PLAN4BUSINESS

A service platform for aggregation, processing and analysis of urban and regional planning data

SEVENTH FRAMEWORK PROGRAMME

Project number:	296282
Project acronym:	*plan4business*
Project title:	A service platform for aggregation, processing and analysis of urban and regional planning data
Instrument:	STREP
Call identifier:	FP7-ICT-2011-SME-DCL
Activity code:	

Start date of Project:	2012-04-01
Duration:	24 month

Deliverable reference number and title (as in Annex 1):	**D5.4 Storage Engine**
Due date of deliverable (as in Annex 1):	**M21**
Actual submission date:	*see "History" Table below*
Revision:	

Organisation name of lead contractor for this deliverable:
Fraunhofer IGD (Fraunhofer)

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium	
CO	Confidential, only for members of the consortium (including the Commission Services)	

European Commission
Information Society and Media

Title:
Storage Engine
Author(s)/Organisation(s):
Christian Malewski / Fraunhofer Tomas Mildorf / UWB
Working Group:
WP5
References:
Grant Agreement No. 296282, Annex I Description of Work

Short Description:
This deliverable summarises the work performed for Task 5.3, the Storage Engine. The Storage Engine is one of three layers implemented in the scope of Work Package 5. The implemented system is composed of three layers, namely the integration layer, the storage layer and the analysis layer. The Storage Engine layer's task is to store and provide data that is related to land use data. It therefore represents the central interface of the data processing pipeline. On the one hand it receives harmonized data from the Integration Engine and on the other hand it provides a central access point for harmonized and various other datasets for the Analysis Engine.
Keywords:
Storage Engine, primary data pool, secondary data pool

History:				
Version	Author(s)	Status	Comment	Date
001	Christian Malewski	new	outline of the report	09.12.2013
002	Christian Malewski	draft	introduction	10.12.2013
003	Christian Malewski	draft	Primary data pool - urban plan data	11.12.2013
004	Christian Malewski	draft	Primary data pool - Statistical data	12.12.2013
005	Christian Malewski	draft	Secondary data pool - Use Cases	13.12.2013
006	Christian Malewski	draft	The Pool storage manager	17.12.2013
007	Simon Templer	rfc	Small textual changes	18.12.2013
008	Tomas Mildorf	draft	Other thematic data	19.12.2013
009	Przemysław Turos	draft	Input for section 5.2	19.12.2013
010	Christian Malewski	Final draft	Available land use data – roadmap and future work	20.12.2013

011	Otakar Cerba	rfc	Minor textual changes on section 3	29.12.2013
012	Christian Malewski	final	Final revisions reflecting the reviewers' comments and additions	07.01.2014

Review:			
Version	Reviewer	Comment	Date
010	Tor Gunnar Øverli	minor comments	23.12.2013
010	Tomas Mildorf	minor comments	06.01.2014

Table of contents

List of acronyms

API	Application Programming Interface
DBMS	Database Management System
GML	Geography Markup Language
GPS	Global Positioning System
HALE	HUMBOLDT Alignment Editor
HILUCS	Hierarchical INSPIRE Land Use Classification System
NUTS	Nomenclature des unités territoriales statistiques
OGC	Open Geospatial Consortium
PDF	Printable Document Format
RDF	Resource Description Framework
REST	Representational State Transfer
SPARQL	SPARQL Protocol And RDF Query Language
SQL	Structured Query Language
WFS	Web Feature Service
WMS	Web Map Service

List of Figures

1 Introduction

This deliverable summarises the work performed for Task 5.3, the Storage Engine. The Storage Engine is one of three layers implemented in the scope of Work Package 5. Figure 1 outlines the role of the Storage Engine layer within the plan4business platform. The implemented system is composed of three layers, namely the integration layer, the storage layer and the analysis layer. The Storage Engine layer's task is to store and provide data that is related to land use data. It therefore represents the central interface of the data processing pipeline. On the one hand it receives harmonized data from the Integration Engine and on the other hand it provides a central access point for harmonized and various other datasets for the Analysis Engine.

As starting point for the work on the Storage Engine, we initially evaluated suitable DBMS that could be used as a basis for the primary and secondary data pool that form the foundation of the Storage Engine. We performed an evaluation on suitable components for the primary data pool and started the development on the database schema as well as the actual implementation, which is explained in more detail in the following section. In a second step we performed an evaluation of promising DBMS to use as a basis for the Secondary Storage Engine and developed several prototypes concerning technical feasibility as well as ease of use and integration for the Neo4j Graph Database. The detailed evaluation can be found in Deliverable 5.1.

The Storage Engine has been realized as a combination of two separated data bases following the relational paradigm on the one hand and the graph paradigm on the other. The main component of the Storage Engine is the relational data base called primary data pool. Although the relational paradigm has been carried out successfully for many years, it lacks in performance when it comes to more complex queries that require lots table joins. Thus, it has been supplemented with a graph data base that runs particular use cases. The graph data base founds the secondary data pool. Both, the relational data base and the graph data base, can be managed via the plan4business web portal, which represents the data base manager. The data base manager allows for storing, deleting or updating both, harmonized and non-harmonized data sets. Despite this, it also provides the functionality to transfer particular data sets from the primary to the secondary data pool.

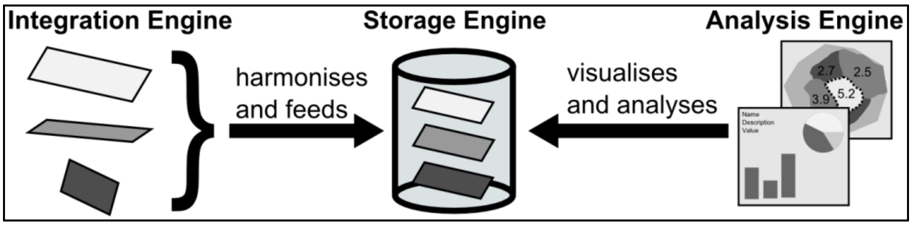


Figure 1 Composition of core engines

This deliverable is structured as follows. In chapter two the primary data pool's structure is presented. An overview on the secondary data pool is given in chapter three. These two are followed by a short summary of the web portal integrated Pool Storage Manager. The pool of acquired land use data is presented in chapter four. A chapter six outlines the remaining roadmap and starting points for further development.

2 Primary Data Pool

This section outlines the structure of the primary data pool. The primary data pool maintains communication interface to both, the Integration Engine and the Analysis Engine. The Integration Engine supplies the primary data pool with land use data. Each integrated vector data set references a metadata entry that also captures the corresponding original source data set. Besides that core data the primary data pool is extended with other data, such as geospatial reference keys, statistical data of a various set of themes, flood related data or metadata.

The Primary Pool Data Model is a SQL based schema for the PostGIS extension to the PostgreSQL Open Source DBMS. It conforms to the OGC Simple Feature Access Specification[1,2] to express spatial data and provides support for spatial indexes that greatly increase spatial query execution performance. Metadata entities within the data base store information about the location of uploaded files. Moreover, these entities are narrowly linked to the files, such that a deletion of an entity triggers removing the file on the file system. This allows for diverting the relational data base in order to manage the file system synchronously.

The Primary Pool Data Model does not impose strict limitations on what file types can be attached to a meta-data entry. Thus it is possible to store the original, untransformed source data sets, more detailed licensing and legal information as well as documents of arbitrary file formats[3]. Thus, it is also possible to store data that cannot be integrated to the planning data model by the integration engine. In particular that encompasses planning data that is not available in a sufficiently structured data format, such as PDF documents or plain raster data.

2.1 Land Use Database Schema

The core data is land use data. According to the INSPIRE Specification[4] on land use data and reflecting the intermediate schema discussed in Deliverable 5.2 land use data is divided either type-specifically or thematically. The former partition distinguishes areas with a homogeneous combination of land use types whereas the other is a collection of the former. Thematically land use data is divided into existing land use and planned land use data.

2.1.1 Main Entities

Figure 2 shows the type-specific fragmentation between the database relations `dataset` and `ft_landuse`. These two relations correspond to the `AbstractDataSet` and `AbstractLandUseObject` classes in the intermediate schema (Deliverable 5.2) and are therefore realized through a one (`dataset`) or many (`ft_landuse`) connection.

---

[1] OGC Simple Feature Access, Part 1: Common Architecture http://www.opengeospatial.org/standards/sfa

[2] OGC Simple Feature Access, Part 2: SQL http://www.opengeospatial.org/standards/sfs

[3] This is in fact a crucial feature as in many cases the legally binding part of spatial plans is often provided as not machine-readable text while the machine readable spatial vector data is just a supplementary and not strictly legally binding representation of the planning document.

[4] http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_LU_v3.0rc2.pdf

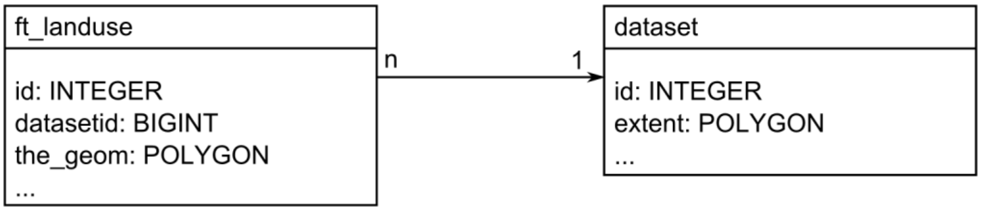


Figure 2 General land use relations

The `the_geom` attribute represents the spatial area of an `ft_landuse` feature. The spatial extent of all `ft_landuse` entities is aggregated in the `extent` attribute of a `dataset` entity. The `datasetid` is the foreign key connecting the `ft_landuse` and `dataset`.

Each of the previously mentioned relations is thematically divided into two sub-relations. The corresponding table relations are named `ft_zoning_element` for planned areas and `ft_existinglu` for present land use areas (Figure 3). The datasets are structured accordingly. There is one relation for both, present land use collections (`ds_existinglu`) and planned land use collections (`ds_spatialplan`) (Figure 4).

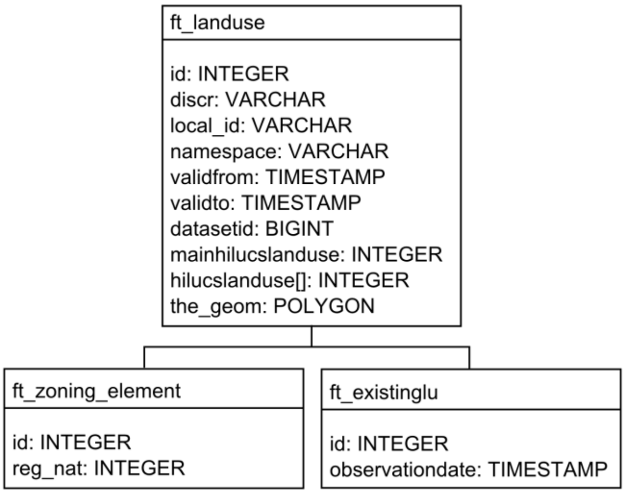


Figure 3 land use areas; planned and existing

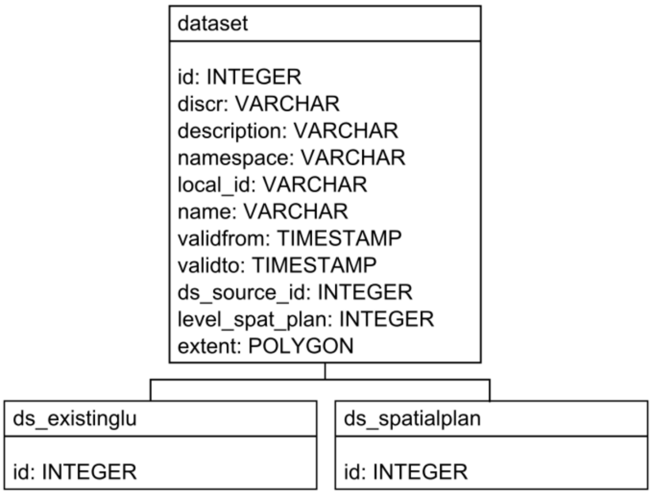


Figure 4 land use data sets; planned and existing

2.1.2 Code lists

Albeit the INSPIRE initiative's web pages provide a registry of code lists for land use data in a linked data manner (http://inspire.ec.europa.eu/codelist) they are mirrored in the data base. This allows for faster access and queries within the primary data pool. Those INSPIRE code lists that are supposed to be extended (e.g. provider specific value on land use) are furthermore made accessible in a linked data manner via a REST API (see Deliverable 5.2).

Figure 5 depicts usage of the internal code lists within the relational database schema. Three of the sketched code lists, namely the HILUCS land use values (`cl_hilucs`), the level of spatial plan (`cl_levelofspatialplan`) and the regulation nature (`cl_regnature`) are one to one representations of the code list determined in the INSPIRE specification on land use.

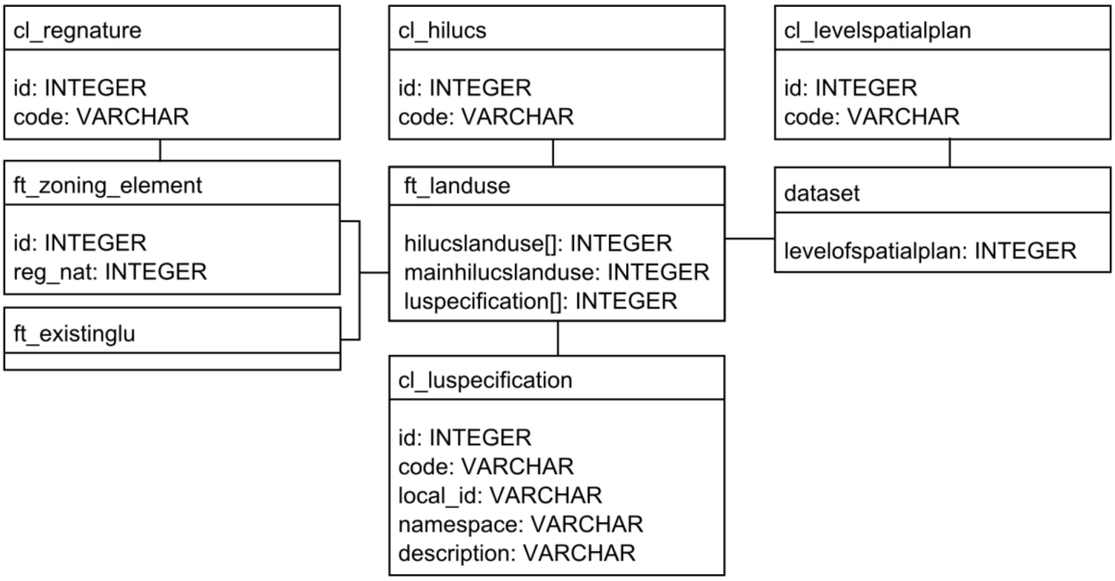


Figure 5 code lists in the primary data pool

2.1.3 Mapping entities

In order to allow for repeating a mapping e.g., when the original source data has been updated or the mapping instructions have been adjusted both, source data and mapping files, are stored as data base entities narrowly connected to files on the file server. Figure 6 sketches the mapping related entities. During a mapping a collection of planned or existing data sets is created from the source data. The spatial source files are represented by the `filedescriptor_sd` relation. These relations contain detailed information about the file, such as location, name, or size. An element of the `ds_source` can have one or many instances of the `filedescriptor_sd` relation. When a HALE[5] mapping project for a source data collection is uploaded the corresponding relation to an instance of `mapping_project` is created and linked to the source data entity.

Similar to source data the representations for the mapping project files are stored in the data base (`filedescriptor_mp`). Once the harmonized data is derived the harmonized data can be published in various ways. These include access via Web Map Service or as an INSPIRE conform GML file. The way a land use data set is published is stored in the `publishoption` relation.

The mapping process is tracked by the integration engine and made accessible through JasperReports[6] either on success or fail. The important information is maintained in the `mapping_report` relation table and the `mapping_report_dataset_detailed` table (Figure 7). These track the mapping process and give information of to several attributes, such as how many default values where automatically created or a reason why the mapping process might have failed.

---

[5] http://www.esdi-community.eu/projects/hale

[6] http://community.jaspersoft.com/

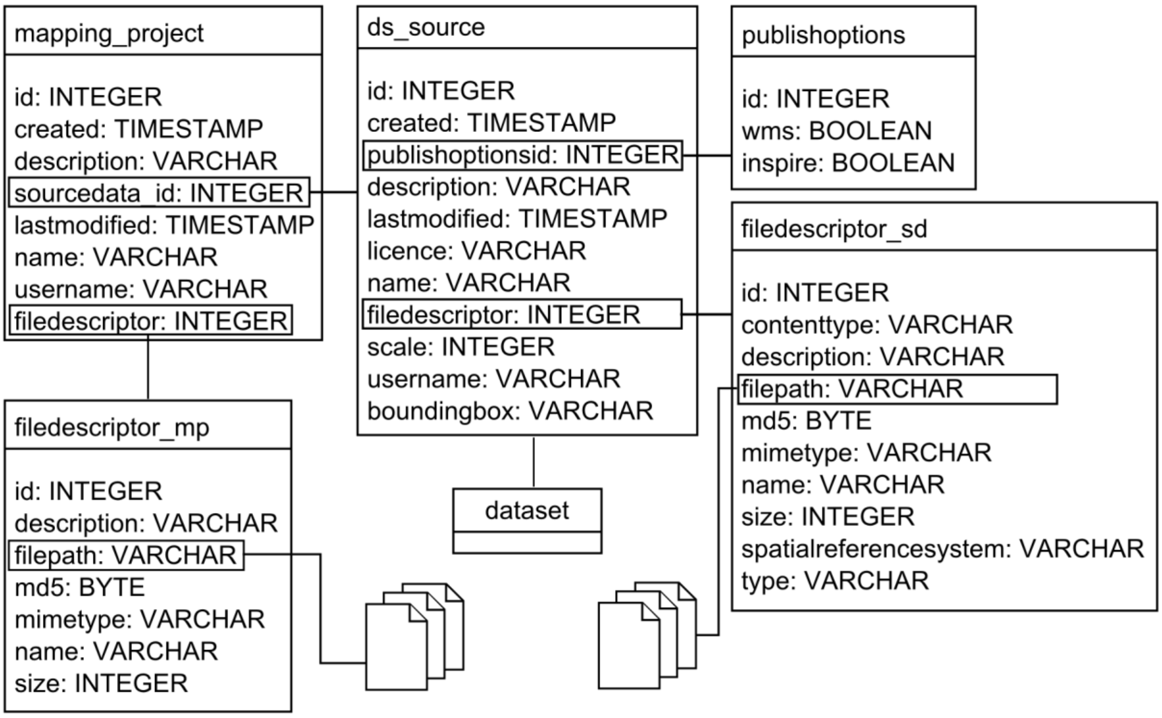


Figure 6 Model of mapping entities and linkage to document store

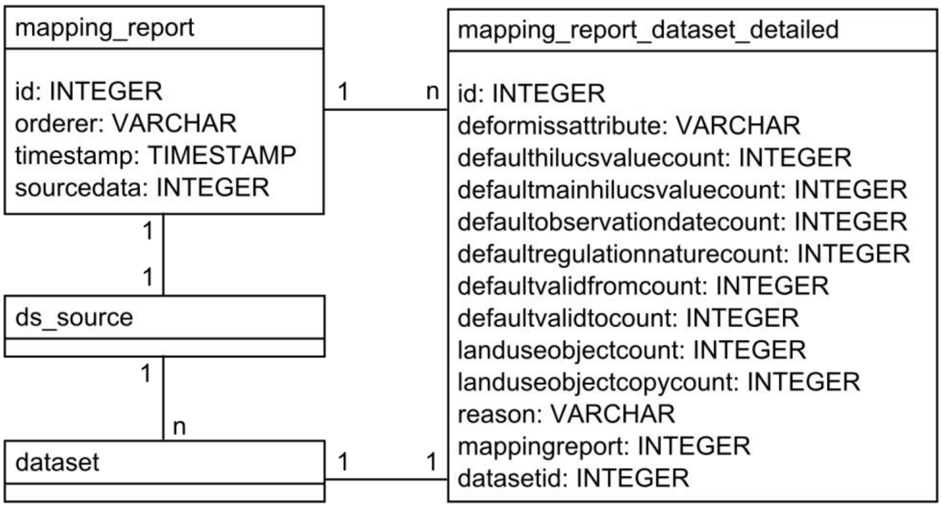


Figure 7 Model of mapping report and its sub-entities

2.1.4 Land use classification mapping

A crucial subtask of the schema mapping process within the context of INSPIRE land use is the re-classification of the land use categories. The data that is required for the web front-end is stored within the primary data pool as a collection of four classes (Figure 8). The basic class is the `hilucs_reclassification_project` that models the collaborative project for mapping land use

categories. Each instance of this relation has one or more `hilucs_reclassifications` that represent an alignment of a HILUCS value with a source data specific land use value. An entity of both, the project or the alignment object, is commented by a user. The comments are stored within the `hilucs_reclassification_comment` relation. Since a reclassification project can be created from a file, the corresponding file descriptor can be stored, too.

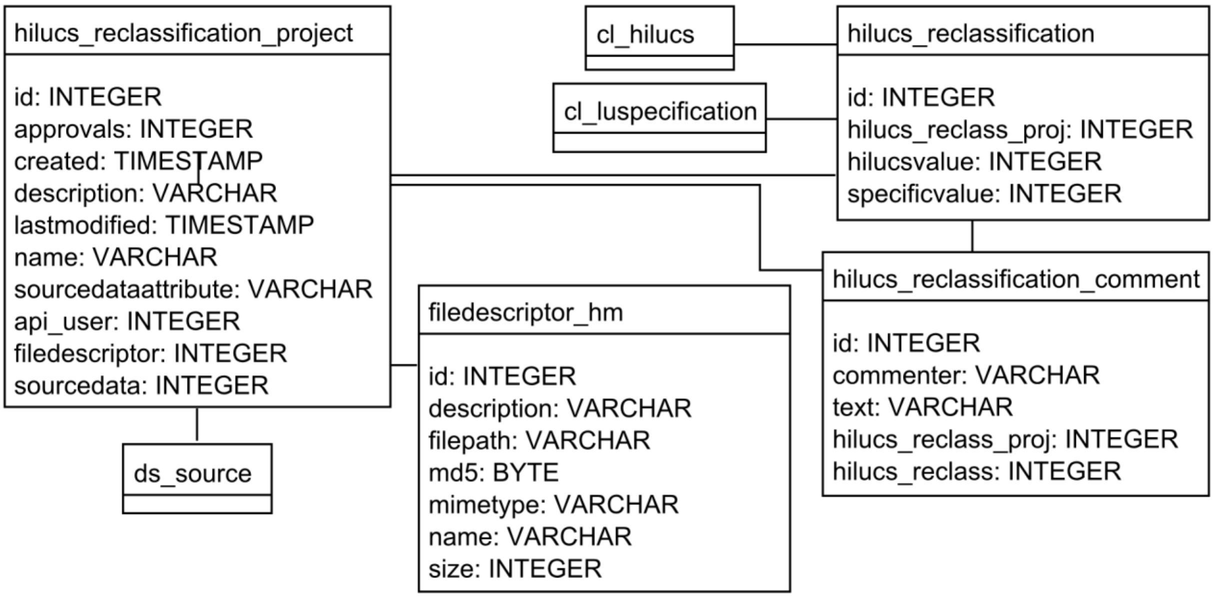


Figure 8 Model of land use reclassification project; linkage to document store and main application entities

2.1.5 Application specific entities

Besides the basic relations the urban plan schema stores application specific information. Figure 9 depicts the relations `api_user` and `publishinfo`. The previous stores information about the current user logged in to the web-frontend application. The latter stores information about the publish options of a dataset within the MICKA service catalogue.

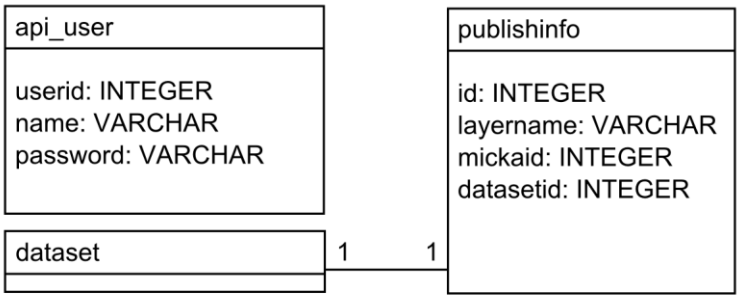


Figure 9 Application specific entities for API user and publish info

2.2 Statistical data

Three data base schemata store information about country specific statistical data, namely `cz_statistics` (Czech Republic), `de_statistics` (Germany), and `pl_statistics` (Poland). These data sets are directly taken from the corresponding country specific agency of statistics. Each of the schemata is a subset of the online available data. This allows for faster data access and guarantees data to be available when needed. However, the primary data pool by no means mirrors the statistical data base of a statistical agency. Its thematic focus is on seven categories that are shown in Figure 10. All attributes that have been selected from the statistical agencies can be aligned to one of them.

Besides a thematic division the data is also partitioned into a geographical level of detail. Figure 11 shows the level of details *Nomenclature des unités territoriales statistiques*[7] (NUTS) 0 to 3, where NUTS 0 is the most general level. A unique identifier in each table specifies the region code of a statistical entity, e.g. DE2 for Hamburg, Germany. These codes are interlinked with the spatial region in another table, e.g. `nuts_regio_transform` within the schema for German statistical data.

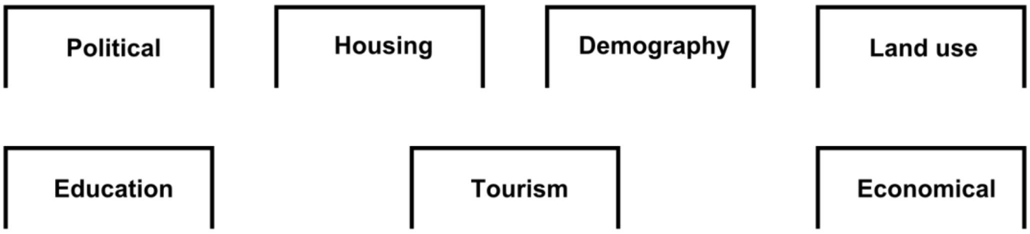


Figure 10 Main categories of statistical themes

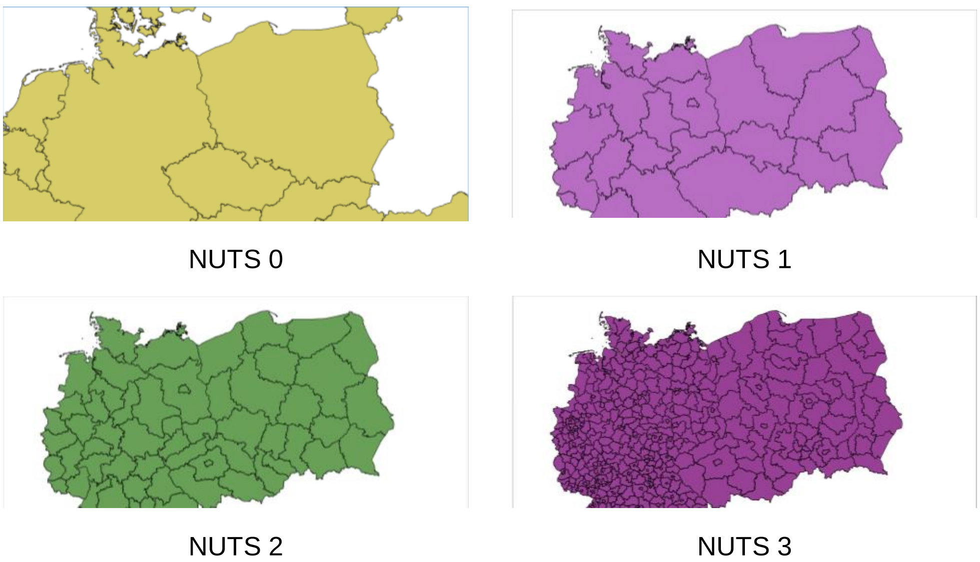


Figure 11 geographic region hierarchy (NUTS levels 0-3)

[7] http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/introduction

The country specific data is provided for more detailed regions. However, it is supplemented by data that is harmonized for whole Europe. This data is contained within the Eurostat schema and mirrors a sub-part of the Eurostat website, for the latest timestamp the data is available. The Eurostat schema is meant to be parsed quickly and thus, pre-computations have been made to get quicker access to the data.

2.3 Other thematic data

Besides the previously detailed main thematic data there are data sets that serve certain specific requests. These data is detailed in the following.

2.3.1 Corine Land Cover

The Corine Land Cover (CLC) data were acquired by the CORINE (Coordination of information on the environment) programme of the European Commission. This land cover/land use database covering 38 European countries includes the 2006 update (CLC2006) which is owned and distributed by the European Environmental Agency (EEA). The CLC2006 data are available as GIS vector data and are visual interpretation of satellite images. These data can be used for commercial and non-commercial purposes free of charge, provided that the source is acknowledged.

2.3.2 Cadastral data Czech Republic

The Czech cadastral data are published according to the INSPIRE Directive. “WFS download service for the theme Cadastral Parcels (CP) is a public download service for provision of data from the Information System of the Cadastre of Real Estates (ISKN) that makes possible repetitive download of these data in files by cadastral units and an online access to these data using the WFS 2.0.0 technology. The download service provides harmonised data from the theme Cadastral Parcels in GML format according to INSPIRE. The service is available only for cadastral units with a digital map (DKM/KMD) – effective for CP.CadastralParcel and CP.CadastralBoundary layers. The layer CP.CadastralZoning is available for the whole territory of the Czech Republic. The service fulfils technical guidance for INSPIRE download services version 3.0.1 and simultaneously fulfils the OGC WFS 2.0.0 standard.” (http://geoportal.cuzk.cz/)

A selected set of cadastral data and other data from various sources (e.g. municipalities, local building offices) are also published through the Register of Territorial Identification, Addresses and Real Estates (RUIAN). RUIAN is one of the four basic registers in the Czech Republic.

2.3.3 Disaster management

Flood data were provided by the T. G. Masaryk Water Research Institute from the Czech Republic. The data are included in the Digital Base of Hydrography Data (DIBAVOD), which is publicly available at http://www.dibavod.cz/index.php?id=27 in the ESRI shape file formats.

The data include 5-year, 20-year, 100-year and the biggest flood zones on the entire territory of the Czech Republic. The level of detail is equivalent to the map scale 1:10,000.

2.3.4 Spatial reference

The OpenStreetMap (OSM) is used as a base map. The map is collected by volunteers all over the globe and is distributed for free for commercial and non-commercial purposes.

OpenStreetMap (OSM) is a collaborative project aiming to create a free editable map of the world. The maps are created using data from portable GPS devices, aerial photography, other free sources or simply from local knowledge. Both rendered images and the vector dataset are available for download under a Creative Commons Attribution-ShareAlike 2.0 licence. Registered users can upload GPS track logs and edit the vector data using the given editing tools [2].

OpenStreetMap can be considered as an example of neogeography. Citizens have an interest in sharing their findings within a community. Therefore, the architecture should foresee collaboration possibilities for people who are not employed by environmental agencies. Data reported by the general public need to be checked and validated before it may be published. Looking at this example, it is clear that the services used to build an Integrated Environmental Information Space need not only “access” functions but also “input” functions [3].

3 Secondary Data pool

Based on the evaluation in deliverable 5.1 we have chosen Neo4j[8] as suitable graph data base for the secondary data Storage Engine. The corresponding prototypes for insert or convert data into a graph data base schema have been developed. Spatial data is expressed according to the OGC Simple Feature Access Specification and can also be indexed spatially, both of which is supported by the Neo4j-spatial extension. Neo4j provides a REST API that allows for making features accessible in a Linked Data manner. Further on a plugin is available that allows for filtering using SPARQL[9] and facilitates linkage to the Semantic Web [1]. At the time of writing this deliverable we still labour on the secondary data pool, since it is one of the last working packages to be finished. However in the following a detailed overview is given of what is planned to be achieved in the last three months (Project months 22 to 24) or in the project following actions.

3.1 Use Case 1 - Spatial Relations

Geospatial operations are very time consuming. An example is filtering by spatial relations, e.g. spatial containment. Computing such relations means evaluating the coordinates of a region or its bounding box amongst all other features in the data base. Pre-ordered graph structures such as R-Trees [4] support these computations and allow for faster results. Despite this, utilizing qualitative expressions defined in the Region Connection Calculus (RCC) [5] (e.g. *contains*) reduces geospatial computation effort to filtering for a particular relation during run-time.

Albeit, qualitative relations can be modelled in a relational database as depicted in Figure 12 it corresponds not to the conceptual approach of a relational database. Furthermore, queries that combine two or more relations require join operations in relational databases. Graph data bases do have a conceptual advantage over relational databases. Thus, a first use case for the secondary data pool is utilization of spatial relations. The following shows a query that might be formulated by a user that intents to find a location for installing renewable energy sources. They might be built around a region where a citizen with average income can afford a more user subsidised but nature friendly form of energy.

```
Detect a region contained in Germany and average income of over
1,500 € and externally connected to a region with minimum area of
X hectares of 6_4_AreasWhereAnyUseIsAllowed
```

Spatial relations are pre-computed on the previously mentioned NUTS regions and whenever an urban plan related dataset is uploaded. Relations can be adopted from the RDF vocabulary NeoGeo[10]. This builds a connection to the Semantic Web and Linked data approach and allows for reasoning over relations at a more advanced stage of the application and therefore gives a starting point for future development.

[8] http://www.neo4j.org/

[9] http://www.w3.org/TR/rdf-sparql-query/

[10] http://geovocab.org/spatial.rdf

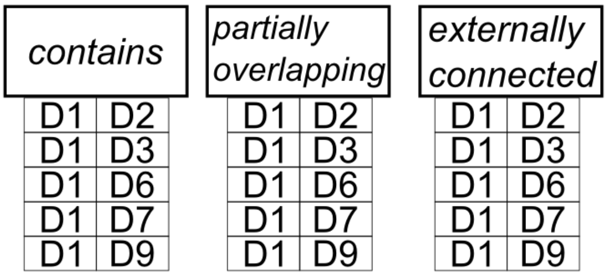


Figure 12 Spatial relations modelled in relational database

3.2 Use Case 2 - Harmonizing Statistical Data

Although statistical data of the Eurostat database is harmonized for EU countries, additional statistics from the state agencies for statistics complement the available data significantly in terms of level of detail or thematic range. As mentioned previously the statistic themes are divided into seven categories depending on the already available data sets. Since, this data module might change in further development organizing that data in a graph data base is advantageous. A graph data base allows for easier adjustment of a data base schema than a relational data base.

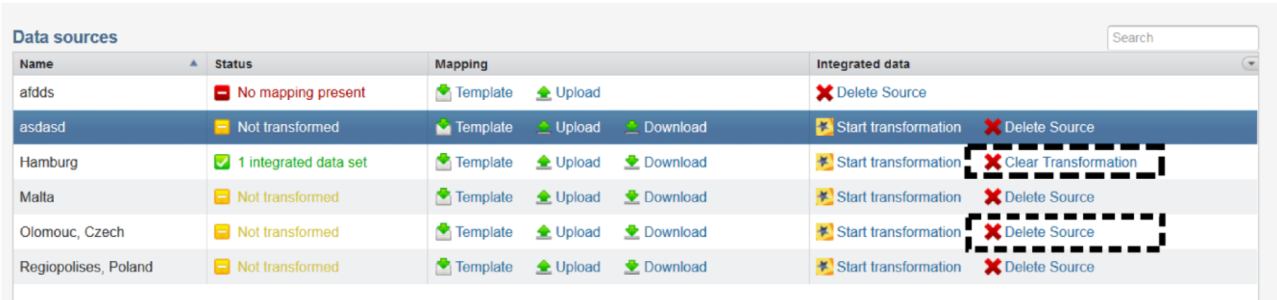
Storage of statistical data in a graph data base furthermore allows for direct linkage of spatial regions with their corresponding statistical data.

3.3 Use Case 3 - Land Use Category Mapping

One subtask of the schema mapping outlined in Deliverable 5.2 is the land use category mapping. The land use category mapping aligns land use values of a source data set to those of the INSPIRE HILUCS code list. A land use value mapping, however, cannot always performed as a direct but rather a nearly match. These graduations in mapping relations lead to application of a graph based storage of land use mappings.

4 The Pool Storage Manager

The Pool Storage Manager represents a third component of the Storage Engine. It acts as an access layer to the Storage Engines and should manage synchronization of those parts of the data needed for analysis. The Pool Storage Manager has been integrated to the web portal (see Deliverable 5.2). It currently allows for deletion or update of both, harmonized or source data sets (Figure 13). An additional feature to this functionality is the already present user management. Thus, we have decided to integrate synchronization of the data in the primary storage pool and the secondary storage pool to the portal or its backend, respectively.



Data sources

Search

Name	Status	Mapping			Integrated data	
afdds	No mapping present	Template	Upload		Delete Source	
asdasd	Not transformed	Template	Upload	Download	Start transformation	Delete Source
Hamburg	1 integrated data set	Template	Upload	Download	Start transformation	Clear Transformation
Malta	Not transformed	Template	Upload	Download	Start transformation	Delete Source
Olomouc, Czech	Not transformed	Template	Upload	Download	Start transformation	Delete Source
Regiopolises, Poland	Not transformed	Template	Upload	Download	Start transformation	Delete Source

Figure 13 Implemented Pool Storage Manager functions

In what follows three requirements are listed that are planned to be implemented and refer to the secondary data pool.

Requirement 1: The Pool Storage Manager automatically synchronizes a pre-configured sub-set of statistical data available in the primary storage pool with the secondary storage pool.

For **Requirement 1** the functionality has to be implemented that checks, whether all data contained in the primary data pool is available in the secondary data storage. This has to be performed only on application start up.

Requirement 2: The Pool Storage Manager is able to fill any source data set into the secondary data pool. This process is activated by the portal administrator.

Requirement 3: The Pool Storage Manager is able to delete any source data set from the secondary data pool. This process is activated by the portal administrator or automatically when data set is removed from primary data pool.

Requirement 2 and **Requirement 3** are implemented parallel to the development of the secondary data pool ontologies. Similar to the process of deletion and update of data from the primary data pool this functionality is integrated to the web portal. However, the division of secondary and primary data pool storage should be transparent to the common user. Thus, they will be added to a particular admin view.

5 Available land use data

Mapping projects have been created for all available land use data sources. Besides mapping of entities and attributes (see Deliverable 5.2) a mapping includes a re-classification from the original land use classification to the INSPIRE HILUCS land use classification. Land use data sources and their mapped INSPIRE land use equivalents have been uploaded and fed into the primary data storage pool. In a next step, the created mapping projects and land use re-classifications need to be evaluated by the original data providers. This section lists land use data by countries.

5.1 Germany

The city state **Hamburg** in Germany maintains a geodata catalogue service. Thus, land use data but also other themes can be searched and are publicly accessible from the Hamburg Metadata Catalogue (www.HMDK.de) provided by *Landesbetrieb Geoinformation und Vermessung* Hamburg. The data is provided from WMS and WFS[11]. The latter has been used to download GML vector data of planned land use from 2012. The data set contains about 2600 land use entities (Figure 14). These entities are comprised to one spatial plan. The harmonized data set is uploaded to the primary data pool.

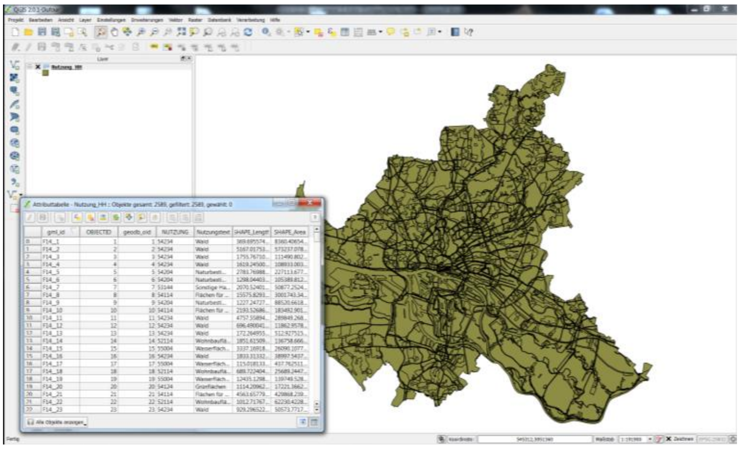


Figure 14 Planned land use in Hamburg, Germany

The *Regionalverband FrankfurtRheinMain*[12] maintains spatial plans for the **metropolitan area of Frankfurt**, in South of the state Hessen, Germany. Raster data is available online[13]. For the plan4business project the vector data subset from the spatial planned area of the city of Frankfurt (2010) is available. The source schema has been mapped to the INSPIRE compliant intermediate schema and a sub-set of the data has been transformed. However, due to licence issues the data is currently not uploaded to the project data pool.

[11] http://gateway.hamburg.de/OGCFassade/DE_HH_WFS_INSPIRE_A3_4_Bodennutzung_gesamtstaedtisch.aspx?SERVICE=WFS&REQUEST=GetCapabilities&VERSION=1.1.0

[12] http://www.region-frankfurt.de/Planung/Regionaler-Fl%C3%A4chennutzungsplan

[13] http://www.region-frankfurt.de/Planung/Regionaler-Fl%C3%A4chennutzungsplan/Hauptkarte

When the licence for public provision of the harmonized data is available the whole data set can be uploaded immediately.

The *Regionalverband Ruhr* maintains planned and existing land use for the **metropolitan region Ruhr** and has provided the project with the planned and existing land use vector data sets of Bottrop (2011), North Rhine Westfalia, Germany. The source schema has been mapped to the INSPIRE compliant intermediate schema and a sub-set of the data has been transformed. However, due to licence issues the data is currently not uploaded to the project data pool. When the licence for public provision of the harmonized data is available the whole data set can be uploaded immediately.

5.2 Poland

Within Poland spatial plans of 19 cities and towns of different size have been acquired. Most of the cities have 100,000 citizens or more. However, only one city (Ruda Śląska) has been completely covered with vector data. Furthermore, data for one city may consist of several hundred individual spatial plans that might not be consistent within one town. Thus, a pre harmonization has been performed. Currently the data of 9 cities have been uploaded to the primary pool (Figure 15). In the following the cities from which spatial plans exist are listed.

- Wrocław
- Bydgoszcz
- Toruń
- Gorzów Wielkpolski
- Kraków
- Tarnobrzeg
- Rzeszów
- Białystok
- Bielsko-Biała
- Tychy
- Kielce
- Poznań
- Koszalin
- Gdynia
- Tarnów
- Kędzierzyn-Koźle
- Częstochowa
- Rybnik
- Ruda Śląska

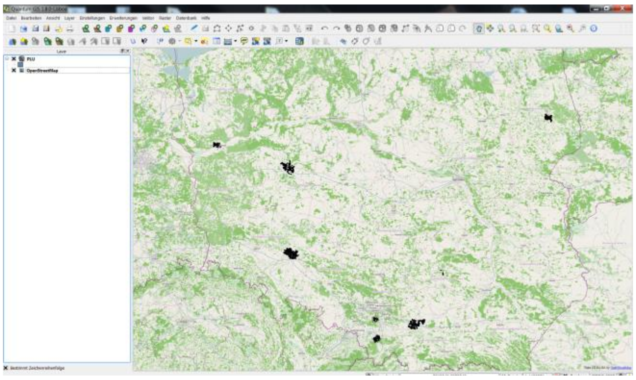


Figure 15 Planned land use in Poland

5.3 Ireland

The portal and data at myplan.ie is owned by the Department of Environment, Community and Local Government of Ireland. Herein land use plans are accessible in both, raster and vector data from a huge area of Ireland. These data sets have been mapped into the INSPIRE intermediate schema and populated to the primary data storage. The catalogue provides more than 260 planned and existing land use data sets. The validity ranges from 2000 to 2020.

5.4 Czech Republic

Within the Czech Republic spatial and existing land use data sets from the regions Liberec and Olomouc have been acquired. The data sets of Olomouc have already be mapped and uploaded to the server. The land use categories of the data sets of Liberec are currently mapped to the INSPIRE HILUCS classification.

5.5 France

A limited set of spatial plans in France as vector format could be acquired from a catalogue service located at http://catalogue.sigloire.fr/catalogue/. These plans are located within the French regions Brittany and Pays-de-las-Loire. The mapping to both HILUCS classification and INSPIRE intermediate schema is currently ongoing.

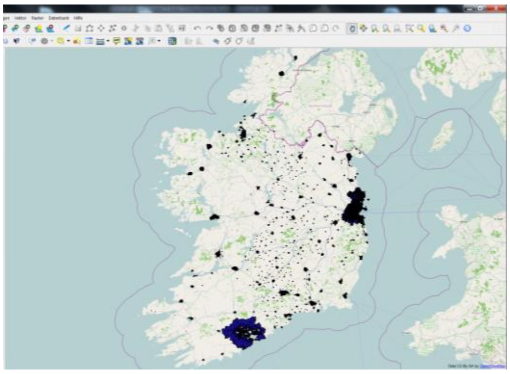


Figure 16 Planned land use Ireland

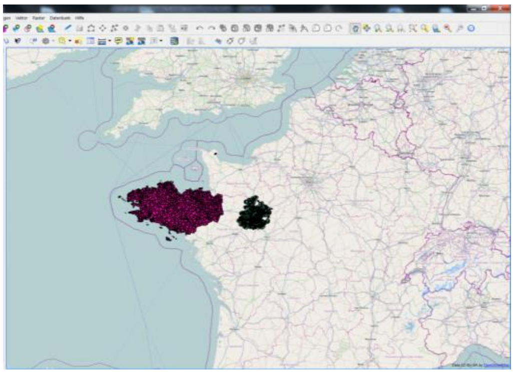


Figure 17 Planned land use West France

5.6 Spain

Up to now two public access points for land use data in Spain have been identified. These provide land use data from the region of Navarra in North Spain (http://idena.navarra.es/busquedas/catalog/main/home.page) and the region around the city Gijon in North West Spain (http://ide.gijon.es/en_index.html). For the Spanish data sets the mapping is not yet finished at the time the report is written.

5.7 Norway

To represent a Norwegian region, existing and planned land use data sets for Arendal have been acquired. These data sets have been mapped and uploaded to the primary data base. A comprehensive dataset with municipal plans for a number of Norwegian municipalities has been provided for the purpose of the project from the Norwegian Mapping Authority and is now being assessed for harmonization, based on the model designed for Arendal.

5.8 Malta

Maltese data has already been acquired for the plan4all project. That data has been reused and covers the complete country of Malta. The data has been mapped during the plan4business project and uploaded to the primary data pool.

6 Roadmap and future work

This section outlines the roadmap that will be performed during the last three month of the project within the scope of the Storage Engine. The Storage Engine also gives starting points for further work done based on the project.

Due to month 21 an instance of Neo4j has been setup. Both, the spatial and SPARQL endpoints have been setup and tested with prototypes. Additionally a Linked Data vocabulary to represent and interlink statistical categories, land use data and geographical data is developed.

Completion of vocabulary for statistical categories, land use data and geographical data is the first step to be achieved. Based on this step the data can be uploaded to the secondary data storage.

Enable synchronization of land use data. For this step a trigger for each data set has to be added to upload and data available in the primary data pool also to the secondary data pool based on the vocabulary from the previous step. At the same time functionality for removing data from the secondary data pool has to be developed. Manual synchronization of urban planning data allows for control the available data sets.

Upload and test land use data. Within this step the available mapped land use data sets will be uploaded to the secondary data pool. The data will be tested in preparing a set of SPARQL queries.

Enable synchronization of geographical data. In order to achieve Use Case 1 (Section 3.1) a subset of the geographical data from the regions NUTS 0 - 3 will be added to the secondary data pool.

Compute and test spatial relations. A separate step has to be performed to compute a network of spatial relations amongst the areas, where the focus is on the general relations, such as *contains*, *is part of*, or *is tangent to*. Spatial relations will be computed for all NUTS regions and land use data. For testing the correctness of spatial relation computation a set of SPARQL queries will be developed and performed.

Enable synchronization of and test statistical data. This step interlinks a subset of statistical data with the geographic regions committed in the previous steps. In order to allow for control of statistical entities the synchronization will firstly be performed manually using buttons within the web portal.

The roadmap presented above refers to obtain data form the secondary Storage Engine. Currently a prototype is developed that supports mapping of land use values to the HILUCS classification. Due to know it uses the relational data base as endpoint for storing land use category mapping projects and results (see section 2.1.4). Within this scope it will be investigated how a graphical data base can better support the flexibility of alignment relations (see section 3.3). HILUCS mappings that may be publicly accessible from the Storage Engine may facilitate efforts to clearly specify land use category mapping for those agencies that will have to transform their data into an INSPIRE conform format in the future.

Section 5 summarized the effort that has been done up to now to harmonize and upload land use data sets to the Storage Engine. However, not all available plans have been mapped so far, since adjustments to the intermediate schema were necessary in certain stages of the Integration Engine. Within the next months it will be investigated to map the remaining data sets and make them accessible from the primary Storage Engine.

Another point for investigating in depth based on the plan4business project is the analysis of land use development and possible dependencies of other thematic data. Regularly updated land use data will serve as a base for this type of analyses.

References

[1] Berners-Lee, Tim, James Hendler, and Ora Lassila. The semantic web. *Scientific american* 284.5 (2001): 28-37.

[2] Bennett, Jonathan. *OpenStreetMap*. Packt Publishing, 2010.

[3] Pillmann, Werner, and Jiří Hřebíček. Information Sources for a European Integrated Environmental Information Space. (2009): 341-352.

[4] Guttman, A. (1984). R-Trees: A Dynamic Index Structure for Spatial Searching. *Proceedings of the 1984 ACM SIGMOD international conference on Management of data - SIGMOD '84*. p. 47.

[5] Renz, J. (2002). Qualitative spatial reasoning with topological information. Springer-Verlag.