reviewed paper

Innovative Approaches to Urban Data Management using Emerging Technologies

Jens Dambruch, Andreas Stein, Veneta Ivanova

(Dipl. Inf. Jens Dambruch, Fraunhofer IGD, Fraunhoferstraße 5, 64283 Darmstadt, Germany, jens.dambruch@igd.fraunhofer.de)
(M. Sc. Wirt.-Inf. Andreas Stein, Fraunhofer IGD, Fraunhoferstraße 5, 64283 Darmstadt, Germany, andreas.stein@igd.fraunhofer.de)
(Veneta Ivanova, Fraunhofer IGD, Fraunhoferstraße 5, 64283 Darmstadt, Germany, veneta.ivanova@igd.fraunhofer.de)

1 ABSTRACT

Many characteristics of Smart cities rely on a sufficient quantity and quality of urban data. Local industry and developers can use this data for application development that improves life of all citizens. Therefore, the handling and usability of this data is a big challenge for smart cities. In this paper we investigate new approaches to urban data management using emerging technologies and give an insight on further research conducted within the EC-funded smarticipate project.

Geospatial data cannot be handled well in classical relational database environments. Either they are just put in as binary large objects or have to be broken down into elementary types which can be handled by the database, in many cases resulting in a slow system, since the database technology is not really tuned for delivery on mass data as classical relational databases are optimized for online transaction processing and not analytic processing.

Document-based databases provide a better performance, but still struggle with the challenge of large binary objects. Also the heterogeneity of data requires a lot of mapping and data cleansing, in some cases replication can't be avoided.

Another approach is to use Semantic Web technologies to enhance the data and build up relations and connections between entities. However, data formats such as RDF use a different approach and are not suitable for geospatial data leading to a lack on usability.

Search engines are a good example of web applications with a high usability. The users must be able to find the right data and get information of related or close matches. This allows information retrieval in an easy to use fashion. The same principles should be applied to geospatial data, which would improve the usability of open data. Combined with data mining and big data technologies those principles would improve the usability of open geospatial data and even lead to new ways to use it. By helping with the interpretation of data in a certain context data is transformed into useful information.

In this paper we analyse key features of open geodata portals such as linked data and machine learning in order to show ways of improving the user experience. Based on the Smarticipate projects we show afterwards as open data and geo data online and see the practical application. We also give an outlook on piloting cases where we want to evaluate, how the technologies presented in this paper can be combined to a usefull open data portal. In contrast to the previous EC-funded project urbanapi, where participative processes in smart cities where created with urban data, we go one step further with semantic web and open data. Thereby we achieve a more general approach on open data portals for spatial data and how to improve their usability.

The envisioned architecture of the smarticipate project relies on file based storage and a no-copy strategy, which means that data is mostly kept in its original format, a conversion to another format is only done if necessary (e.g. the current format has limitations on domain specific attributes or the user requests a specific format). A strictly functional approach and architecture is envisioned which allows a massively parallel execution and therefore is predestined to be deployed in a cloud environment.

The actual search interface uses a domain specific vocabulary which can be customised for special purposes or for users that consider their context and expertise, which should abstract from technology specific peculiarities.

Also application programmers will benefit form this architecture as linked data principles will be followed extensively. For example, the JSON and JSON-LD standards will be used, so that web developers can use results of the data store directly without the need for conversion. Also links to further information will be provided within the data, so that a drill down is possible for more details.

The remainder of this paper is structured as follows. After the introduction about open data and data in general we look at related work and existing open data portals. This leads to the main chapter about the key

ISBN 978-3-9504173-0-2 (CD), 978-3-9504173-1-9 (print)
Editors: Manfred SCHRENK, Vasily V. POPOVICH, Peter ZEILE, Pietro ELISEI, Clemens BEYER

technology aspects for an easy-to-use open data portal. This is followed by Chapter five, an introduction of the EC-funded project smarticipate, in which the key technology aspects of chapter four will be included.

2 INTRODUCTION

Publishing data on city portals as open data is a growing trend, which is also promoted by the European Commission. However, the prevalent experience when investigating such data is that it is just a more or less raw dump of data that was created sometimes as a by-product due to technical reasons. It's hard for laymen to interpret the data sets, yet even impossible to tell if the content is useable for anything as mostly no meta-information is given. Also keeping up-to-date can be hard under such circumstances. Especially when talking about data, that has not been issued by public authorities also the question of trust in data rises: who collected it and when? Is there some attempt to defraud? This is very important as building services on faulty data consequently leads to faulty results and this undermines the reputation of Open Data as a whole. Another important aspect is in which way should data be published? Many open data portals publish them in PDF, Excel or some image format. It's very much like the traditional data, which was published as printout. Geospatial 2D data such as zoning plans are often published as ESRI-shapefiles, which need special tools such as Q-GIS to work with them. There are already a lot of tools like Geoserver or CityServer3D, which can work with these data sets. However, the problem of linking into the data remains the same. The solution can be to reduce the format to atomic entities capable to represent the data to manage. The semantic web initiative proposes a whole technology stack, but it has not gained the momentum to innovate information systems as needed. In this context the term "5 star open data principle"[1] has been coined to set out the ultimate goal for data availability. While the levels 2 or 3 are quite easy to reach, the most useful layer 5 needs a lot more involvement. In our view it ultimately does boil down not to a question of standardised file formats but to interoperable dynamic information systems, considering the needs of users and developers first. Such a system has to deal with aspects of heterogeneity, distribution of data and systems, support for multiple formats and standards, whilst providing openness for innovation.

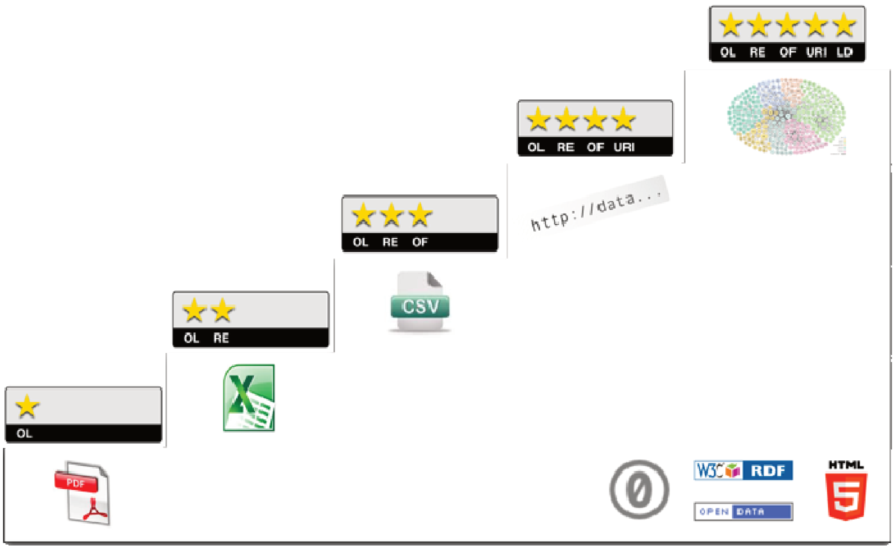

Fig. 1: Five-star open data principle as of Tim Berners-Lee.

3 RELATED WORK

3.1 The urbanAPI project

The urbanAPI project[2] innovated public participation through the use of interactive 3D-Web technology, as shown by Dambruch and Krämer (2014). Furthermore, a new way of interaction with data was proposed by Malewski, Dambruch and Krämer (2015) using Domain-Specific Languages. The biggest issue identified as obstacle to wide application of such systems was that there is some substantial effort required in pre-

[1] 5stardata.info
[2] www.urbanapi.eu

processing and annotating the data to be used. Most of these steps are not supported well by tools and gathering input for a particular use case had several steps to be done over and over again – with very little but important differences. The results of the pre-processing was not useable in other use cases as it was tied to the special needs of the use case. So the major observation was that the basically the same steps had to be carried out again and again, as the parameters changed. This is the motivation for the service platform in smarticipate.

3.2 The semantic web

The semantic web was originally conceived by Tim Berners-Lee (Berners-Lee 1998) as a universal standard for describing, encoding and annotating data along with formal ontology and interlinking data sets with other data. For this a lot of technologies for advanced information encoding and processing in a standardised way were proposed such as XML/RDF[3], RDFS, OWL[4], SPARQL[5].

Based on this, ontologies and vocabularies have been defined to model aspects such as a thesaurus via SKOS[6] or physical units, quantities and dimension as QUDT[7]. Also meta data in general has been described as the Dublin Core initiative[8] and many others also supported by government authorities and standardisation bodies. But several aspects needed by nature to remain vague as the relations of things to describe are not as clear as one might suspect from a data perspective. It turned out that it is hard to grasp the essence of a thing relevant in general, therefore it is also hard to encode it in a formal system. Certain aspects of a thing might only be relevant in certain contexts or use cases, others might also be only of historical interest. A taxonomy like SKOS therefore defines vague concepts of identity such as closeness or exact match which are quite hard to understand from a programmer's point of view, but are needed from the ontology-engineering perspective.

3.3 Existing open data portals

Schäfer (2016) gives an overview of existing open data portals and catalogue applications. A key observation is that most of them are not designed well and usability wasn't taken into account appropriately. Major concern is that there is no overview provided and exploration not supported as references between data items is not given, if at all, only references between datasets as a whole may be provided.

A necessary key feature of open data portals is a well performing search function. The user usually uses domain specific search terms which needs to be used correctly by the search engine. The availability of open data in existing portals is another key feature. If the users are able to find the data they are looking for the data must be available for download immediately, ideally without registration or verification of mail address or other individual data.

4 KEY TECHNOLOGY ASPECTS

Open data should be provided in an open and widely established format. Especially accessability on fact level without the need for proprietary tools is important. The web technology and especially the semantic web initiative provided mechanisms to do this (see 3.2). The typical user first needs to get an overview on data he needs for his purpose and which data is available at all, therefore the first step is exploration of data. Current experiences show that the immense amount of data available, the heterogenity of datasets and distribution of data as well as the lack of well maintained catalogues makes the exploration a tedious and time consuming process. This is especially true for open data as this data often is published „as is", e.g. a dump of results originating from arbitrary activity.

Today people expect search engines to be as easy as using google, but most open data and geo portals are miles away from this expectation. Finding data by simple keyword matching works in some cases, but to improve results and especially to find related data sets needs some more sophisticated approaches. Semantic relationships can already be modelled by semantic web technologies, e.g. RDF and ontologies, but this has

[3] https://www.w3.org/RDF/
[4] https://www.w3.org/2001/sw/wiki/OWL
[5] https://www.w3.org/2001/sw/wiki/SPARQL
[6] http://www.w3.org/2004/02/skos/vocabs
[7] http://www.qudt.org
[8] http://dublincore.org

ISBN 978-3-9504173-0-2 (CD), 978-3-9504173-1-9 (print)
Editors: Manfred SCHRENK, Vasily V. POPOVICH, Peter ZEILE, Pietro ELISEI, Clemens BEYER

not gained much prominence although being developed for more than 10 years. One big problem is that data has to be transformed in RDF, and therefore data has to be replicated in this particular format, which means maintenance of doubled data sets. Apart from costs regarding additional infrastructure mainly the efforts for dedicated personnel is a clear limit for such an approach.

Another interesting question is about the search results. The range of possible results is also mandated by the form of query that can be issued. E.g. from simple facts such as how many people live in the city to more complex answers such as what is the best cycling path to work or even complex 3D visualizations. A query is the first step in this by saying "what" I want to get and the second step is "how" I want to view it.

Data sets that do not adhere to a common structure with normed semantics are most likely using a vocabulary of their own. This necessitates a preprocessing step called harmonisation and mapping where common terms are identified and how the terms will map on terms defined in other data sets. The consequence of this approach is that a normal ontology is to be defined in which all terms can be mapped.

Ultimately such an approach leads to endeavours such as the insprire directive, mandated be the European Commission, which has the goal to define a common ontology for spatial infrastructure for all European countries. However, such an approach has severe drawbacks in terms of usability (Janowicz et al. 2013) as it tends to make projects and applications overly complex and draws attention away from the real problems users are facing.

Instead Janowicz et al. propose the concept of Microontologies where the conceptualisation of data is closely tied to the problem to solve, the domain of it and the application or service to be developed. The consequence for such an approach is that the mapping of vocabularies and ontologies has to be provided on a case by case basis and can build on smaller units. With this the solution of the problem can be defined in a more convenient way, as generalized ontologies need to define globally valid definitions, which is a hard task, if possible at all. So the vocabulary to use in the query considers the "what" or use case they are designed for and secondly a special vocabulary for visualizing the results can be defined as defining the "how".

The interesting part comes into play when these definitions can be changed: A new set of "what" and "how" for a completely new domain can view data in a completely different context. Based on this experts and laymen can consider observations of other people easily and switch perspectives.

As already said results need to be presented in a convenient way to users. For geospatial data this means to visualize it in a map in 2D or 3D along with some context. This context can be determined by the vocabulary/ontology used in the query considering different visual styling or highlighting important aspects.

The next important thing is to support data exploration at a more useful layer. Most classical GIS do just provide one layer of results and no further refinement. Integrating other data sources in a mash-up style provides new opportunities to drill down on results in a convenient "web surfing" style, which means to provide related datasets for some particular feature or search results.

4.1 Search technology

Search engines are a crucial part of today's web architecture. Finding the right data and getting information of related or close matches guide people in information retrieval in an easy to use fashion. The same principles should be applied to geospatial data. Combined with data mining or big data technologies this would improve the usability of open geospatial data and even lead to new ways to use it. By helping with the interpretation of data in a certain context data is transformed into useful information. As mentioned above users expect a search engine similar to google and therefore they expect search results in a comparable quality and time. To achieve this, state-of-the-art search technology has to be used. Fortunately, real-time data analytics and search engines have been improved rapidly over the past years. A combination of elastic search for the analytics engine, cabana for the visualization of data and logstash for data preparation is used by multinational companies (e.g. Netflix, Wikipedia, Facebook) that face similar problems.

4.2 Linked Data, JSON LD

Linked data is in principle already available via web technology and especially semantic web technology. However, it is not as simple as this. HTML and websites are ubiquitous nowadays and have gained a lot of success. But the sites are not interoperable and reusable as needed for data services. This issue is mostly addressed via the RDF encoding standards and related semantic web facilities. But the paradigm behind this seems to be not very tempting for most developers and users.

Manu Sporny points out a lot of concerns in his blog[9] about the origin of the JSON LD standard, especially if non elementary things are to be modelled. A very specific domain may use terms and concept in such a way, that it does not make sense to even discuss it a general level. Also Janowicz and Hitzler conclude that standardized meaning is a misconception in many cases: Mostly it is to be considered in a certain context or you also may say use case. Our conclusion regarding this is that such a technology needs to involve experienced ontology engineers on a local level to cope with the inherent complexity. Otherwise the problem of losing focus on practical application and problem solving and overloading of users with irrelevant details dismantles the reputation of such a system in general. Apart from this the RDF and OWL specifications are not easy to understand and use, as they apply paradigms not very familiar with web development. JSON LD on the other hand just extends the web programming model in a simple and pragmatic way, without breaking the minds of users or technology.

The semantic web has yet played a minor role for geo information and GIS but also geo information standards play a minor role in other communities. This is because of nature of standardisation bodies such as the Open Geospatial Consortium (OGC) as there are only GI experts designing for their own community (Hart and Dolbear, 2013, p. 40). Also the standardisation itself is mostly perceived as a barrier to modern development approaches as a top-down definition of governmental driven use cases such as the INSPIRE[10] directive tend to be long-lasting and tedious and off scope in many cases. Especially the goal to set normative semantics is totally against the ideas of the semantic web and the needs and goals of users and creators of geo information.

On the other hand, publishing data should be made easy, without the need for a lot of cleansing and preparation. If the data published has no special purpose than just publishing it, it is hard to say what people can make out of it. So the best approach for this seems to keep it as simple as possible and choose pragmatic approaches. For example, JSON LD as web affine encoding standard, which could be used right away with any modern browser.

4.3 Machine Learning

Data pre-processing and especially semantic annotation is a time consuming and tedious task. Especially maintaining meta data is a crucial point in many geospatial applications to make data usable. Therefore, such tasks should be automated as much as possible. A key technology which could help in this comes from the Artificial Intelligence domain and deals with machine learning. With this a computer can be trained to recognise patterns in data and extract information from datasets by applying sophisticated statistical methods. Thus, annotating data sets and finding relations can benefit from this to some extent. The data analysis can range from text mining and natural language processing to advanced image recognition of features in aerial images. Common patterns and correlations can be analysed and contribute to building assumptions or assumptions can be checked on real data. It is yet not clear to what extend the analysis can be automated, thus also aided operation or crowd sourcing can be used to verify results of such analysis.

4.4 Database Systems

Since open data is often published on distributed systems it is necessary to create an index referring to the available data. Since geospatial data is usually considered big data there are different advantages and disadvantages for the different types of databases.

SQL Databases

Even though new database concepts and systems were developed recently most data is still stored in SQL databases. This fact is not astonishing since it is false to assume SQL database lack the ability to scale. Facebook for example release Presto, a SQL Database that interacts with petabytes of data.[11] Since a system providing big geospatial data is a very heterogeneous database environment it is important how queries can

[9] http://manu.sporny.org/2014/json-ld-origins-2/

[10] EU Parliament Directive 2007/2/EC

[11] https://www.facebook.com/notes/facebook-engineering/presto-interacting-with-petabytes-of-data-at-facebook/10151786197628920

ISBN 978-3-9504173-0-2 (CD), 978-3-9504173-1-9 (print)
Editors: Manfred SCHRENK, Vasily V. POPOVICH, Peter ZEILE, Pietro ELISEI, Clemens BEYER

perform across different database systems. User queries might need information from different databases which requires a well performing read-operation.

No-SQL Databases

Currently there are many different No-SQL Databases. This is explained by Brewer's Theorem (CAP-theorem). According to this theorem a distributed system can only provide two out of three characteristics: consistency, availability and partition tolerance. Different No-SQL Databases focus on different trade-offs and the provider of data has to choose the best database solution for his data. Since No-SQL Databases have a better performance on average they are usually the first choice if there are no critical transactional or consistency requirements. However there are huge differences between the different systems.[12]

NewSQL Databases

NewSQL Databases are a new type of database system. They support relational data model and use SQL but they target systems that need a large number of transactions. Especially when providing open data this will be the case. Not only a large number of transactions but the transactions will target only a small subset of the available data which is a strength of NewSQL Databases.

Besides the type of database system, the provider has to evaluate what system will be the most beneficial for the requirements of his system. The most important requirements are the ability to search quickly and on multiple, distributed databases at the same time and a fast reading mechanism.

4.5 Cloud technology

Geospatial data and open data are often used in the context of big data. Therefor the usage of cloud technology is almost mandatory. To evaluate the importance of cloud technology we use the definition of the National Institute of Standards and Technology.[13] According to this definition there are five characteristics of a cloud computing system.

On-demand self-service

The computing abilities have to be scalable in real time to match the requirements of the environment. This is very important for a provider of open data since some events (e.g. Olympic Games) can trigger a rapid increase of open data usage.

Broad network access

The data has to be available for different client platforms (e.g. mobile phones, tablets, laptops and workstations). This is critical as well because most applications based on open data demand a noteworthy part of the available open data. Especially mobile applications depend on a fast transfer of the required data since they usually don't provide the necessary memory to store this data.

Resource pooling

The available resources of a cloud will be shared among the different users. This can be a critical characteristic since the user of a cloud has no information where the data is stored or where the data is processed. In case of the appropriation of open data this aspect accentuates the importance of cloud technology. Resource pooling is one of the main reason why cloud resources are low priced.

Rapid elasticity

Even an unexpected increase of user queries can be handled by the cloud computing system. Ideally the scaling of the used resources will be fully automatic. In this case the provider of open data does not have to be available at all times to scale the resources himself.

Measured service

Especially with an automatic system that handles unexpected increase of user queries it is important to have some kind of cost control. Typically this is done on a pay-per-use or charge-per-use basis. Even though this

---

[12] https://www.google.de/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0ahUKEwjwi__nktnLAhUHVBQKHY6UDfAQFgg7MAE&url=http%3A%2F%2Fwww.datastax.com%2Fwp-content%2Fuploads%2F2013%2F02%2FWP-Benchmarking-Top-NoSQL-Databases.pdf&usg=AFQjCNEdcXHWXW78i5J5CL4Frzk1VdykBw&sig2=bP33uyudgh-CSp72cl2_Hw&cad=rja

[13] http://dx.doi.org/10.6028/NIST.SP.800-146

is the least important characteristics of a cloud computing system for providing open data it is mandatory to avoid cost spikes.

4.6 Methods of interaction: Domain Specific Languages

Users that analyse geospatial data sets are confronted with a lot of GIS technology driven terminology and also computing terminology, which is mostly reflected in the language they could use for this (e.g. Python for ArcGIS). On the other hand, they naturally have a different vocabulary driven by the need of their working domain. Even each city or department can have different definitions of seemingly common terms. These deviations in definitions stem from the fact that as each organisation only needs to cover aspects they deal with and ignore others which are not relevant to their work. In many cases every department in an administration has different definitions for the same entity: Customer Relationship Management systems for example define addresses as immutable entities denoting a location. Billing systems often see addresses as a person and a location where a bill can be sent. In short, such definitions depend on the context they are used in and this is not a drawback since the focus is put on the aspects necessary. This is a common principle in IT systems, which can be paraphrased as: Do only one thing – but do it right. At second glance there are some implications when users need to understand another person's vocabulary and they encounter conflicts of definitions of terms. So a mapping between both domains is vital to get a common understanding.

For GIS expert users' analysis tasks on geospatial data are common and easy but they also need to deal with a lot of technology issues, which they need to resolve in order to get the results demanded. They are fluent with both the technology driven vocabulary (which uses terms like Feature, TerrainGrid, Layer, etc.) and the domain vocabulary (which may use terms like Street, Quarter, River, etc.) What would happen if we decouple the user vocabulary from the technology vocabulary by using semantic annotation and mapping? As a consequence of the mapping the domain vocabulary can be flexible and highly adapted to special circumstances. How we approach this is detailed in the following chapter. For useful interaction just a vocabulary is not sufficient – syntax for example is also required. Such a language is constructed based on the concept of Domain Specific Languages - DSL (Fowler 2010). Such DSLs have already been applied successfully within the urbanAPI project for data preparation and policy modelling (Krämer, Ludlow, Khan 2013). Most of this was done with classical desktop-GIS software (Krämer and Stein 2014). In UrbanAPI we did show how we extended the DSL concept to a web-based analysis tool which enables users to work in their language level without the need to have strong geodata and technology domain knowledge.

Our prototype is inspired by the paper of Janowicz and Hitzler (2013) and elaborates an alternative approach to using Semantic Web technologies and related generic ontologies, which is driven by the idea of simple-to-the-point models and not a normative semantic world model.

5 THE SMARTICIPATE PROJECT

The driver of the smarticipate project is the need to publish open data which is clearly understandable, serves a defined purpose and is trustable and authoritative. This data should enable people and business organisations to build up new public information services to participate and understand what is going on in their city.

The envisioned software is a data-rich citizen dialogue system, transforming public data into useful information. Tools will be developed and implemented to make data trustworthy, by using state-of-the-art web and GIS technology.

New services provided with smarticipate can put people and non- governmental organisations in charge of several tasks selected by and under the supervision of municipalities, thus transforming the administration processes and turning the administration into a partner. Within the municipal departments smarticipate will facilitate collaboration, as city administrations will work on the same data basis, which as well can be shared with business and citizens on the smarticipate platform. Data security and trust in data is responsibility of the city government and therefore is considered as legitimate.

5.1 Concept

smarticipate fosters a bottom-up approach of local governance. All citizens like entrepreneurs, students, workers, members of NGO's, estate owners and investors have the chance thanks to open government and access to open data to propose technically qualified ideas/initiatives on an expert/professional level.

Increased connectivity leads to new coalitions between different citizens and produces new, unexpected and innovative ideas for new public services. Initiative can be taken by the inhabitants of the city, such as city officials, entrepreneurs and regular citizens who have issues and ideas of how to solve them. The initiators create their own applications available for everyone using their interface. To achieve the concept described above, smarticipate has the following principles of development:

- Interdisciplinary – participatory stakeholder involvement, focussing on innovation processes
- iterative development – the observation that a planning process is evolutionary is to be taken into account when discussing with citizens and stakeholders.
- development for openness – existing frameworks and technologies will be reused, standards examined and applied where suitable and useful
- extensive piloting – the platform is developed in an open environment with direct feedback options for stakeholders

See also Fröhlich and Vogt (2016) for further details on engaging people.

5.2 Tech approach

In order to enable wide distribution and technological readiness, existing software frameworks will be used as a basis for all developments. To facilitate this approach, the Service Platform is the central data hub and data processing component providing methods to build services using several proven technologies. The services will provide information in a suitable way, which will be used in the user interface and apps. On the other hand, the services can be used to gather feedback from users, file issues or report like in open311 or propose completely new things. The core of this is a component that enables semantic integration and enrichment of various datasets as shown in Figure 2 below. Based on the experiences with past research projects such as urbanAPI and Plan4Business we identified the need for more interoperability, exchange of data between stakeholders and the municipality and especially a holistic view on data to consider side effects and impact on others.

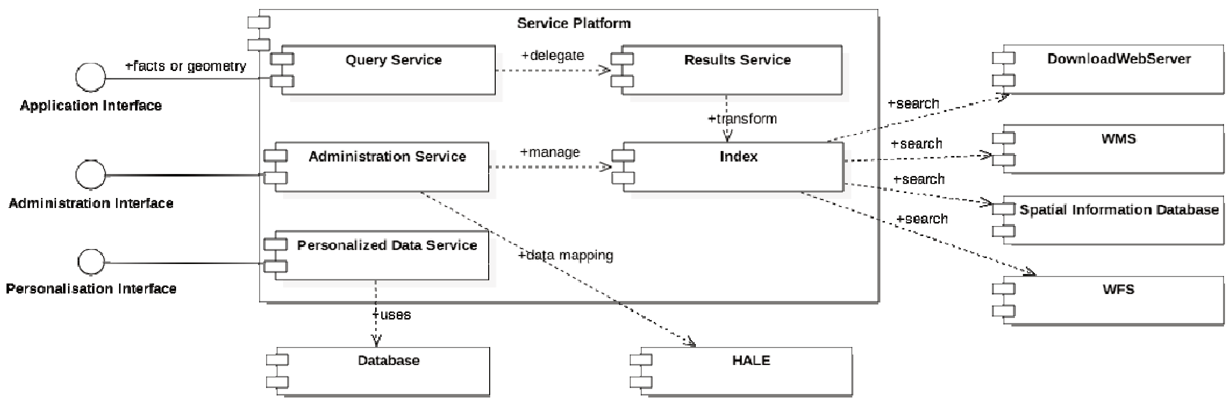

Fig. 2: Smarticipate Data Service Platform components.

The Service Platform is made up of the following parts: Technical interfaces for users and applications

- Application Interface – An interface to data for apps, providing linked data as facts or geometry data. This is the main interface for the aforementioned user interaction
- Administration Interface – An interface for users which maintain contents on the platform or also for integration of new data
- Personalisation Interface – Stores user related personalized data, can also contribute to analysis

On a functional level the platform will be made up by the following parts to be developed

- Query Service – the core to semantically enriched an integrated, holistic data, served in standardised formats like JSON-LD. Different data sets use different vocabularies for describing objects and circumstances. If these should be integrated, a common scheme needs to be developed and described. Technologies based on the Semantic Web initiative will be explored and a fit gap analysis will be carried out matching to requirements of real users.

- Results Service – Can analyse data, can draw conclusions and can check violations of conditions. A rule based query technology enables users to phrase a holistic question to the system. Also users can interact with such rules to work or change data. Which rule based systems are suitable is a question to be answered in the analysis and design phase. Also can enrich the result data set with a context, e.g. provide a complete 3D scenario with buildings related.
- Administration Service – Used to maintain data elements, import of new data and aids in semantically annotating data
- Personalized Data Service – Captures personalised context of users, preferences and settings. Every person has a different set of experiences and therefore a different approach to problems. An integrated semantically enriched system allows to connect different views on the same problems and provides methods to bridge the gap between different understandings or, at least opens up appreciation for the different view of people.
- HALE – Humboldt Alignment Editor is a tool for defining and evaluating conceptual schema mappings. With HALE domain experts will create logically and semantically consistent mappings and consequently harmonize geodata in the smarticipate platform.
- Databases, WMS, WFS, webservers for file downloads and other spatial information databases – The actual data sources for the platform. Contains the raw data, which is processed by the results service.

5.3 Piloting Cities

Three major European cities are committed to the project during the whole runtime: London - the borough of Kensington and Chelsea; Rome and Hamburg. At the beginning several workshops take place in each city gathering the requirements and use cases of each city and thus defining the scope of the developments. When the first usable software is available, an extensive joint piloting phase is started together with citizens, urban planners, stakeholders and the smarticipate team to work with the software under realistic conditions. The experiences made will directly feedback to development and assure the relevance of the software. To have a broad perspective of application each city has a different focus on the piloting use cases. However, the cases of the three cities fulfil common criteria such as:

- The selected used cases support citizens in taking over services from government, or in developing entirely new public services
- smarticipate users will receive direct, instant digital feedback that is customized to his or her own proposal. Used cases will provide the possibility for this kind of interactive relationship.
- smarticipate will provide continuous support and feedback to citizen initiatives; The used cases will not be single or periodic events, but ongoing activities.

Through the pilots and their transfer potential to other European cities, smarticipate will stimulate the creation, delivery and use of new services on a variety of devices, utilizing new web technologies, coupled with open public data.

6 CONCLUSION

Open data portals for spatial data are usually of low usability which is caused by the lack of semantic web technologies. Semantic web technologies could help to improve the usability but also improve the value of open data itself.

But the classical top-down approach imposed by many standardisation organisations and governmental agencies to normative semantics in geoinformation is a barrier for technologies such as the semantic web to be applied in its original sense. Apart from this the technology is very complex and not easy to use. Bottom-up initiatives need freedom of choice what is appropriate in their use case and heterogeneity is what people have to deal with. Easy to use approaches and simple access to data using existing technologies and concepts as search engines, folksonomies, tagging and automated annotation via machine learning should be applied in research projects to cater the needs of the public. Otherwise geodata solutions development will continue to be driven by data or commercial technology, which is a barrier for wide application, thus foiling the intention of any open data initiative on an European level. The goal of the smarticipate project is to find adequate software solutions to publish open data, make the data more visible and searchable and improve the user experience to some extent. In the scope of the smarticipate project the key technology aspects will be included and evaluated with piloting cities in real life scenarios given by citizens and cities.

7 REFERENCES

ISO/IEC JTC 1 Information technology: Smart cities – Preliminary Report 2014. Geneva 2015.

BERNERS-LEE, Tim: Semantic Web Roadmap. http://www.w3.org/DesignIssues/Semantic.html.

MCKEE, Lance; DE LATHOUWER, Bart; JACKSON, Mike: Information Technology Standards for Sustainable Development, OGC White Paper, 14-095. Open Geospatial Consortium 2015.

MALEWSKI, Christian; DAMBRUCH, Jens; KRÄMER, Michel: Towards interactive geodata analysis through a combination of Domain-Specific Languages and 3D geo applications in a web portal environment. In: REAL CORP 2015 Proceedings, pp. 60-616. Vienna, 2015

FRÖHLICH, Peter; VOGT, Marek: Smarticipate: Citizen involvement – the Key for Innovation in Approach and Empowering. In: REAL CORP 2016 Proceedings, Vienna, 2016.

DAMBRUCH, Jens; KRÄMER, Michel: Leveraging public participation in urban planning with 3D web technology. In: Proceedings of the Nineteenth International ACM Conference on 3D Web Technologies (Web3D '14), pp. 117-124. DOI=10.1145/2628588.2628591 http://doi.acm.org/10.1145/2628588.2628591. New York. 2014.

JANOWICZ, Krzysztof; HITZLER, Pascal: Thoughts on the complex relation between linked data, semantic annotations, and ontologies. In: Proceedings of the sixth international workshop on Exploiting semantic annotations in information retrieval (ESAIR ‘13) , pp. 41-44, New York, 2013.

JANOWICZ, Krzysztof; SCHNEIDER, Simon; PEHLE, Todd; HART, Glen: Geospatial semantics and linked spatiotemporal data–Past, present, and future. In: Semantic Web, 3(4), pp. 321-332. 2012.

KOVALERCHUCK, Boris; SCHWING, James: Visual and Spatial Analysis – Advances in data Mining, Reasoning and Problem Solving. Ellensburg 2005.

SWARTZ, Aaron: A Programmable Web – An Unfinished Work, Morgan Claypool. 2013.

SCHÄFER, Marco: Kozeption und Realisierung einer vertikalen Microservice Architektur für einen Online-Geodatenkatalog. Masterarbeit am Fachbereich Mathematik, Naturwissenschaften und Informatik and der Technischen Hochschule Mittelhessen. Gießen 2016.

FOWLER, Martin: Domain-specific languages. Addison-Wesley. Longman, Amsterdam, 2010.

KRÄMER, Michel; LUDLOW, David; KHAN, Zaheer: Domain-Specific Languages For Agile Urban Policy Modelling. In: Proceedings of the 27th European Conference On Modelling and Simulation (ECMS), edited by Webjørn Rekdalsbakken, R.T. Bye and H. Zhang, pp. 673-680. Ålesund, Norway, 2013.

KRÄMER, Michel; STEIN, Andreas: Automated Urban Management Processes: Integrating a Graphical Editor for Modular Domain-Specific Languages into a 3D GIS. In: Proceedings REAL CORP 2014, Vienna, Austria, pp. 99-108. Vienna 2014.

HART, Glen; DOLBEAR, Catherine: Linked Data: a Geographic Perspective. Taylor Francis. Boca Raton 2013.