/triplify/node2498136757>
<http://linkedgeodata.org/ontology/country> "RO" .
```

I uploaded this file to the Sesame server through the Sesame Workbench. The initial number of Romanian POIs was calculated by counting the POIs that were tagged in various ways as belonging to Romania (I manually analyzed the data to determine these country-specific triples). I found 891 POIs that had the properties *lgdo:is_in:country*, *lgdo:addr:country*, *lgdo:country* or *lgdo:is_in:country_code* with the values “Romania” (for the first one) and “RO” for the latter ones.

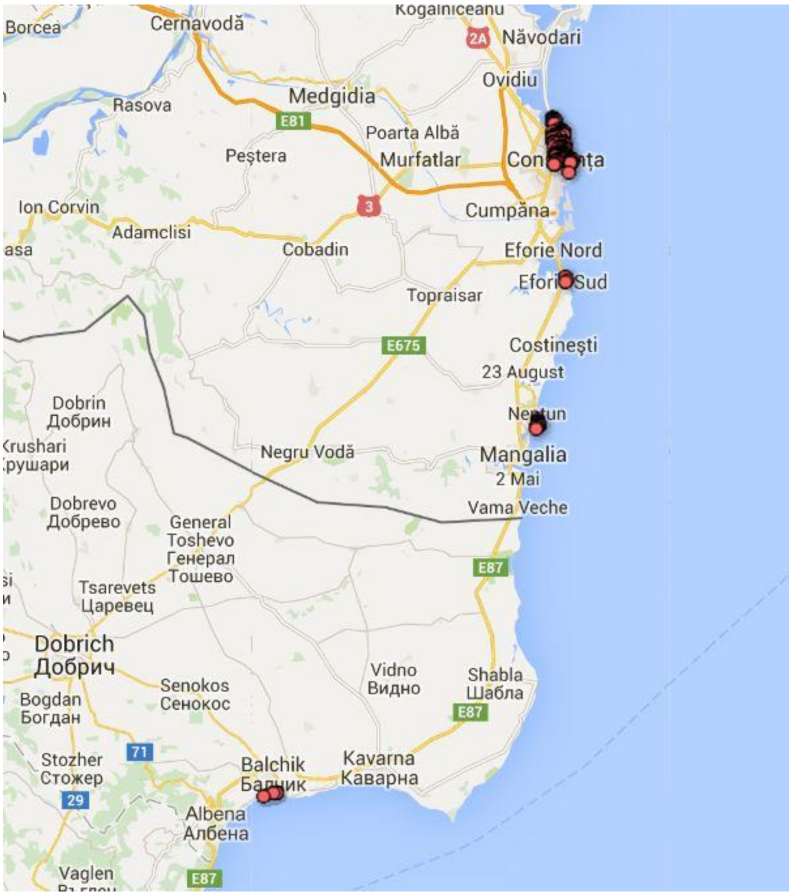


Figure 6-1 Unidentified POIs from LinkedGeoData (Screenshot © 2015 Google - Map data © 2015 Google, used with permission)

I ran the following query for this:

```
SELECT (COUNT(distinct ?s) AS ?no)
WHERE
{{
?s <http://www.w3.org/2003/01/geo/wgs84_pos#long> ?long1 .
?s <http://www.w3.org/2003/01/geo/wgs84_pos#lat> ?lat1 .
?s <http://linkedgeodata.org/ontology/is_in%3Acountry> "Romania" .
```

```
}
UNION
{
?s <http://www.w3.org/2003/01/geo/wgs84_pos#long> ?long2 .
?s <http://www.w3.org/2003/01/geo/wgs84_pos#lat> ?lat2 .
?s <http://linkedgeodata.org/ontology/addr%3Acountry> "RO" .
}
UNION
{
?s <http://www.w3.org/2003/01/geo/wgs84_pos#long> ?long3 .
?s <http://www.w3.org/2003/01/geo/wgs84_pos#lat> ?lat3 .
?s <http://linkedgeodata.org/ontology/country> "RO" .
}
UNION
{
?s <http://www.w3.org/2003/01/geo/wgs84_pos#long> ?long4 .
?s <http://www.w3.org/2003/01/geo/wgs84_pos#lat> ?lat4 .
?s <http://linkedgeodata.org/ontology/is_in%3Acountry_code> "RO" .
}}
```

The resulting number of Romanian POIs, after importing the generated RDF triples, is 23360, which is more than 25 times the initial value.

The final number was calculated with the following query:

```
SELECT (COUNT(distinct ?s) AS ?no)
WHERE
{
?s <http://www.w3.org/2003/01/geo/wgs84_pos#long> ?long .
?s <http://www.w3.org/2003/01/geo/wgs84_pos#lat> ?lat .
?s <http://linkedgeodata.org/ontology/country> "RO" .
}
```

The technique described above can be used to generate correct country tags for all the geotagged POIs, either for DBpedia or for LinkedGeoData, and thus improve these widely used sources of information.

Having the POIs' location correctly identified, I proceeded with analyzing the structure and the content of the information, which corresponds to step five in the general methodology, which is data analysis. The fourth step of the methodology has been skipped, as the dataset obtained previously is small enough to not require sampling.

I calculated some statistics for all the three sources of information and only for the POIs that are placed in Romania, as determined by using the steps mentioned above. To generate the desired statistics, I used code similar to the one below.

```
import csv
import urllib
import urllib2
```

```
import time
import json

f = open('lista_predicate_linkedgeodata_ro.csv')
csv_f = csv.reader(f)

stats = []
for row in csv_f:
  data = {}
  temp = 'SELECT (COUNT(distinct ?s) AS ?no) WHERE { ?s
<http://www.w3.org/2003/01/geo/wgs84_pos#long> ?long . ?s
<http://www.w3.org/2003/01/geo/wgs84_pos#lat> ?lat . ?s
<http://linkedgeodata.org/ontology/country> "RO" . ?s
<http://linkedgeodata.org/ontology/operator> ?thing . } LIMIT 100000'
  data['query'] =
temp.replace('http://linkedgeodata.org/ontology/operator', row[0])
  data['Accept'] = 'application/sparql-results+json'
  data['infer'] = 'false'
  url_values = urllib.urlencode(data)
  url = 'http://193.226.10.115/openrdf-sesame/repositories/svlod'
  full_url = url + '?' + url_values
  data = urllib2.urlopen(full_url).read(10000)
  stats.append([row[0],
json.loads(data)['results']['bindings'][0]['no']['value']])

with open('stats.csv', 'wb') as csvfile:
  csvwriter = csv.writer(csvfile, delimiter=' ')
  for row in stats:
    csvwriter.writerow(row)
```

6.3.3. Statistics for DBpedia

The DBpedia Ontology (dbo) is a shallow, multi-domain ontology which was extracted from Wikipedia through hand-made rules [54]. As a result of replicating Wikipedia, DBpedia contains information about POIs that are rather well-known in their area and are, according to internal rules, notable subjects [214]. As such, it mostly contains information about important institutions and touristic venues, and less about shopping places or utilities.

Analyzing the RDF predicates in the DBpedia dataset, to determine how well the POIs are described, I selected several of them that are significantly interesting for an augmented reality application and categorized them as such: name, description, picture, contact, address, category and provenance. The detailed list of predicates and their occurrences is presented in Table 6-1.

Of course, there were a lot more properties that could be included in these categories (I found 1694 predicates in total). However, in general, I ignored the predicates that occurred extremely infrequently (in less than 1% of the number of POIs that were geotagged, which is 3564).

Table 6-1 Categorization and number of occurences of RDF predicates in DBpedia

Category	Description	Occurrences	Predicate
Name	label	3417	http://www.w3.org/2000/01/rdf-schema#label
	name	706	http://xmlns.com/foaf/0.1/name
	name	553	http://dbpedia.org/property/name
	official name	122	http://dbpedia.org/property/officialName
	other name	60	http://dbpedia.org/property/otherName
Description	comment	856	http://www.w3.org/2000/01/rdf-schema#comment
	abstract	856	http://dbpedia.org/ontology/abstract
Picture	thumbnail	439	http://dbpedia.org/ontology/thumbnail
	depiction	439	http://xmlns.com/foaf/0.1/depiction
	photo collection	832	http://dbpedia.org/property/hasPhotoCollection
Contact	website	157	http://dbpedia.org/property/website
	homepage	224	http://xmlns.com/foaf/0.1/homepage
	external link	377	http://dbpedia.org/ontology/wikiPageExternalLink
	wikipedia	856	http://xmlns.com/foaf/0.1/isPrimaryTopicOf
Address	city	136	http://dbpedia.org/ontology/city
	address	28	http://dbpedia.org/ontology/address
	address	28	http://dbpedia.org/property/address
	latitude	3564	http://www.w3.org/2003/01/geo/wgs84_pos#lat
	longitude	3564	http://www.w3.org/2003/01/geo/wgs84_pos#long
Category	type	1002	rdf:type
	type	474	http://dbpedia.org/ontology/type
	subject	856	http://purl.org/dc/terms/subject
Provenance	wiki page ID	856	http://dbpedia.org/ontology/wikiPageID

In DBpedia, while almost all of the POIs have a name, only a minority of them have a shorter (abstract) or longer (comment) description. Even fewer have attached a photo of the POI, an important asset for an augmented reality application, as a photo greatly helps the user to identify the POI that she is searching for. The *address* category lacks information on the actual address of the POI (street, number, house), so the only reliable information consists of the GPS coordinates. For the *contact* category, I notice that no information is included for email or phone numbers, as it is almost inexistent in DBpedia. For the *provenance* category, I considered the property *http://dbpedia.org/ontology/wikiPageID*, which points to the ID of the Wikipedia page where the information was generated from.

Starting from here, one can find out, theoretically, the user(s) that created the information. As it turns out, this is a very indirect and rather unusable provenance information for an augmented reality application. The *category* category shows that only about half of the POIs are categorized somehow, a fact that hinders one of the most important aspects of a good augmented reality feature, the filtering option.

6.3.4. Statistics for LinkedGeoData

The LinkedGeoData Ontology (lgdo) is a lightweight OWL ontology that was crafted based on the tags existent for each POI in LinkedGeoData, which work as key-value pairs for the pieces of information [59]. As a result of LinkedGeoData replicating the data from OpenStreetMap, it usually contains information about a great variety of POIs, from ATMs to museums, information which is, however, rather shallowly described.

Analyzing the predicates in the LinkedGeoData dataset, to determine how well the POIs are described, I selected several of them that are significantly interesting for an augmented reality application and categorized them as such: name, description, address, accessibility and parking, contact, category and provenance. The detailed list of predicates and their occurrences is presented in Table 6-2.

Table 6-2 Categorization and number of occurences of RDF predicates in LinkedGeoData

Category	Description	Occurrences	Predicate
Name	label	15736	rdfs:label
Description	comment	347	rdfs:comment
Address	city	1591	http://linkedgeodata.org/ontology/addr%3Acity
	street	2432	http://linkedgeodata.org/ontology/addr%3Astreet
	house number	2116	http://linkedgeodata.org/ontology/addr%3Ahousenumber
	latitude	23360	http://www.w3.org/2003/01/geo/wgs84_pos#lat
	longitude	23360	http://www.w3.org/2003/01/geo/wgs84_pos#long
Accessibility and parking	wheelchair	386	http://linkedgeodata.org/ontology/wheelchair
	parking	408	http://linkedgeodata.org/ontology/parking
Contact	homepage	474	foaf:homepage
	phone	463	foaf:phone
	opening hours	600	http://linkedgeodata.org/ontology/opening_hours
Category	type	23360	rdf:type
	type (ro)	109	http://linkedgeodata.org/ontology/_Tip_
Provenance	contributor	23360	http://purl.org/dc/terms/contributor
	last modified	23360	http://purl.org/dc/terms/modified
	source of info	2132	http://linkedgeodata.org/ontology/source
	link to source	465	http://linkedgeodata.org/ontology/source%3Alink

Of course, there were a lot more properties that could be included in these categories (I found in total 483 predicates). However, in general, I ignored the predicates that occurred extremely infrequently (in less than 1% of the number of POIs that were geotagged, which is 23360).

LinkedGeoData features many predicates that vary only slightly by name (e.g. *short_name*, *_Nume_*, *_nume_*, *old_name%3Aen,* etc.) which is a result of the fact that users can add tags, as they wish, when describing a POI in OpenStreetMap. There are no properties for images or pictures of the POIs. The *address* (except GPS coordinates) and *contact* related properties show up rather infrequently. More than half of the POIs have some kind of label, although very few have a description. It is interesting to note the presence of some small pieces of information on the accessibility of the POIs. Contrarily to DBpedia, all the POIs are categorized and provenance is well described is terms of contributor and date of last modification (for all POIs) and source of information and link to it (for some POIs).

6.3.5. Statistics for the museums dataset from the Romanian Open Data portal

To publish this governmental dataset as linked open data, I employed some parts of the FOAF and Basic Geo vocabularies, as well as a part of the DBpedia ontology, along with custom defined properties.

Analyzing the predicates in the museums dataset, I similarly selected several of them that are significantly interesting for an augmented reality application and categorized them as such: name, description, address, contact and category. The detailed list of predicates and their occurrences is presented in Table 6-3.

Table 6-3 Categorization and number of occurences of RDF predicates in the museums dataset

Category	Description	Occurrences	Predicate
Name	label	951	rdfs:label
Description	comment	951	rdfs:comment
Address	city	951	http://tom7.cm.upt.ro/onto/cityvalue
	address	673	http://tom7.cm.upt.ro/onto/addressvalue
	latitude	951	http://www.w3.org/2003/01/geo/wgs84_pos#lat
	longitude	951	http://www.w3.org/2003/01/geo/wgs84_pos#long
Contact	website	534	foaf:homepage
	phone	730	foaf:phone
	opening hours	834	http://tom7.cm.upt.ro/onto/hoursvalue
	email	451	foaf:mbox
Category	type	951	rdf:type

The dataset is well described, with names and descriptions for all 951 geotagged POIs (in total there were 967 POIs in the dataset). Half or more than half have addresses, websites, phones, emails and opening hours specified. All the information is generally given in both Romanian and English.

As a result of the fact that the dataset was released specifically as the list of museums in Romania, the category information is clear: all the POIs are museums. In addition, as it was published on the National Open Data Portal, this is a good indication, although not complete, of its provenance.

6.4. Comparison with similar applications on exploited data

Several projects based on augmented reality visualization techniques (and not only) have tackled the integration of linked open data sources, mainly from general knowledge repositories. These projects have been described in detail in chapter 4. A short overview on their purpose and exploited data is shown in Table 6-4.

Table 6-4 Overview of similar projects (purpose and exploited data)

Project	Year	AR	Purpose	Exploited data
mSpace Mobile [141]	2005	Non-AR	Tourism	Open Guide to London, IMDB, BBC
DBpedia Mobile [142]	2009	Non-AR	Tourism (urban)	DBpedia, Geonames, flickr wrappr and other sources reachable from DBpedia
Cultural heritage mobile guide [149]	2010	AR	Providing cultural heritage resources for an end user	General knowledge platforms (GeoNames, LinkedGeodata, Freebase, DBPedia) and platforms specialized on cultural heritage (e.g. Art and Architecture Thesaurus, Union List of Artist Names) or platforms of individual cultural institutions
Smart Reality [215]	2012	AR	Young people interested in listening to music and attending concerts	Play.fm (more than 18000 DJ mixes and live recordings); other sources crawled, starting from the URI defined by the person who is annotating the poster
Mobile mountain guide [37]	2012	AR	Visualizing mountain-specific data	Geonames, LinkedGeoData
ARCAMA-3D [147]	2013	AR	Generic surroundings discovery focusing on topic experiences	Direct linking to DBpedia
LOD4AR	2014	AR	Tourism	DBpedia, LinkedGeoData, Romanian Open Data Portal

These projects relate on their findings in terms of challenges and approaches in integrating linked open data in augmented reality applications. However, none of them gives a detailed overview of the structure, content and appropriateness of the integrated data for an augmented reality-based application.

However, some of them report on the experiences of dealing with real-world data, such experiences being already presented in subchapter 4.2.5. Their findings are similar to those that showed up during the implementation of LOD4AR.

It is hard to compare the amount of data exploited in these projects. In general, the projects use one or more of the major hubs of the LOD cloud (DBpedia, LinkedGeoData, Geonames) and some other sources of information. Data is mainly fetched as Linked Data. Sometimes, the classical HTML crawling is employed. Data is fetched live (with the exception of the mobile mountain guide [37]) and only for the area that the user is in. All the projects (including LOD4AR) declare that they are extensible and that data sources can be added as needed.

The principle for fetching the data differs from project to project. The cultural heritage mobile guide [149] fetches the data "on the spot", with the argument that the loading time is not that critical if the mobile application starts retrieving the data when the user begins going towards a certain POI. The mobile mountain guide [37] has the data already replicated on the smartphone, and only does the merging and alignment on the spot, with processing times that are considered reasonable by the researchers. The SmartReality project uses a combination of cached data and live crawling of the data. LOD4AR works only with cached data.

LOD4AR, as opposed to other projects, focuses on open data released by governments worldwide, mostly in the form of tabular data, such as XLS or CSV. It shows how such datasets can be mixed up with generic, well-known user-generated datasets such as DBpedia and LinkedGeoData and what the results look like.

6.5. Comparison with similar applications from other points of view

A comparison between these projects can be made from other points of view as well.

Regarding the architecture used, all the projects employ a typical client-server architecture. In the case of the mobile mountain guide [37], both the client and the server reside on the same machine, the mobile phone, as the RDF data is replicated on the smartphone. This is also the only project that makes the data available offline, a choice that supports the use case chosen, a guide for a mountain region, where there is usually a lack of internet connection. In general, the projects were developed to be showcased in an urban space, so the premise on which they were built was that an internet connection was always available.

The projects studied above vary greatly in terms of availability. The applications that are built as mobile websites (this includes LOD4AR) have a better chance to stay available online for a long time. They can be accessed anytime and anywhere, if a data connection is available, without the need to download an application on the smartphone. Native applications are usually developed only as a prototype and do not get to the point of being released publicly before the project ends.

LOD4AR is available on the latest versions of Chrome and Firefox on Android devices (as tested by the author of this thesis), theoretically being available on the latest browsers on many mobile devices, including smartphone, tablets, eye glasses and so on. Other projects seem to be available only on the smartphones that they were implemented and tested on.

As opposed to other projects, LOD4AR gives a recipe (through an example) of how to integrate linked open data in an augmented reality application and gives some deeper statistics regarding this integration. Other projects explain more or less how they integrated the data and what results they obtained overall (not only for a certain POI presented as study case).

6.6. Discussion

DBpedia and LinkedGeoData are two of the backbones of the Web of Data, in general, and of the world of geo linked data, in particular. They complement each other, due to the purpose for which their original counterparts were created: Wikipedia features rich information on notable POIs, while OpenStreetMap strives to equally cover smaller and bigger POIs, although not that deep. Geonames is also regarded as a big player in this field; however, I considered it not to be that interesting for an augmented reality application for tourists, as it mostly contains information about administrative regions in Romania (counties, cities, villages etc.).

Due to the nature of the open content, these big user-generated hubs of information expose data that is rather unpredictable, incomplete and error-prone.

Except the textual information, these platforms lack the really useful elements for an interactive and eye-catching augmented reality application: images, videos or 3D content. While a small part of the POIs features an image, videos or 3D content are almost non-existent.

Such elements would be interesting not only for the richness of the application, but also for enabling the applications to use image-based recognition techniques, which would lead to a better augmented reality experience. The fact that almost all the similar projects (the exception being SmartReality) use location-based augmented reality is a proof of the lack of images, especially, in user-generated platforms.

GPS data is poor on user-generated portals of information, a fact that is clearly linked to the technical difficulty for the common contributor in obtaining richer GPS information. Augmented reality applications rely on 3D models of the buildings for proper identification and augmentation of the surroundings, while these big platforms lack even GPS coordinates for the boundaries of the POIs (the only information provided is a single pair of GPS coordinates, as if the POI is just a point on the ground).

On the other side, the governmental dataset is well built, with more complete information, at least for the properties that it features. Also, it is inherently more reliable as a source of information due to its governmental origins. However, the range of properties that such information has is rather limited, as the government usually collects only some small data. It also lacks the same useful elements for a good augmented reality application.

Based on the previous statistics and on the discussion above, I can conclude that datasets should contain more specific GPS information (at least the boundary and the height), with a 3D model of the POI being the ideal target; they should feature more photos and videos, as interactive elements; and they should include URIs of the same objects as they are described in the Linked Open Data cloud, for proper linking and information crawling. These guidelines for publishing (governmental) open data suitable for exploitation in augmented reality applications should lead to better and more useful applications for the end user.

Regarding the application itself, I observe the universality of the client-server type of architecture in the reviewed projects, which has been proven successful on the WWW. In addition, I note that web-based technologies used for creating augmented reality applications seem to provide a longer lifetime for the application itself and a wider range of platforms on which it can run.

6.7. Conclusions

Integrating linked open data in mobile augmented reality applications has certain benefits, most of them related to the removal of the limitations imposed by classical databases that are used nowadays in augmented reality applications.

In this chapter, to assess the appropriateness of the linked open data for augmented reality applications, I proceeded to profile the data in order to get an overview about the structure and the content of the data sources.

To put the effort in context, I reviewed the research literature on profiling the Web of Data, which is clearly needed due to the fact that metadata is very shallow or non-existent in linked datasets. Specific challenges of profiling this data include heterogeneity of vocabularies and performance times for large datasets.

The literature review did not reveal techniques for profiling data that report on the appropriateness of it for exploitation in augmented reality applications. Consequently, I proposed a methodology for assessing this appropriatness based on known data profiling techniques. I described this methodology, noting the foundation that it is based on and the criteria that we look for in the Web of Data.

To apply the methodology to the datasets that I exploited in the augmented reality application described in the previous chapter, LOD4AR, I first needed to make sure that the datasets covered the exact same geographic area, so I could make a fair comparison between them. I proposed and described a process for obtaining the data for just one country, in our case Romania, from DBpedia and LinkedGeoData.

Afterwards, consistent with the methodology process, I generated statistics for the properties that occur in the data and categorized them on criteria of relevance for augmented reality applications.

I found that the user-generated hubs of information have shallow GPS data, only a small part features photos (videos and 3D content are non-existent), categorization varies in quality and coverage, hindering proper filtering of information, provenance information depends on how the linked source of information was built, and other information, such as contact, is only partially present.

Open data from governmental sources, although well built and more reliable, still lacks information necessary for a good and useful augmented reality application, but can complement nicely the user-generated information.

After analyzing my own sources of data, I proceeded to compare LOD4AR with other similar projects, first from the point of view of the exploited data. I note there is not enough information in the literature to be able to asses the quantity of the data being exploited in similar projects and, in general, there is almost no assessment about the approapriatness of the data for augmented reality applications, except some experiences of dealing with real-world data.

The projects use one or more of the major hubs of the LOD cloud and some other sources of information, accessed as linked data or crawled as HTML. The data is either cached and served afterwards, or is fetched live. A combination of both methods was also employed.

In the end, I compared the projects from other points of view, such as architecture, availability and reproducibility. I noted that all the projects use a client-server architecture, the most common approach on web today. Also, I observed that LOD4AR is better positioned regarding availability, due to the fact that it works as a website in the mobile browser, and regarding reproducibility, as the model used to develop it is described in detail in the present thesis, as opposed to similar projects.

The theoretical contributions of this chapter consist in proposing a methodology for profiling linked open data sources and assessing the information depending on its usefulness for augmented reality scenarios. The practical contributions consist in developing a process for detecting and correctly labeling a POI with the country that it belongs to.

7. CONTRIBUTIONS AND CONCLUSIONS

7.1. Conclusions

The research that backs up this thesis was done as part of the Multimedia Research Centre, Faculty of Electronics and Telecommunications, Politehnica University Timisoara. The research activity was complemented by volunteer activities that I was involved in, for promoting open data principles, as Ambassador for Romania of the Open Knowledge Foundation and cofounder of the open data oriented Smart City Association in Timisoara. The practical research contributions were done in the context of the candidacy of the city of Timisoara for the title of European Capital of Culture in 2021.

Therefore, my research can be summarized as a proposal for a model and an implementation of an application that proves the integration of linked open data in augmented reality applications helps tourists to discover more and better information in a smart city that they visit. The exact flow of the research and the main findings are detailed below.

My initial focus was on augmented reality technologies. I analyzed the general landscape of AR, the main research areas in this field and the current status and evolution of the technologies. I focused on augmented reality browsers, highlighting through a critical review their main limitations today, of which I consider the static, limited content to be a major obstacle in their evolution.

In this context, I believe that there is great potential in exploiting the vast amount of open data existing today on the Web in various forms, including in augmented reality applications. For this exploitation to be feasible, I proposed the use of Linked Data tools and principles. I analyzed the current status of the Linked Data landscape, including its founding principles, the growth of the Linked Open Data cloud, its core standards (RDF and SPARQL) and models for publishing and consuming Linked Data. I focused on the open government data landscape, showing how a smart city application can be built using such data. My conclusion was that Linked Data is the recommended approach nowadays for exploiting and integrating heterogeneous data on the Web, thus making it my chosen approach for improving content in augmented reality applications for tourists.

Therefore, I continued by doing a critical review of research projects that already tackled such integration and studying in detail the issues that need to be approached in this endeavor. While there are some lessons learned, I found that there is a lack of a clear, straightforward model for doing such an integration.

The next main part of the thesis consists in the proposal of a model for integrating linked open data in augmented reality applications for tourists, which I have shown that it involves existing techniques, patterns and tools in the area, so it is not difficult to apply and implement by researchers and developers. To demonstrate its viability, I implemented a real-case mobile augmented reality web application that integrates linked open government data and user-generated open data for the benefit of tourists that visit Romania (and for citizens, as well). In the process, I highlighted challenges in this endeavor and showed approaches for handling them appropriately.

Afterwards, I analyzed the appropriateness of the exploited sources of linked open data for augmented reality applications and I derived some conclusions and some guidelines for publishing better data, in this sense.

I will attempt to answer the research questions that I raised at the beginning of the thesis, based on my research findings:

1. *What is the current status of AR browsers in terms of content for tourism?*
   A detailed analysis of the current landscape of augmented reality technologies was done in chapter 2, with a focus on mobile augmented reality browsers. Several limitations were identified, including poor registration with the real world, poor user experience, limited content and heterogeneous architectures and data formats. Regarding content, nowadays AR browsers make use of isolated, limited silos of information, which leads to applications that are not dynamic and adaptive as the users expect them to be, as shown in chapter 4. This is especially significant in touristic applications, which are by nature information rich.

2. *How mature and appropriate are Linked Data principles and tools for enhancing AR applications?*
   An overview of the Linked Data landscape was done in chapter 3. I described the strengths of Linked Data in terms of compliance with the actual Web architecture, standardization of the data model and the data query language and the evolution of real world Linked Open Data sets on the Web. The conclusion, as indicated in the same chapter, is that Linked Data principles and tools are nowadays the recommended approach for integrating heterogenous data, such as open government data and user-generated open data that might enhance touristic applications.

3. *What are the challenges in integrating linked data in mobile augmented reality applications?*
   The challenges were found by analyzing the projects that already addressed this integration and reviewing other significant research papers related to these domains. To summarize the findings from chapter 4, these challenges are related to geodata integration, data quality, provenance and trust issues.

4. *Is there a model for this integration and what are its particularities?*
   There is a clear model for this integration and it is presented in detail in chapter 5. The model considers all of the issues mentioned above and is based on known tools, architectures and techniques, so it is

straightforward to put in practice, as demonstrated by the implementation developed and described in the thesis.

5. *How appropriate are the content and the structure of linked open data for augmented reality applications?*
   The analysis of the exploited sources of data in chapter 6 led me to the conclusion that open user-generated data are moderately appropriate for integration in augmented reality applications. Such platforms offer some textual information but lack consistency and more interactive information such as photos and videos. Open governmental data is more complete and coherent and can complement the user-generated data, given that some guidelines are followed in the process of publicating it.

7.2. Theoretical contributions

(1) Overview of augmented reality and critical study of augmented reality components and applications

I presented the main pillars of an augmented reality system, both from a hardware and from a software point of view, mainly regarding AR for the visual sense, but not excluding other senses, such as the auditory one, the olfactory one or the kinesthetic one. The focus was on assessing the status of augmented reality applications for mobile devices.

(2) Overview of linked data and critical study of tools and architectures for publishing and consuming linked data

I presented the evolution of linked data, its relation with the bigger concept of Semantic Web, its main standards (RDF, SPARQL), the growth of the actual cloud of linked data sets on the Web and the principles and the models for publishing and consuming linked data for / by applications.

(3) Critical analysis of Linked Open Government Data and implementations, mainly in Smart City applications

I have done a critical analysis of the landscape linked open government data, including the movement of open government data, the lifecycle of Linked Open Government Data and some implementations from around the world (subchapter 3.9)

(4) Critical study of the challenges of integrating linked open data in mobile augmented reality applications for tourism

I reviewed the projects that already addressed the integration of linked open data in augmented reality applications and other research in the context of these domains and presented the main challenges faced in this endeavor: geodata integration, data quality, provenance and trust issues (chapter 4).

(5) Proposal for a model to integrate linked open data in mobile augmented reality applications for tourists

I proposed a model for implementing linked open data in mobile augmented reality applications for tourists, showing that the model resembles known tools and architectures and that it is straightforward to implement, describing the challenges that need to be addressed and some approaches to overcome them (chapter 5).

(6) Proposal for a methodology to assess the appropriateness of linked open data for exploitation in augmented reality applications

I proposed, in chapter 6, a methodology to profile linked open datasets in order to get an overview about the structure and the content of the data sources and, consequently, to assess the appropriateness of this data for exploitation in augmented reality applications for tourism.

7.3. Practical contributions

(1) The implementation of a native mobile augmented reality application for tourism based on open culture data

I developed, as coordinator for a team of students, a native mobile augmented reality application called Timisoara City Art. The application features street art in Timisoara, which the tourist can explore via a map, a list or an augmented reality interface (subchapter 2.7).

(2) The implementation of a Smart City application based on Linked Open Government Data

I showed how a simple, yet effective, application for a Smart City, called Timisoara Street History, could be developed based on linkifying open government data sets that have been published and are of interest (subchapter 3.8.4).

(3) The implementation of a web-based mobile augmented reality application that integrates various linked open data for tourism purposes

I showed how the model proposed could be implemented in a real-case scenario, a mobile augmented reality web application for tourists visiting Romania. I showed the challenges involved and approaches to overcome them (chapter 5).

(4) Description and implementation of a process for detecting and correctly labeling a POI with the country that it belongs to

Points of Interest on LinkedGeoData and DBpedia are incompletely tagged with the country information. As such, I implemented a process to detect and label the downloaded POIs with the country that they belong to, based on the GPS position that all of them have attached (subchapter 6.3.2).

(5) Analyzing the data profile of the linked open datasets that were integrated in the LOD4AR application

Based on the proposed methodology, I profiled the linked open datasets that were integrated in the LOD4AR application and analyzed their appropriateness for exploitation in augmented reality applications for tourism. A discussion on the findings can be read in subchapter 6.6.

7.4. Future research directions

Several research directions can be approached, that build on the research findings of this thesis. I describe them in brief below.

Towards Linked Open Repositories for augmented reality content. We should be able to publish augmented reality content, which we create, to open repositories. At the same time, we should be able to easily access content that someone else created, from such open repositories, so we don't have to recreate it [22, p. 262]. Such repositories would best work using Linked Data principles and technologies, because they offer the framework for heterogeneous content to be queried, processed and published using the same standards.

Smarter integration of geodata. I imagine an effort towards integrating geodata at a large scale, with algorithms for schema alignment and entity reconciliation that have good performance and precision. Integration should work flawlessly, not only with very simple geographic objects, like coordinates for a point, but also with complex geometries (such as lines and polygons). The framework that enables this integration should support dynamically adding or removing of datasets, as well as ranking the datasets, based on data quality and relevance for the task that the user has at hand.

Towards a generic augmented reality-based linked data browser. There are several Linked Data browsers developed as research projects, such as Tabulator, Disco and OpenLink [216]. I envision a generic augmented reality-based Linked Data browser which people would use to browse content, described using Linked Data principles, that is geo-tagged around them. This would benefit the Linked Data domain, as it will allow users to browse an otherwise unfamiliar machine-readable content in a more natural and engaging way.

Towards more personalization and contextualization. Linked Data has the potential of offering a huge amount of information. Even when using closed databases for content delivery, the number of Points of Interest returned to the user can be overwhelming. Semantic Web and Linked Data principles can enable reasoning over the data, such that results of users' queries can be more relevant. Results can be also personalized, for example by leveraging the FOAF profile of the user.

BIBLIOGRAPHY

[1] T. Langlotz, T. Nguyen, D. Schmalstieg, and R. Grasset, "Next-Generation Augmented Reality Browsers: Rich, Seamless, and Adaptive," *Proceedings of the IEEE*, vol. 102, no. 2, pp. 155–169, 2014.

[2] H. ten Wolde, "Apple buys German augmented-reality software maker Metaio," *Reuters*, 29-May-2015.

[3] Reuters News, "Blippar Acquires Layar, Creating World's Largest AR Userbase," *Reuters*, 18-Jun-2014.

[4] Open Knowledge, "The Open Definition." [Online]. Available: http://opendefinition.org/. [Accessed: 06-Sep-2014].

[5] EU Directorate General For Internal Policies, "Mapping the smart cities in the UE, 2014. Timisoara Commitment," 2014. [Online]. Available: http://eu-smartcities.eu/commitment/7711. [Accessed: 20-Nov-2014].

[6] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," *IEICE TRANSACTIONS on Information and Systems*, vol. 77, no. 12, pp. 1321–1329, 1994.

[7] Russell Freeman, "Virtuality Continuum," *Wikipedia, the free encyclopedia*. 22-Nov-2007.

[8] S. Mann and J. Fung, "Videoorbits on eye tap devices for deliberately diminished reality or altering the visual perception of rigid planar patches of a real world scene," *EYE*, vol. 3, p. P3, 2001.

[9] B. A. Parviz, "Augmented reality in a contact lens," *IEEE spectrum*, vol. 9, pp. 1–4, 2009.

[10] Morton Leonard Heilig, *Sensorama simulator*. Google Patents, 1962.

[11] B. Laurel, *Computers as theatre*. Addison-Wesley, 2013.

[12] Morton Leonard Heilig, *Sensorama*. Wikimedia Commons. 26-Feb-2008.

[13] I. E. Sutherland, "A head-mounted three dimensional display," in *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, 1968, pp. 757–764.

[14] H. McLellan, "Virtual realities," *Handbook of research for educational communications and technology*, pp. 457–487, 1996.

[15] B. Furht, *Handbook of Augmented Reality*. Springer New York, 2011.

[16] T. P. Caudell and D. W. Mizell, "Augmented reality: an application of heads-up display technology to manual manufacturing processes," in *Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences, 1992*, 1992, vol. ii, pp. 659–669 vol.2.

[17] L. B. Rosenberg, "The Use of Virtual Fixtures as Perceptual Overlays to Enhance Operator Performance in Remote Environments.," DTIC Document, 1992.

[18] R. T. Azuma, "A survey of augmented reality," *Presence*, vol. 6, no. 4, pp. 355–385, 1997.

[19] H. Kato and M. Billinghurst, "Marker tracking and hmd calibration for a video-based augmented reality conferencing system," in *Augmented Reality, 1999.(IWAR'99) Proceedings. 2nd IEEE and ACM International Workshop on*, 1999, pp. 85–94.

[20] B. Thomas, B. Close, J. Donoghue, J. Squires, P. De Bondi, M. Morris, and W. Piekarski, "ARQuake: an outdoor/indoor augmented reality first person application," in *The Fourth International Symposium on Wearable Computers*, 2000, pp. 139–146.

[21] C. Arth, R. Grasset, L. Gruber, T. Langlotz, A. Mulloni, and D. Wagner, "The History of Mobile Augmented Reality," *arXiv preprint arXiv:1505.01319*, 2015.

[22] A. B. Craig, *Understanding Augmented Reality: Concepts and Applications*. Newnes, 2013.

[23] J. P. Rolland, L. Davis, and Y. Baillot, "A survey of tracking technology for virtual environments," *Fundamentals of wearable computers and augmented reality*, vol. 1, pp. 67–112, 2001.

[24] F. Zhou, H. B.-L. Duh, and M. Billinghurst, "Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR," in *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, 2008, pp. 193–202.

[25] E. Costanza, A. Kunz, and M. Fjeld, *Mixed reality: A survey*. Springer, 2009.

[26] V. Kindratenko, "Calibration of electromagnetic tracking devices," *Virtual Reality*, vol. 4, no. 2, pp. 139–150, 1999.

[27] D. K. Bhatnagar, "Position trackers for Head Mounted Display systems: A survey," *University of North Carolina, Chapel Hill TR93*, vol. 10, 1993.

[28] A. Henrysson and M. Ollila, "UMAR: Ubiquitous mobile augmented reality," in *Proceedings of the 3rd international conference on Mobile and ubiquitous multimedia*, 2004, pp. 41–45.

[29] J. Carmigniani and B. Furht, "Augmented Reality: An Overview," in *Handbook of Augmented Reality*, B. Furht, Ed. Springer New York, 2011, pp. 3–46.

[30] O. Bimber and R. Raskar, *Spatial augmented reality: merging real and virtual worlds*. CRC Press, 2005.

[31] Sensics Inc., *xSight HMD*. Wikipedia, the free encyclopedia. 3-Sep-2010.

[32] Glogger, *Augmented reality map on iPhone*. Wikimedia Commons. 2009.

[33] Z. Huang, P. Hui, C. Peylo, and D. Chatzopoulos, "Mobile Augmented Reality Survey: A Bottom-up Approach," *arXiv preprint arXiv:1309.4413*, 2013.

[34] A. Harma, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, and H. Nironen, "Techniques and applications of wearable augmented reality audio," in *Audio Engineering Society Convention 114*, 2003.

[35] D. Van Krevelen and R. Poelman, "A survey of augmented reality technologies, applications and limitations," *International Journal of Virtual Reality*, vol. 9, no. 2, p. 1, 2010.

[36] Dhiraj Amin and Sharvari Govilkar, "Comparative Study of Augmented Reality SDK's," *IJCSA*, vol. 5, no. 1, pp. 11–26, Feb. 2015.

[37] S. Zander, C. Chiu, and G. Sageder, "A computational model for the integration of linked data in mobile augmented reality applications," in *Proceedings of the 8th International Conference on Semantic Systems*, 2012, pp. 133–140.

[38] T. Vincenty, "Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations," *Survey Review*, vol. 23, no. 176, pp. 88–93, 1975.

[39] C. Veness, "Calculate distance and bearing between two Latitude/Longitude points using haversine formula in JavaScript." [Online]. Available: http://www.movable-type.co.uk/scripts/latlong.html. [Accessed: 01-Apr-2015].

[40] S. Vert and R. Vasiu, "School of the Future: Using Augmented Reality for Contextual Information and Navigation in Academic Buildings," in *Proceedings of the IEEE 12th International Conference on Advanced Learning Technologies (ICALT 2012)*, Rome, Italy, 2012, pp. 728–729.

[41] N. Shadbolt, W. Hall, and T. Berners-Lee, "The semantic web revisited," *Intelligent Systems, IEEE*, vol. 21, no. 3, pp. 96–101, 2006.

[42] Marobi1, *Semantic Web Stack*. Wikimedia Commons. 29-Aug-2014.

[43] B. Dragulescu, "Tehnologii semantic web in mediul educational," PhD Thesis, Politehnica University of Timisoara, Timisoara, Romania, 2012.

[44] B. Glimm, A. Hogan, M. Krötzsch, and A. Polleres, "OWL: Yet to arrive on the Web of Data?," in *WWW2012 Workshop on Linked Data on the Web, Lyon, France, 16 April, 2012*, 2012.

[45] 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, pp. 1–22, 33 2009.

[46] Max Schmachtenberg, Christian Bizer, Anja Jentzsch, and Richard Cyganiak, "Linking Open Data cloud diagram," 2014. [Online]. Available: http://lod-cloud.net/. [Accessed: 26-Nov-2014].

[47] J. Weaver and P. Tarjan, "Facebook Linked Data via the Graph API.," *Semantic Web*, vol. 4, no. 3, pp. 245–250, 2013.

[48] T. Berners-Lee, "Linked Data - Design Issues," 2006. [Online]. Available: http://www.w3.org/DesignIssues/LinkedData.html. [Accessed: 24-Apr-2014].

[49] Ian Jacobs and Norman Walsh, Eds., "Architecture of the World Wide Web, Volume One - Fragment identifiers and content negotiation." 15-Dec-2004.

[50] L. Sauermann and R. Cyganiak, "Cool URIs for the semantic web." 03-Dec-2008.

[51] T. Heath and C. Bizer, "Linked Data: Evolving the Web into a Global Data Space," *Synthesis Lectures on the Semantic Web: Theory and Technology*, vol. 1, no. 1, pp. 1–136, Feb. 2011.

[52] D. Wood, Marsha Zaidman, L. Ruth, and M. Hausenblas, *Linked Data*. Manning, 2014.

[53] M. Schmachtenberg, C. Bizer, and H. Paulheim, "Adoption of the linked data best practices in different topical domains," in *The Semantic Web–ISWC 2014*, Springer, 2014, pp. 245–260.

[54] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann, "DBpedia-A crystallization point for the Web of Data," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 7, no. 3, pp. 154–165, 2009.

[55] DBpedia, "About DBpedia," 2014. [Online]. Available: http://dbpedia.org/About. [Accessed: 24-Feb-2014].

[56] Freebase, "RDF Overview - Freebase API," 2014. [Online]. Available: https://developers.google.com/freebase/v1/rdf-overview. [Accessed: 27-Feb-2014].

[57] O. Thereaux, S. Angeletou, M. Smethurst, and J. Tarling, "Opening up the BBC's data to the Web," presented at the Open Data on the Web, London, UK, 2013.

[58] Geonames, "GeoNames Ontology - Geo Semantic Web," 2014. [Online]. Available: http://www.geonames.org/ontology/documentation.html. [Accessed: 26-Feb-2014].

[59] C. Stadler, J. Lehmann, K. Höffner, and S. Auer, "Linkedgeodata: A core for a web of spatial open data," *Semantic Web*, vol. 3, no. 4, pp. 333–354, 2012.

[60] National Center for Biotechnology Information, "Home - PubMed - NCBI," 2014. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/. [Accessed: 25-May-2014].
[61] Association for Computing Machinery, "ACM RKB Explorer," 2014. [Online]. Available: http://acm.rkbexplorer.com/. [Accessed: 25-Feb-2014].
[62] L3S Research Center, "D2R Server publishing the DBLP Bibliography Database," 2014. [Online]. Available: http://dblp.l3s.de/d2r/. [Accessed: 24-Feb-2014].
[63] K. Alexander, R. Cyganiak, M. Hausenblas, and J. Zhao, "Describing Linked Datasets.," in *LDOW*, 2009.
[64] Nature Publishing Group, "Nature.com Developers | Linked Data Platform," 2014. [Online]. Available: http://www.nature.com/developers/documentation/linked-data-platform/. [Accessed: 25-May-2014].
[65] Library of Congress, "LC Linked Data Service," 2014. [Online]. Available: http://id.loc.gov/. [Accessed: 26-Aug-2015].
[66] Bibliothèque nationale de France, "BNF Semantic Web," 2014. [Online]. Available: http://data.bnf.fr/semanticweb. [Accessed: 27-Feb-2014].
[67] New York Times, "New York Times - Linked Open Data," 2010. [Online]. Available: http://data.nytimes.com/. [Accessed: 27-Feb-2014].
[68] EUscreen Consortium, "EUscreen Linked Open Data Pilot," 2014. [Online]. Available: http://lod.euscreen.eu/. [Accessed: 26-Feb-2014].
[69] A. Isaac, R. Clayphan, and B. Haslhofer, "Europeana: Moving to linked open data," *Information Standards Quarterly*, vol. 24, no. 2/3, 2012.
[70] Open Library, "Welcome to Open Library," 2014. [Online]. Available: https://openlibrary.org/. [Accessed: 24-May-2014].
[71] Michael Hausenblas, "5 star Open Data," Apr-2012. [Online]. Available: http://5stardata.info/. [Accessed: 11-Jun-2015].
[72] W3C RDF Working Group, "RDF 1.1 Primer," 25-Feb-2014. [Online]. Available: http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140225/. [Accessed: 17-Jun-2014].
[73] Marcelo Arenas, "RDF and SPARQL: Two basic components of the Semantic Web," Universidad de Valladolid, Spain, Apr-2013.
[74] Richard Cyganiak, David Wood, and Markus Lanthaler, "RDF 1.1 Concepts and Abstract Syntax - Section 5: Datatypes." 25-Feb-2014.
[75] Tim Berners-Lee, "Giant Global Graph," 21-Nov-2007. [Online]. Available: http://dig.csail.mit.edu/breadcrumbs/node/215. [Accessed: 11-Jun-2015].
[76] Dan Brickley and Ramanathan V. Guha, Eds., "RDF Schema 1.1." 25-Feb-2014.
[77] Mike Dean and Guus Schreiber, Eds., "OWL Web Ontology Language Reference." 10-Feb-2004.
[78] D. Allemang and J. Hendler, *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*. Elsevier Science, 2011.
[79] Fabien Gandon and Guus Schreiber, Eds., "RDF 1.1 XML Syntax." 25-Feb-2014.
[80] David Beckett, Tim Berners-Lee, Eric Prud'hommeaux, and Gavin Carothers, "RDF 1.1 Turtle." 25-Feb-2014.
[81] David Beckett, "RDF 1.1 N-Triples." 25-Feb-2014.
[82] Ivan Herman, Ben Adida, Manu Sporny, and Mark Birbeck, Eds., "RDFa 1.1 Primer - Third Edition." 17-Mar-2015.

[83] Manu Sporny, Dave Longley, Gregg Kellogg, Markus Lanthaler, and Niklas Lindström, "JSON-LD 1.0." 16-Jan-2014.
[84] W3C SPARQL Working Group, "SPARQL 1.1 Overview," 21-Mar-2013. [Online]. Available: http://www.w3.org/TR/2013/REC-sparql11-overview-20130321/. [Accessed: 17-Jun-2014].
[85] Paul Gearon, Alexandre Passant, and Axel Polleres, Eds., "SPARQL 1.1 Update." 21-Mar-2013.
[86] Steve Harris and Andy Seaborne, Eds., "SPARQL 1.1 Query Language." 21-Mar-2013.
[87] Chief Technology Officer Council, "Designing URI sets for the UK public sector - Publications - GOV.UK," 05-Oct-2010. [Online]. Available: https://www.gov.uk/government/publications/designing-uri-sets-for-the-uk-public-sector. [Accessed: 17-Jun-2015].
[88] Richard Cyganiak, "Creating an RDF vocabulary: Lessons learned," 07-Mar-2011. [Online]. Available: http://richard.cyganiak.de/blog/2011/03/creating-an-rdf-vocabulary/. [Accessed: 17-Jun-2015].
[89] R. Cyganiak, H. Stenzhorn, R. Delbru, S. Decker, and G. Tummarello, *Semantic sitemaps: Efficient and flexible access to datasets on the semantic web*. Springer, 2008.
[90] K. Alexander and M. Hausenblas, "Describing linked datasets-on the design and usage of void, the'vocabulary of interlinked datasets," in *In Linked Data on the Web Workshop (LDOW 09), in conjunction with 18th International World Wide Web Conference (WWW 09*, 2009.
[91] J. Volz, C. Bizer, M. Gaedke, and G. Kobilarov, "Silk-A Link Discovery Framework for the Web of Data.," *LDOW*, vol. 538, 2009.
[92] A.-C. N. Ngomo and S. Auer, "Limes-a time-efficient approach for large-scale link discovery on the web of data," *integration*, vol. 15, p. 3, 2011.
[93] EUCLID project, "Module 5 - Building Linked Data Applications (http://www.slideshare.net/EUCLIDproject/building-linked-data-applications-27768679)," 2013.
[94] C. Bizer and A. Schultz, "The R2R Framework: Publishing and Discovering Mappings on the Web," in *Proceedings of the First International Workshop on Consuming Linked Data, Shanghai, China, November 8, 2010*, 2010.
[95] C. Bizer and A. Schultz, *The berlin sparql benchmark*. 2009.
[96] RDF Web Applications Working Group, "RDF API." 11-May-2011.
[97] D. Lathrop and L. Ruma, *Open government: Collaboration, transparency, and participation in practice*. O'Reilly Media, Inc., 2010.
[98] C. P. Geiger and J. von Lucke, "Open Government and (Linked)(Open)(Government)(Data).," *eJournal of eDemocracy & Open Government*, vol. 4, no. 2, 2012.
[99] B. Hyland and D. Wood, "The Joy of Data - A Cookbook for Publishing Linked Government Data on the Web," in *Linking Government Data*, D. Wood, Ed. Springer New York, 2011, pp. 3–26.
[100] Web Foundation, "Open Data Barometer—2013 Global Report," 2013.
[101] D. Dell'Aglio, I. Celino, and E. Della Valle, "Urban mashups," in *Semantic Mashups*, Springer, 2013, pp. 287–319.
[102] D. DiFranzo, A. Graves, J. S. Erickson, L. Ding, J. Michaelis, T. Lebo, E. Patton, G. T. Williams, X. Li, J. G. Zheng, J. Flores, D. L. McGuinness, and J. Hendler, "The Web is My Back-end: Creating Mashups with Linked Open Government Data," in *Linking Government Data*, D. Wood, Ed. Springer New York, 2011, pp. 205–219.

[103] A. Caragliu, C. Del Bo, and P. Nijkamp, *Smart cities in Europe. Research Memoranda Series 0048 (VU University Amsterdam, Faculty of Economics, Business Administration and Econometrics)*. CRC Press, Boca Raton, 2009.

[104] S. Vert and R. Vasiu, "Relevant Aspects for the Integration of Linked Data in Mobile Augmented Reality Applications for Tourism," in *Information and Software Technologies*, vol. 465, G. Dregvaite and R. Damasevicius, Eds. Springer International Publishing, 2014, pp. 334–345.

[105] B. Villazón-Terrazas, L. M. Vilches-Blázquez, O. Corcho, and A. Gómez-Pérez, "Methodological guidelines for publishing government linked data," in *Linking Government Data*, Springer, 2011, pp. 27–49.

[106] A. Gonzalez, B. Villazon-Terrazas, and J. M. Gomez, "A Linked Data Lifecycle for Smart Cities in Spain," in *Proceedings of the Fifth Workshop on Semantics for Smarter Cities a Workshop at the 13th International Semantic Web Conference (ISWC 2014)*, Riva del Garda, Italy, 2014, vol. 1280, pp. 9–14.

[107] P. Vassiliadis, "A survey of Extract–transform–Load technology," *International Journal of Data Warehousing and Mining (IJDWM)*, vol. 5, no. 3, pp. 1–27, 2009.

[108] J.-P. Calbimonte, H. Y. Jeung, Ó. Corcho, and K. Aberer, "Enabling query technologies for the semantic sensor web," *International Journal on Semantic Web and Information Systems*, vol. 8, no. EPFL-ARTICLE-183971, pp. 43–63, 2012.

[109] L. Ding, T. Lebo, J. S. Erickson, D. DiFranzo, G. T. Williams, X. Li, J. Michaelis, A. Graves, J. G. Zheng, Z. Shangguan, J. Flores, D. L. McGuinness, and J. A. Hendler, "{TWC} LOGD: A portal for linked open government data ecosystems," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 9, no. 3, pp. 325 – 333, 2011.

[110] N. Shadbolt, K. O'Hara, T. Berners-Lee, N. Gibbins, H. Glaser, W. Hall, and others, "Linked open government data: Lessons from data. gov. uk," *IEEE Intelligent Systems*, vol. 27, no. 3, pp. 16–24, 2012.

[111] L. Ding, V. Peristeras, and M. Hausenblas, "Linked Open Government Data [Guest editors' introduction]," *Intelligent Systems, IEEE*, vol. 27, no. 3, pp. 11–15, 2012.

[112] I. Ermilov, C. Stadler, M. Martin, and S. Auer, "Towards an Open-Governmental Data Web," 2012.

[113] B. Villazón-Terrazas, L. M. Vilches-Blázquez, O. Corcho, and A. Gómez-Pérez, "Methodological Guidelines for Publishing Government Linked Data," in *Linking Government Data*, D. Wood, Ed. Springer New York, 2011, pp. 27–49.

[114] N. Veltman, "Mapping the History of Street Names." [Online]. Available: https://source.opennews.org/en-US/articles/mapping-history-street-names/. [Accessed: 01-Sep-2014].

[115] N. Veltman, "veltman/streets," *GitHub*. [Online]. Available: https://github.com/veltman/streets. [Accessed: 01-Sep-2014].

[116] S. Vert, "Linked Open Government Data for Smart City Applications," presented at the International Conference Social Media in Academia: Research and Teaching (SMART 2014), Timisoara, Romania, 2014.

[117] S. Vert and D. Andone, "Open Educational Resources in the Context of the Linked Data Web," in *Proceedings of the 10th International Scientific Conference eLearning and software for Education*, Bucharest, Romania, 2014, vol. 1, pp. 304–310.

[118] N. Komninos, H. Schaffers, and M. Pallot, "Developing a policy roadmap for smart cities and the future internet," in *eChallenges e-2011 Conference Proceedings, IIMC International Information Management Corporation*, 2011.

[119] U. Gretzel, R. Law, and M. Fuchs, *Information and Communication Technologies in Tourism 2010: Proceedings of the International Conference in Lugano, Switzerland, February 10-12, 2010*. Springer, 2010.

[120] F. Fritz, A. Susperregui, and M. Linaza, "Enhancing cultural tourism experiences with augmented reality technologies," in *The 6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST*, 2005, pp. 1–6.

[121] M. F. Tutunea, "Augmented Reality - State of Knowledge, Use and Experimentation," *USV Annals of Economics & Public Administration*, vol. 13, no. 2, 2013.

[122] T. Jung and D.-I. Han, "Augmented Reality (AR) in Urban Heritage Tourism," *e-Review of Tourism Research*, 2014.

[123] S. Feiner, B. MacIntyre, T. Höllerer, and A. Webster, "A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment," *Personal Technologies*, vol. 1, no. 4, pp. 208–217, 1997.

[124] T. Höllerer, S. Feiner, T. Terauchi, G. Rashid, and D. Hallaway, "Exploring MARS: developing indoor and outdoor user interfaces to a mobile augmented reality system," *Computers & Graphics*, vol. 23, no. 6, pp. 779–785, 1999.

[125] D. Gray, I. Kozintsev, Y. Wu, and H. Haussecker, "WikiReality: augmenting reality with community driven websites," in *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on*, 2009, pp. 1290–1293.

[126] G. Takacs, M. El Choubassi, Y. Wu, and I. Kozintsev, "3D mobile augmented reality in urban scenes," in *Multimedia and Expo (ICME), 2011 IEEE International Conference on*, 2011, pp. 1–4.

[127] Z. Yovcheva, D. Buhalis, and C. Gatzidis, "Overview of Smartphone Augmented Reality Applications for Tourism.," *E-review of Tourism Research*, vol. 10, no. 2, 2012.

[128] D. Boyer and J. Marcus, "Implementing mobile augmented reality applications for cultural institutions," *Museums and the Web, Philadelphia, USA*, 2011.

[129] "Streetmuseum: Q&A with Museum of London." [Online]. Available: http://blog.variousbits.net/2010/06/01/streetmuseum-qa-with-vicky-lee-museum-of-london/. [Accessed: 23-Apr-2014].

[130] E. Hutchings, "Time Travel Through Paris With Augmented Reality App," *PSFK*, 05-Jul-2013. [Online]. Available: http://www.psfk.com/2013/07/paris-travel-augmented-reality-app.html. [Accessed: 23-Apr-2014].

[131] B.-K. Seo, K. Kim, J. Park, and J.-I. Park, "A tracking framework for augmented reality tours on cultural heritage sites," in *Proceedings of the 9th ACM SIGGRAPH Conference on Virtual-Reality Continuum and its Applications in Industry*, 2010, pp. 169–174.

[132] L. Dorrzapf, N. Kratz, and M. Schrenk, "LIMES App - Mobile applications as an opportunity for cultural tourism along the Roman Limes in Europe," presented at the International Conference on Cultural Heritage and New Technologies, Vienna, 2012.

[133] M. T. Linaza, D. Marimón, P. Carrasco, R. Álvarez, J. Montesa, S. R. Aguilar, and G. Diez, "Evaluation of mobile augmented reality applications for tourism destinations," in *Information and Communication Technologies in Tourism 2012*, Springer, 2012, pp. 260–271.

[134] D. Marimon, C. Sarasua, P. Carrasco, R. Álvarez, J. Montesa, T. Adamek, I. Romero, M. Ortega, and P. Gascó, "MobiAR: Tourist Experiences through Mobile Augmented Reality," *Telefonica Research and Development, Barcelona, Spain*, 2010.

[135] B. Butchart, "Augmented reality for smartphones," 2011.

[136] C. D. Kounavis, A. E. Kasimati, and E. D. Zamani, "Enhancing the Tourism Experience through Mobile Augmented Reality: Challenges and Prospects.," *International Journal of Engineering Business Management*, vol. 4, 2012.

[137] C. Perey, T. Engelke, and C. Reed, "Current Status of Standards for Augmented Reality," in *Recent Trends of Mobile Collaborative Augmented Reality Systems*, L. Alem and W. Huang, Eds. Springer New York, 2011, pp. 21–38.

[138] C. Emmanouilidis, R.-A. Koutsiamanis, and A. Tasidou, "Mobile guides: Taxonomy of architectures, context awareness, technologies and applications," *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 103–125, 2013.

[139] V. Reynolds, M. Hausenblas, A. Polleres, M. Hauswirth, and V. Hegde, "Exploiting linked open data for mobile augmented reality," in *W3C Workshop: Augmented Reality on the Web*, 2010, p. 12.

[140] D. Schmalstieg and G. Reitmayr, "The World as a User Interface: Augmented Reality for Ubiquitous Computing," in *Location Based Services and TeleCartography*, 2005.

[141] M. L. Wilson, A. Russell, D. A. Smith, A. Owens, and others, "mspace mobile: A mobile application for the semantic web," 2005.

[142] C. Becker and C. Bizer, "Exploring the geospatial semantic web with dbpedia mobile," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 7, no. 4, pp. 278–286, 2009.

[143] V. Hegde, V. Reynolds, J. X. Parreira, and M. Hauswirth, "Utililising Linked Data for Personalized Recommendation of POI's," presented at the International AR Standards Meeting, Barcelona, Spain, 2011.

[144] L. Nixon, J. Grubert, and G. Reitmayr, "SmartReality: Augmented Reality + Services + Semantics," presented at the International AR Standards Meeting, Barcelona, Spain, 2011.

[145] L. J. Nixon, J. Grubert, G. Reitmayr, and J. Scicluna, "SmartReality: Integrating the Web into Augmented Reality.," in *I-SEMANTICS (Posters & Demos)*, 2012, pp. 48–54.

[146] B. Aydin, J. Gensel, P. Genoud, S. Calabretto, and B. Tellez, "Extending Augmented Reality Mobile Application with Structured Knowledge from the LOD Cloud," in *3rd International Workshop on Information Management for Mobile Applications*, 2013, vol. 13.

[147] B. Aydin, J. Gensel, P. Genoud, S. Calabretto, and B. Tellez, "An architecture for surroundings discovery by linking 3D models and LOD cloud," in *Proceedings of the Second ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems*, 2013, pp. 9–16.

[148] P. Grillo, S. Likavec, and I. Lombardi, "Using mobile phone cameras to interact with ontological data," in *Computer Aided Systems Theory–EUROCAST 2011*, Springer, 2012, pp. 568–576.

[149] C. Van Aart, B. Wielinga, and W. R. Van Hage, "Mobile cultural heritage guide: location-aware semantic search," in *Knowledge Engineering and Management by the Masses*, Springer, 2010, pp. 257–271.

[150] T. Matuszka and A. Kiss, "Alive Cemeteries with Augmented Reality and Semantic Web Technologies," *International Journal of Computer, Information Science and Engineering*, vol. 8, no. 2, pp. 32 – 36, 2014.

[151] R. A. Fouad, N. Badr, H. Talha, and M. Hashem, "On Location-Centric Semantic Information Retrieval in Ubiquitous Computing Environments.," *International Journal of Electrical & Computer Sciences*, vol. 10, no. 6, 2010.

[152] D. Martín-Serrano, R. Hervás, and J. Bravo, "Telemaco: Context-aware System for Tourism Guiding based on Web 3.0 Technology," in *1st Workshop on Contextual Computing and Ambient Intelligence in Tourism, Riviera Maya, Mexico*, 2011.

[153] M. Hausenblas, "Exploiting Linked Data to Build Web Applications," *IEEE Internet Computing*, vol. 13, no. 4, pp. 68–73, Jul. 2009.

[154] P. Shvaiko and J. Euzenat, "Ontology matching: state of the art and future challenges," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 25, no. 1, pp. 158–176, 2013.

[155] M. Emaldi, O. Pena, J. Lazaro, S. Vanhecke, E. Mannens, D. Lopez-de-Ipina, and others, "To Trust, or not to Trust: Highlighting the Need for Data Provenance in Mobile Apps for Smart Cities," in *Proceedings of the 3rd International Workshop on Information Management for Mobile Applications*, Riva del Garda, Italy, 2013, vol. 15, pp. 68–71.

[156] T. De Nies, S. Coppens, R. Verborgh, M. Vander Sande, E. Mannens, R. Van de Walle, D. Michaelides, and L. Moreau, "Easy Access to Provenance: an Essential Step Towards Trust on the Web," in *Computer Software and Applications Conference Workshops (COMPSACW), 2013 IEEE 37th Annual*, 2013, pp. 218–223.

[157] O. Hugues, J.-M. Cieutat, and P. Guitton, "Gis and augmented reality: State of the art and issues," in *Handbook of Augmented Reality*, Springer, 2011, pp. 721–740.

[158] C. R. Rivero, A. Schultz, C. Bizer, and D. Ruiz, "Benchmarking the Performance of Linked Data Translation Systems.," in *LDOW*, 2012.

[159] S. Massmann, S. Raunich, D. Aumüller, P. Arnold, and E. Rahm, "Evolution of the coma match system," *Ontology Matching*, p. 49, 2011.

[160] Q. Ji, P. Haase, G. Qi, P. Hitzler, and S. Stadtmüller, "RaDON—repair and diagnosis in ontology networks," in *The semantic web: research and applications*, Springer, 2009, pp. 863–867.

[161] R. Karam, F. Favetta, R. Laurini, and R. K. Chamoun, "Uncertain geoinformation representation and reasoning: A use case in lbs integration," in *Database and Expert Systems Applications (DEXA), 2010 Workshop on*, 2010, pp. 313–317.

[162] R. Karam, F. Favetta, R. Kilany, and R. Laurini, "Integration of similar location based services proposed by several providers," in *Networked Digital Technologies*, Springer, 2010, pp. 136–144.

[163] B. Martins, "A supervised machine learning approach for duplicate detection over gazetteer records," in *GeoSpatial Semantics*, Springer, 2011, pp. 34–51.

[164] S. Hahmann and D. Burghardt, "Connecting linkedgeodata and geonames in the spatial semantic web," in *6th International GIScience Conference*, 2010.

[165] R. Parundekar, C. A. Knoblock, and J. L. Ambite, "Linking and building ontologies of linked data," in *The Semantic Web–ISWC 2010*, Springer, 2010, pp. 598–614.

[166] R. Parundekar, C. A. Knoblock, and J. L. Ambite, "Aligning ontologies of geospatial linked data," in *Workshop On Linked Spatiotemporal Data, in*

conjunction with the 6th International Conference on Geographic Information Science (GIScience 2010). Zurich, 14th September.(forthcoming 2010), 2010.

[167] R. Parundekar, J. L. Ambite, and C. A. Knoblock, "Aligning unions of concepts in ontologies of geospatial linked data," in *Proceedings of the Terra Cognita 2011 Workshop in Conjunction with the 10th International Semantic Web Conference. Bonn, Germany*, 2011.

[168] J. Salas and A. Harth, "Finding spatial equivalences accross multiple RDF datasets," in *Proceedings of the Terra Cognita Workshop on Foundations, Technologies and Applications of the Geospatial Web*, 2011, pp. 114–126.

[169] R. Battle and D. Kolas, "GeoSPARQL: Enabling a Geospatial Semantic Web," *Semantic Web Journal*, vol. 3, no. 4, pp. 355–370, 2011.

[170] GeoKnow Consortium, "Deliverable 2.1.1 - Market and research overview." GeoKnow, 2013.

[171] G. A. Atemezing and R. Troncy, "Comparing vocabularies for representing geographical features and their geometry," in *Terra Cognita 2012 Workshop*, 2012, p. 3.

[172] T. Matuszka and A. Kiss, "Geodint: Towards Semantic Web-Based Geographic Data Integration," in *Intelligent Information and Database Systems*, Springer, 2014, pp. 191–200.

[173] W3C RDF Core Working Group, "Resource Description Framework (RDF): Concepts and Abstract Data Model," 2002. [Online]. Available: http://www.w3.org/TR/2002/WD-rdf-concepts-20020829/#xtocid48014. [Accessed: 23-Apr-2014].

[174] A. Flemming and O. Hartig, "Quality Criteria for Linked Data sources," 2010. [Online]. Available: http://sourceforge.net/apps/mediawiki/trdf/index.php?title=Quality_Criteria_for_Linked_Data_sources. [Accessed: 22-Apr-2014].

[175] P. Jain, P. Hitzler, P. Z. Yeh, K. Verma, and A. P. Sheth, "Linked Data Is Merely More Data.," in *AAAI Spring Symposium: linked data meets artificial intelligence*, 2010.

[176] C. Fürber and M. Hepp, "Using SPARQL and SPIN for data quality management on the Semantic Web," in *Business Information Systems*, 2010, pp. 35–46.

[177] C. Fürber and M. Hepp, "Using semantic web resources for data quality management," in *Knowledge Engineering and Management by the Masses*, Springer, 2010, pp. 211–225.

[178] C. Fürber and M. Hepp, "Towards a vocabulary for data quality management in semantic web architectures," in *Proceedings of the 1st International Workshop on Linked Web Data Management*, 2011, pp. 1–8.

[179] M. F. Goodchild, "Citizens as sensors: the world of volunteered geography," *GeoJournal*, vol. 69, no. 4, pp. 211–221, 2007.

[180] D. Zielstra and A. Zipf, "A comparative study of proprietary geodata and volunteered geographic information for Germany," in *13th AGILE international conference on geographic information science*, 2010, vol. 2010.

[181] M. Haklay, "How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets," *Environment and Planning B: Planning and Design*, vol. 37, no. 4, pp. 682 – 703, 2010.

[182] M. (Muki) Haklay, S. Basiouka, V. Antoniou, and A. Ather, "How Many Volunteers Does it Take to Map an Area Well? The Validity of Linus' Law to

Volunteered Geographic Information," *The Cartographic Journal*, vol. 47, no. 4, pp. 315–322, Nov. 2010.

[183] I. Celino, S. Contessa, M. Corubolo, D. Dell'Aglio, E. Della Valle, S. Fumeo, and T. Krüger, "Linking smart cities datasets with human computation–the case of urbanmatch," in *The Semantic Web–ISWC 2012*, Springer, 2012, pp. 34–49.

[184] Y. Gil and D. Artz, "Towards content trust of web resources," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 5, no. 4, pp. 227–239, 2007.

[185] O. Hartig, "Provenance Information in the Web of Data.," in *LDOW*, 2009.

[186] L. Moreau, B. Clifford, J. Freire, J. Futrelle, Y. Gil, P. Groth, N. Kwasnikowska, S. Miles, P. Missier, J. Myers, and others, "The open provenance model core specification (v1. 1)," *Future Generation Computer Systems*, vol. 27, no. 6, pp. 743–756, 2011.

[187] D. Garijo, B. Villazón-Terrazas, and O. Corcho, "A provenance-aware linked data application for trip management and organization," in *Proceedings of the 7th International Conference on Semantic Systems*, 2011, pp. 224–226.

[188] T. Berners-Lee, "Consistent User Interface, Section - The 'Oh yeah?' button," 1997. [Online]. Available: http://www.w3.org/DesignIssues/UI.html. [Accessed: 23-Apr-2014].

[189] T. De Nies, S. Coppens, E. Mannens, and R. Van de Walle, "Modeling uncertain provenance and provenance of uncertainty in W3C PROV," in *Proceedings of the 22nd international conference on World Wide Web companion*, 2013, pp. 167–168.

[190] S. Vert and R. Vasiu, "Integrating Linked Data in Mobile Augmented Reality Applications," in *Information and Software Technologies*, vol. 465, G. Dregvaite and R. Damasevicius, Eds. Springer International Publishing, 2014, pp. 324–333.

[191] C. L. Franco, "CONCERT: A New Framework for Contextual Computing in Tourism to Support Human Mobility," UNIVERSIDAD DE DEUSTO (SPAIN), 2010.

[192] A.-S. Dadzie and M. Rowe, "Approaches to visualising linked data: A survey," *Semantic Web*, vol. 2, no. 2, pp. 89–124, 2011.

[193] S. Vert and R. Vasiu, "Integrating Linked Open Data in Mobile Augmented Reality Applications - a Case Study," *TEM JOURNAL - Technology, Education, Management, Informatics*, vol. 4, no. 1, Feb. 2015.

[194] Open Knowledge, "Conformant Licenses - Open Definition." [Online]. Available: http://opendefinition.org/licenses/. [Accessed: 14-Jan-2015].

[195] D. Huynh and S. Mazzocchi, *OpenRefine*. 2012.

[196] A. Schultz, A. Matteini, R. Isele, P. N. Mendes, C. Bizer, and C. Becker, "LDIF-A Framework for Large-Scale Linked Data Integration," in *21st International World Wide Web Conference (WWW 2012), Developers Track, Lyon, France*, 2012.

[197] F. M. Suchanek, G. Kasneci, and G. Weikum, "Yago: a core of semantic knowledge," in *Proceedings of the 16th international conference on World Wide Web*, 2007, pp. 697–706.

[198] H. Gonzalez, A. Halevy, C. S. Jensen, A. Langen, J. Madhavan, R. Shapley, and W. Shen, "Google fusion tables: data management, integration and collaboration in the cloud," in *Proceedings of the 1st ACM symposium on Cloud computing*, 2010, pp. 175–180.

[199] P. N. Mendes, H. Mühleisen, and C. Bizer, "Sieve: Linked Data Quality Assessment and Fusion," in *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, New York, NY, USA, 2012, pp. 116–123.
[200] S. Vert, B. Dragulescu, and R. Vasiu, "LOD4AR: Exploring Linked Open Data with a Mobile Augmented Reality Web Application," in *Proceedings of the ISWC 2014 Posters & Demonstrations Track, within the 13th International Semantic Web Conference (ISWC 2014)*, Riva del Garda, Italy, 2014, vol. 1272, pp. 185–188.
[201] Silviu Vert and BuildAR, "Awe.js with geolocation · Issue #3 · buildar/awe.js · GitHub," Jun-2014. [Online]. Available: https://github.com/buildar/awe.js/issues/3. [Accessed: 26-Jun-2015].
[202] J. W. Harris and H. Stöcker, *Handbook of mathematics and computational science*. Springer Science & Business Media, 1998.
[203] F. Naumann, "Data profiling revisited," *ACM SIGMOD Record*, vol. 42, no. 4, pp. 40–49, 2014.
[204] A. Jentzsch, "Profiling the Web of Data," *ISWC-DC 2014 Doctoral Consortium at ISWC 2014*, p. 32.
[205] A. Langegger and W. Woss, "Rdfstats-an extensible rdf statistics generator and library," in *Database and Expert Systems Application, 2009. DEXA'09. 20th International Workshop on*, 2009, pp. 79–83.
[206] A. Langegger, W. Woss, and M. Blöchl, *A semantic web middleware for virtual data integration on the web*. Springer, 2008.
[207] S. Khatchadourian and M. P. Consens, "Explod: Summary-based exploration of interlinking and rdf usage in the linked open data cloud," in *The Semantic Web: Research and Applications*, Springer, 2010, pp. 272–287.
[208] S. Auer, J. Demter, M. Martin, and J. Lehmann, "LODStats–an extensible framework for high-performance dataset analytics," in *Knowledge Engineering and Knowledge Management*, Springer, 2012, pp. 353–362.
[209] Z. Abedjan, T. Gruetze, A. Jentzsch, and F. Naumann, "Profiling and mining RDF data with ProLOD++," in *2014 IEEE 30th International Conference on Data Engineering (ICDE)*, 2014, pp. 1198–1201.
[210] C. Bohm, F. Naumann, Z. Abedjan, D. Fenz, T. Grutze, D. Hefenbrock, M. Pohl, and D. Sonnabend, "Profiling linked open data with ProLOD," in *Data Engineering Workshops (ICDEW), 2010 IEEE 26th International Conference on*, 2010, pp. 175–178.
[211] P. Ristoski, C. Bizer, and H. Paulheim, "Mining the web of linked data with rapidminer," in *International Semantic Web Conference (ISWC)*, 2014.
[212] B. Fetahu, S. Dietze, B. Pereira Nunes, M. Antonio Casanova, D. Taibi, and W. Nejdl, "What's all the data about?: creating structured profiles of linked data on the web," in *Proceedings of the companion publication of the 23rd international conference on World wide web companion*, 2014, pp. 261–262.
[213] Business Data Quality Ltd, "Data Profiling Best Practices." 2010.
[214] Wikipedia, "Notability," *Wikipedia, the free encyclopedia*. 06-May-2015.
[215] L. J. B. Nixon, J. Grubert, G. Reitmayr, and J. Scicluna, "SmartReality: Integrating the Web into Augmented Reality," presented at the I-SEMANTICS (Posters & Demos), 2012, pp. 48–54.
[216] C. Bizer, T. Heath, K. Idehen, and T. Berners-Lee, "Linked data on the web (LDOW2008)," in *Proceedings of the 17th international conference on World Wide Web*, 2008, pp. 1265–1266.

APPENDIX

Summary of permission for third party copyright works

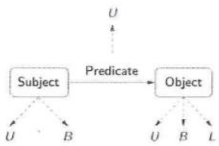
Figure Number	**Type of work**	**Source**	**Copyright Holder & Year**	**Permission to reproduce**
Figure 3-3	Figure	M. Arenas, "RDF and SPARQL: Two basic components of the Web of data" (presentation, slide #5)	© 2013 Marcelo Arenas	Included in this appendix
Figure 3-4	Figure	T. Heath and C. Bizer, "Linked Data: Evolving the Web into a Global Data Space," Synthesis Lectures on the Semantic Web: Theory and Technology, vol. 1, no. 1, Figure 5.1	© 2011 Morgan & Claypool Publishers	Included in this appendix
Figure 5-2	Screenshot	Google Fusion Tables	© 2015 Google - Map data © 2015 Google	According to http://www.google.com/permissions/geoguidelines.html
Figure 6-1	Screenshot	Google Fusion Tables	© 2015 Google - Map data © 2015 Google	According to http://www.google.com/permissions/geoguidelines.html

6 September 2015

Dear Professor Marcelo Arenas,

I am completing my PhD thesis at Politehnica University of Timisoara (Romania) entitled 'Integrating Linked Open Data in Mobile Augmented Reality Applications'.

I seek your permission to reprint, in my thesis, an image from http://web.ing.puc.cl/~marenas/talks/edi-07.pdf (slide #5). The image to be reproduced is displayed below:



I would like to include the image in my thesis, with proper attribution, of course. The thesis will be printed by my university's publishing house and will be distributed to academic libraries for non-commercial purposes.

If you are happy to grant me the permissions requested, please return a signed copy of this letter. If you wish to grant only some of the permissions requested, please list these and then sign.

Yours sincerely,

Silviu Vert

Permission granted for the use requested above:

I confirm that I am the copyright holder of the image above and hereby give permission to include it in your thesis which will be printed for non-commercial purposes.

[please edit the text above if you wish to grant more specific permission]

Signed:

Name: Marcelo Arenas

Organisation: Pontificia Universidad Católica de Chile

Job title: Associate Professor

Subject **RE: permission to re-use copyrighted material**
From Jovan Carreon <carreon@morganclaypool.com>
To <silviu.vert@cm.upt.ro>
Date 2015-09-09 02:32

Dear Silviu Vert:

My name is Jovan Carreon. I am the editorial assistant at Morgan & Claypool and I can assist you with the request to reuse material from one of our books.

Thank you for your email. After reviewing the details of your planned use of figure 5.1 from *Linked Data,* I`m pleased to grant you permission to reuse the material in your thesis, *Integrating Linked Open Data in Mobile Augmented Reality Applications.* Please treat this email as an official approval from Morgan & Claypool Publishers.

Best of luck with your thesis,
Jovan

--
Jovan Carreon
Editorial Assistant
Morgan & Claypool Publishers
1210 Fifth Ave. Ste. 250
San Rafael, CA 94901
carreon@morganclaypool.com
(415) 785-8003

From: silviu.vert@cm.upt.ro [mailto:silviu.vert@cm.upt.ro]
Sent: Saturday, September 5, 2015 1:22 PM
To: info@morganclaypool.com
Subject: permission to re-use copyrighted material

Dear Sir or Madam,

I'm requesting permission to reuse copyrighted material from Morgan & Claypool Publishers (Synthesis). The required information is listed below.

===

Name: Silviu Vert
Title: Engineer, PhD Candidate
Institution: Politehnica University of Timisoara (Romania) www.upt.ro
E-mail address: silviu.vert@cm.upt.ro
Mailing address: Piata Victoriei No. 2, 300006 Timisoara, Romania
Phone number: +40 723 139869
Description of the content you wish to reuse: Figure
Title of book in which the material you would like to use appears: "Linked Data: Evolving the Web into a Global Data Space" (ISBN: 9781608454303)

Series: Synthesis Lectures on the Semantic Web: Theory and Technology
Authors name: Tom Heath, Christian Bizer
Specific figure number: Figure 5.1 Linked Data Publishing Options and Workflows, Link to illustration: http://linkeddatabook.com/editions/1.0/images/linkeddat_small.png)
Purpose of re-use: PhD Thesis
Title of new work in which MCP material will appear: Integrating Linked Open Data in Mobile Augmented Reality Applications
Title of chapter: Linked Data
Authors: Silviu Vert (coordinated by Prof. Radu Vasiu)
Expected publication date: September 2015
Publisher: Politehnica Publishing House (http://www.editurapolitehnica.upt.ro/)
Intended audience for the work: academia
Expected circulation: 50 copies
Retail price: 0

===

I look forward to a positive answer.

Yours sincerely,

Silviu Vert
PhD Candidate
Politehnica University of Timisoara
+40 723 139869