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):	**D5.2 Integration 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

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:
Integration Engine
Author(s)/Organisation(s):
Simon Templer, Christian Malewski, Eva Klien, Robert Gregor / Fraunhofer IGD Tomas Mildorf / UWB Michal Sredl / HSRS
Working Group:
WP5
References:
Grant Agreement No. 296282, Annex I Description of Work D4.1.1 Operational System V1 D4.1.2 Operational System V2 D5.1 Interim report on Integration, Analysis and Storage Engines D5.3 Analysis Engine D5.4 Storage Engine

Short Description:
This deliverable summarises the work performed and progress achieved within Task 5.1 Integration Engine Development from WP5. The report demonstrates the progress of the task's development towards the goals given by the Description of Work and stakeholder requirements. It describes the server side Integration Engine for the *plan4business* platform that is responsible for the integration of harmonised data into the primary data pool.
Keywords:
integration engine, layer manager, land use, API, *plan4business*

History:				
Version	Author(s)	Status	Comment	Date
001	Simon Templer Robert Gregor Eva Klien Tomas Mildorf Michal Sredl	new	Initial content based on D5.1	03.12.2013
002	Simon Templer	rfc	Major updates and addenda to all chapters	18.12.2013
003	Christian Malewski	rfc	Minor changes	18.12.2013
004	Simon Templer	final	Updates based on comments (mainly sections 4.1, 4.3.1 and 4.5.2).	06.01.2014

Review:			
Version	Reviewer	Comment	Date
003	Karel Charvat		19.12.2013
003	Tomas Mildorf		19.12.2013

Table of Contents

List of Figures

1 Introduction

1.1 *plan4business*

plan4business is a European project running from April 2012 until March 2014 and co-financed by the 7th Framework Programme of the European Commission. The full title is *plan4business – a Service Platform for Aggregation, Processing and Analysing of Urban and Regional Planning Data*.

plan4business develops a service platform for aggregation, processing and analyses of urban and regional planning data in Europe. Harmonised data will be integrated into a seamless, homogenous, constantly growing and updated trans-border dataset. The platform will enable spatial analyses across European datasets. The platform should serve not only as a catalogue of planning data but also as their integrator enabling user to search, view, analyse and download spatial planning data on European and regional levels. The main project objectives are the automation of harmonisation processes and possibilities of complex analyses.

The *plan4business* consortium comprises six organisations securing the project execution:

- Fraunhofer IGD - Fraunhofer Institute for Computer Graphics Research, Germany
- UWB - University of West Bohemia in Pilsen, Czech Republic
- HSRS - Help Service - Remote Sensing, s. r. o., Czech Republic
- ISOCARP - International Society of City and Regional Planners, The Netherlands
- GEOSYS - GEOSYSTEMS Polska, Poland
- AVINET - Asplan Viak Internet as, Norway

1.2 Aim of the Report

The *plan4business* project should significantly contribute to decision making processes on various governmental levels and in cross-border activities. Next to issues of data availability and business modelling, the *plan4business* consortium is facing challenging problems of data integration, storage and analysis. These are the main issues that are tackled within WP5. This document focuses on the Integration Engine component that is part of the plan4business platform, which is responsible for integrating harmonised data into the platform's data pool.

This deliverable summarises the work performed and progress achieved in Task 5.1 Integration Engine Development in WP5 of the *plan4business* project. It builds upon D5.1 which is the interim report on the Integration, Analysis and Storage Engine components. This is the final report on Task 5.1 and describes the developed Integration Engine, its data model, features and API, and the integration with other platform components.

1.3 Structure of the Report

The document is structured in **Fehler! Verweisquelle konnte nicht gefunden werden.** chapters. The content of the chapters is as follows:

Chapter 1 contains a brief summary of the project, the main objectives of WP5 and Task 5.1 and the structure of the document.

Chapter 2 contains definition of terms that are essential for understanding of the content of this document.

Chapter 3 describes the objectives of the Integration Engine in relation to the plan4business project and platform.

Chapter 4 describes the status and progress of the Integration Engine and its relation to other platform components.

Chapter 5 summarises the report and the development status and concludes the document with an outlook.

2 Terminology

Application Programming Interface (API) – “An application programming interface (API) is a protocol intended to be used as an interface by software components to communicate with each other. An API is a library that may include specification for routines, data structures, object classes, and variables.” (Wikipedia contributors 2013a)

Atomicity, Consistency, Isolation and Durability (ACID) - a common behavioural model for databases and distributed systems that support concurrent transactions.

Geography Markup Language (GML) - “OGC`s XML-based language for describing and encoding geospatial information. An application of XML, a specification developed by members of the Open GIS Consortium. http://www.opengis.org/techno/specs/00-029/GML.html ". GML is an XML encoding for spatial data. In a sense, it is a schema-writing language for spatial information.” (OGC 2012)

HUMBOLDT Alignment Editor (HALE) – “a tool for defining and evaluating conceptual schema mappings. The goal of HALE is to allow domain experts to ensure logically and semantically consistent mappings and consequently transformed geodata. Furthermore, a major focus is put on documentation of the schema transformation process and its impacts, e.g. in the form of lineage information attached to the resultant transformed data.” (Data Harmonisation Panel 2013)

Infrastructure for Spatial Information in the European Community (INSPIRE) – a European directive ensuring that the spatial data infrastructures of the Member States are compatible and usable in a Community and trans-boundary context.

Shapefile – vector data format for spatial data developed by ESRI.

Unified Modelling Language (UML) – a standardized general-purpose modelling language in the field of object-oriented software engineering. The Unified Modelling Language includes a set of graphic notation techniques to create visual models of object-oriented software-intensive systems. (Wikipedia contributors 2013b)

3 Objectives

The Integration Engine is one of the core components of the *plan4business* platform. It is intended to perform the transformations necessary to seamlessly integrate heterogeneously structured, externally provided spatial vector data into the *plan4business* pool of homogeneously structured vector planning data. The integrated data is the base for several services and applications within the platform.

The *plan4business* platform is composed of three layers as depicted in Figure 1:

1. **Client layer (portal)** - user interface for performing data integration, management, analyses and visualisation.
2. **Integration and analysis layer** - engines for data harmonisation and data analyses. It includes API for better exploitation of the *plan4business* platform features in other applications.
3. **Storage layer** - optimised storage for data and metadata.

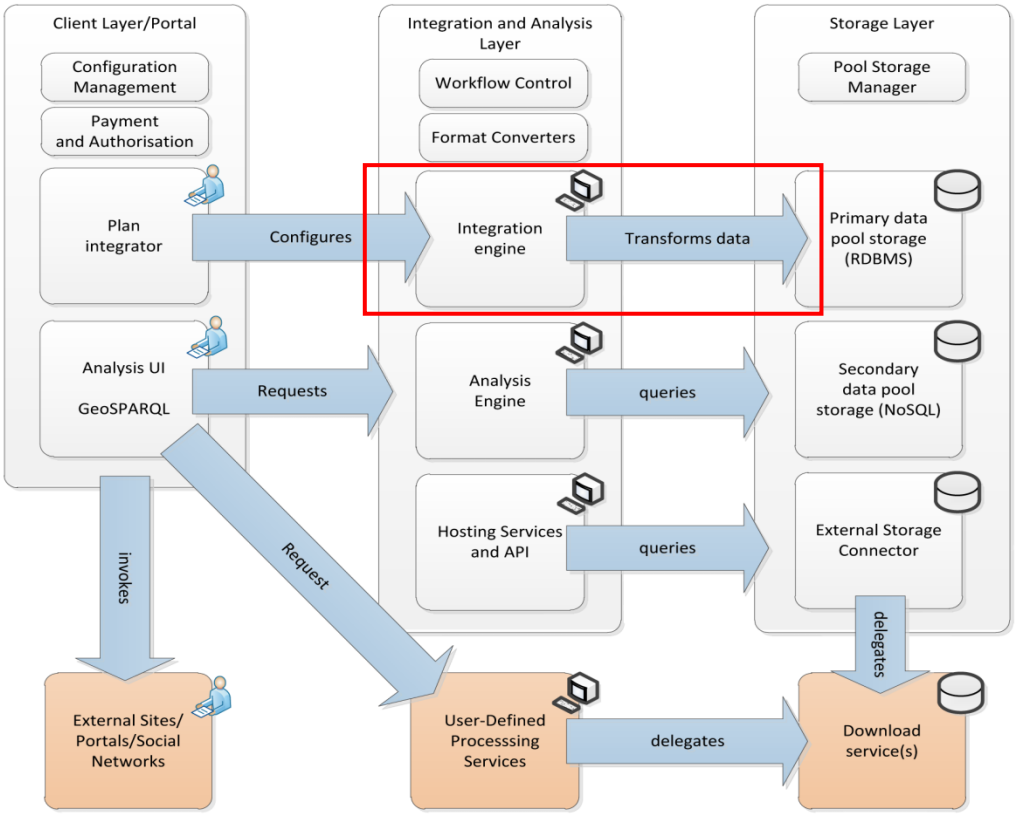

Figure 1 Three tier system of the plan4business platform (Fraunhofer 2012)

This report's focus is on the Integration Engine component of the integration and analysis layer. Reports on the Analysis and Storage Engines are in separate deliverables (D5.3 and D5.4). The client layer is developed in the frame of WP4 Plan Integration & Analysis Clients (see D4.1.2).

The main objectives for the Integration Engine are:

[O1] To support the **integration of planning data into the primary data pool** of the Storage Engine, based on the configuration performed through the Plan Integrator component. The Plan Integrator is the main interface for users interacting with the Integration Engine (see sections 4.2 and 4.3).

[O2] To **publish the integrated data** for use in the platform's services and applications. Other components of the platform, especially the Analysis Engine, need to be made aware of the available integrated data sets (see section 4.4).

[O3] To **provide an external API** for Integration Engine configuration and integrated data access (see section 4.5).

Fulfilling these objectives requires close cooperation with tasks responsible for other platform components, especially the Plan Integrator (Task 4.1), the Storage Engine (Task 5.3) and the Analysis Engine (Task 5.2).

4 Work Done and Progress Achieved

4.1 Overview

Based on the input data supplied through the Plan Integrator user interface (WP4, see D4.1.2) and a user provided schema alignment, the Integration Engine as part of the *plan4business* platform transforms input data to the schema used by the Storage Engine (WP5, see D5.4). In this process various conversions and transformation steps such as mapping between different data formats, geometric representations and conceptual schemas are performed by the integration engine. This functionality is implemented on top of the functionality of the Humboldt Alignment Editor[1] (HALE) and its Conceptual Schema Transformer (CST) component. HALE is a desktop software that was initiated within the EU funded HUMBOLDT project (FP6) and is continuously developed as open source software under the lead of the Data Harmonisation Panel[2].

In the scope of *plan4business*, the actual data transformation is performed on the server. Several data transformation processes that were initiated by different users can be executed at the same time. Furthermore the transformed data is not written to an output file within the file system but directly to the database that is the main component of the primary data pool. Thus the data can be concurrently retrieved, analysed or extended in a resource-efficient and consistent way. To support this, the definition of a common data model and database schema was done in close coordination with the work on the Storage Engine (Task 5.3).

Once data is integrated into the primary data pool, it is automatically published and made available according to the user's options. For each individual spatial plan or existing land use data set metadata is published to the MICKA metadata catalogue. Optional is the publication as individual WMS layer through the LayerManager API (see section 4.5.2) or the publication as INSPIRE compliant GML through a download service.

The Plan Integrator, LayerManager and MICKA metadata catalogue components the Integration Engine interacts with are described in detail in D4.1.2 Operational System V2. More details on HALE and how it is used in context of plan4business can be found in section 4.3.1.

[1] http://dhpanel.eu/humboldt-framework/hale.html

[2] http://dhpanel.eu

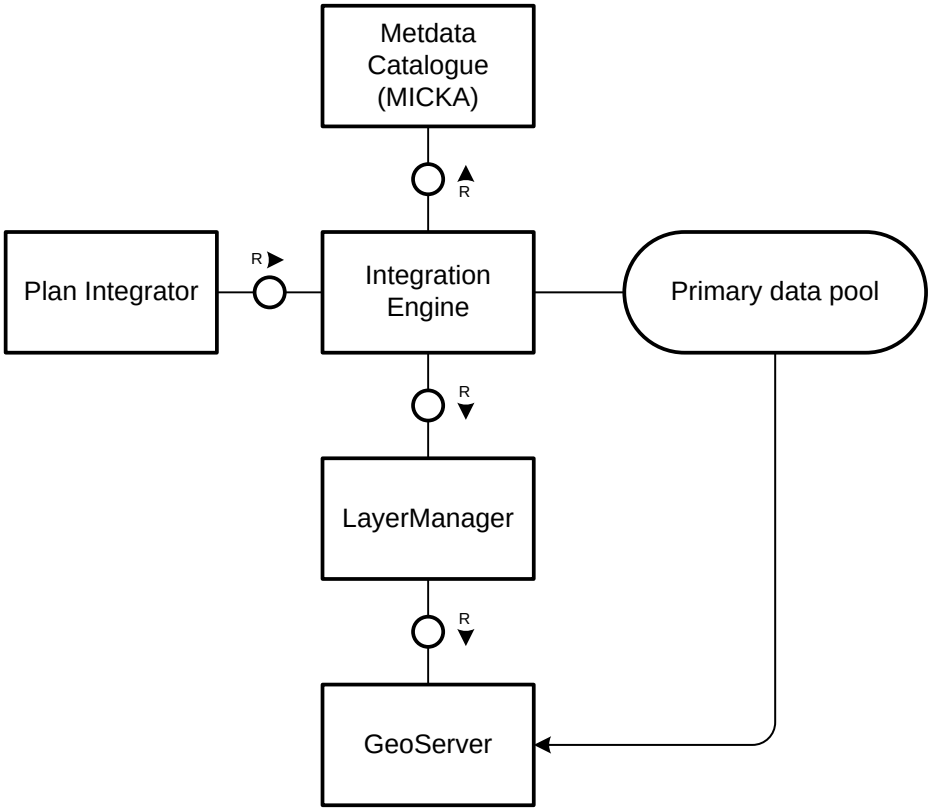

Figure 2 FMC block diagram on interaction of the Integration Engine with other components of the plan4business platform – other components may be linked indirectly to the Integration Engine through GeoServer or primary data pool access.

4.2 Data Model Design

Early on in the project it was decided to focus on land use data as core planning data to integrate into the platform. The Integration Engine is responsible for the transformation of the data to a common data model and storage in the primary data pool. The Analysis Engine requires this data model to be able to perform analyses on all integrated land use data sets and spatial plans in a uniform way.

Figure 3 gives an overview of the data models for land use data within the *plan4business* platform. The models directly related to the Integration Engine will be addressed in the following subsections, while the derived models for the primary and secondary data pool of the Storage Engine are addressed in D5.4.

The *INSPIRE Data Specification on Land Use*[3] was chosen as a functional basis for the land use data models in *plan4business*. The data specification is developed in the scope of the INSPIRE initiative that is targeted to develop legal and technical frameworks for a common European spatial data infrastructure (SDI) and a common European pool of geospatial data that can be easily integrated with various geospatial information services. As such this encompasses, from a technical point of view, lots of goals that are also relevant for the *plan4business* platform. Since the results of the INSPIRE initiative will prospectively have legal character for public authorities, the *plan4business* data model is designed with compatibility to that future standard in mind.

The INSPIRE data specifications encompass 34 themes[4] for each of which a detailed data model is described by the aid of extensive UML models. Amongst other applications, these UML models are used to derive GML based application schemes that can be used to exchange data between different systems or software components in a standardized way. As basis for the work in *plan4business*, version 3.0rc2[5] of the INSPIRE data specifications were used. Later on the models were subject to minor updates based on version 3.0rc3 of the INSPIRE data models that was published during the project runtime. At the time of this deliverable the INSPIRE data specifications have just been released in version 3.0, though the corresponding data models are not made available yet.

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

[4] An overview of the themes is given under: http://inspire.jrc.ec.europa.eu/index.cfm/pageid/2/list/7

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

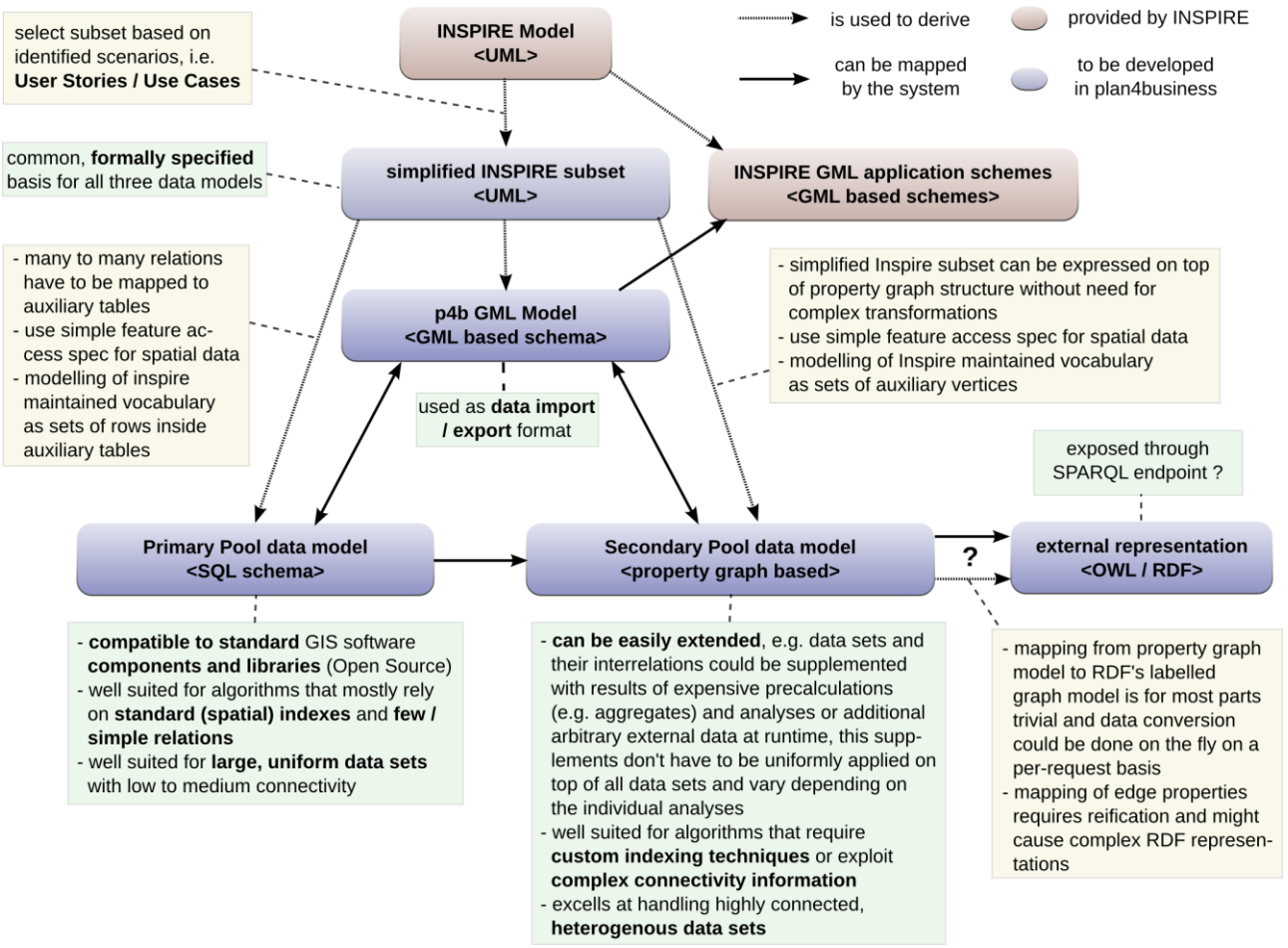

Figure 3 Relations between plan4business and INSPIRE data models for vector and metadata

4.2.1 Simplified INSPIRE Subset

The *plan4business* focuses on land use data as core planning data. Representing land use data from heterogeneous sources in a common harmonised data model is the foundation for enabling analyses across data sets. The *plan4business* data model for land use builds upon the INSPIRE Data Specification on Land Use, which was created by the experts of the INSPIRE Thematic Working Group Land Use. It not only provides a definition of common feature types, but includes also a classification system for land use, the *Hierarchical INSPIRE Land Use Classification System (HILUCS)*. Harmonising data sets to this common classification system makes them comparable – with limitations. The common classification system cannot overcome all differences in the various data sources, e.g. if original classifications are more general than the HILUCS classification.

As shown in Figure 3, the Simplified INSPIRE Subset is used as a logical basis for all other technology specific land use data models of *plan4business*. The model consists of the main features types and contains all mandatory components of the INSPIRE model, plus a few selected additional properties. The model was extended several times during the project, to cope with project requirements and changes to the INSPIRE data specification. Figure 4 shows an overview of the Simplified INSPIRE Subset model.

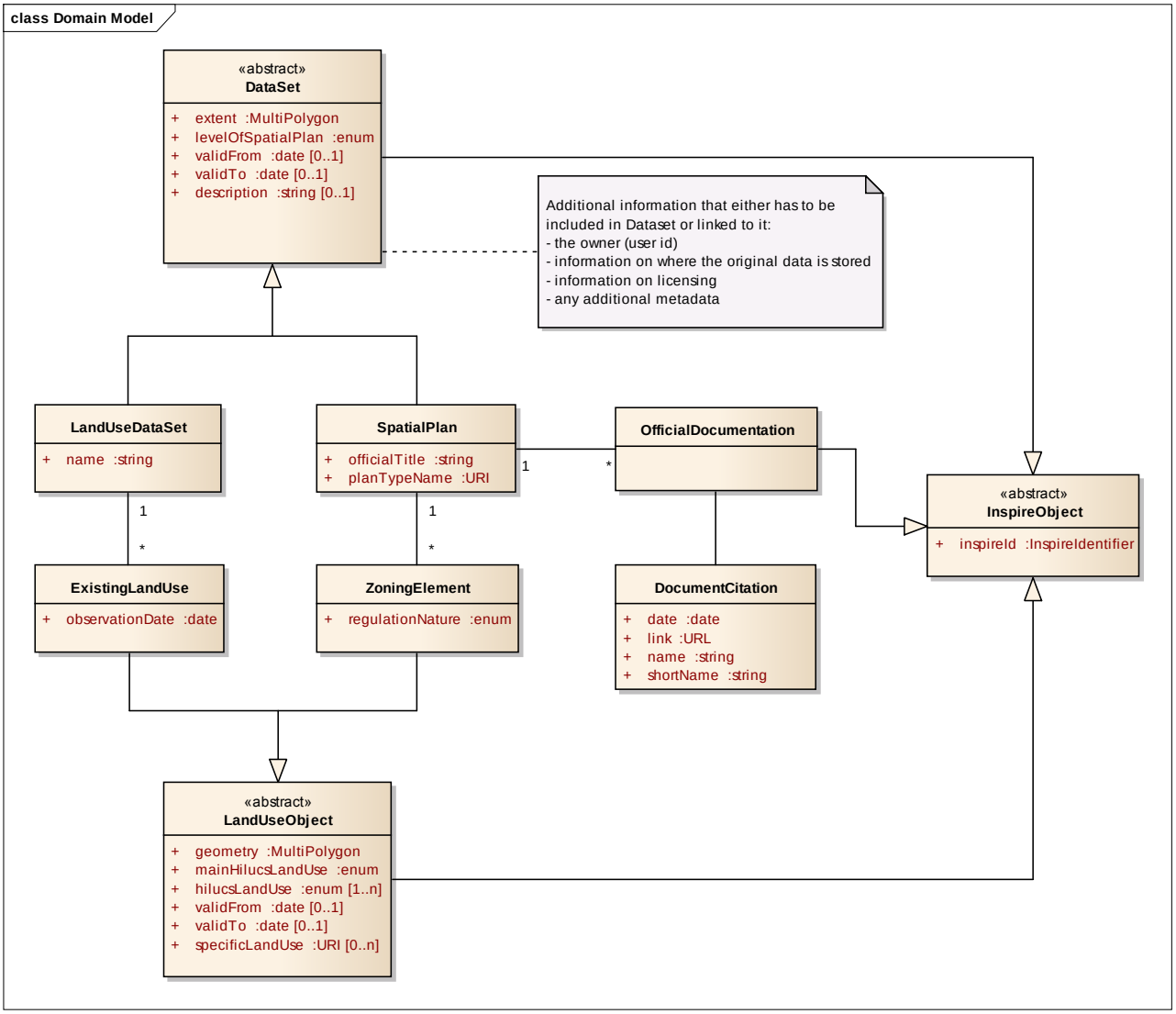

Figure 4 UML Overview of the simplified INSPIRE Subset that serves as basis for the plan4business Intermediate Model

4.2.2 The *plan4business* Intermediate Model

In a similar way as the GML based INSPIRE application schemes are derived from the INSPIRE UML Models, the *plan4business* intermediate model is derived from the selected simplified INSPIRE UML Subset described in the previous section. The Intermediate Model is defined as XML Schema. XML Schema is a schema format that is widely supported and has the additional benefit that the data can be represented as XML if needed. GML 3.2.1[6] is used to model the geometries, although the schema is not strictly a GML Application Schema. The *data sets* and *land use objects* are not modelled as GML feature types by intent, as this would introduce additional properties that are not part of the model but are defined in

[6] http://www.opengeospatial.org/standards/gml

AbstractFeatureType in GML. Figure 5, Figure 6, Figure 7 and Figure 8 showcase the structure of the XML Schema of the intermediate model for existing land use data sets and spatial plans and their dependent subtypes of *LandUseObject*. The schema for the intermediate model is published on the plan4business platform:

http://www.whatstheplan.eu/integrator/schemas/landuse/LandUse.xsd

The *plan4business* intermediate model is used as target schema for the integration of Shapefiles, XML, GML or CSV[7] based input data. The intermediate model can be used as *plan4business* native data format for import and export from, to and among components of the *plan4business* system. Once a data set is structured according to the *plan4business* intermediate model it can be transformed to the database specific models of the primary and secondary data pool. During the transformation the data is not actually encoded as XML in the intermediate model, but transformed directly in-memory and stored in the data pool (see section 4.3.2).

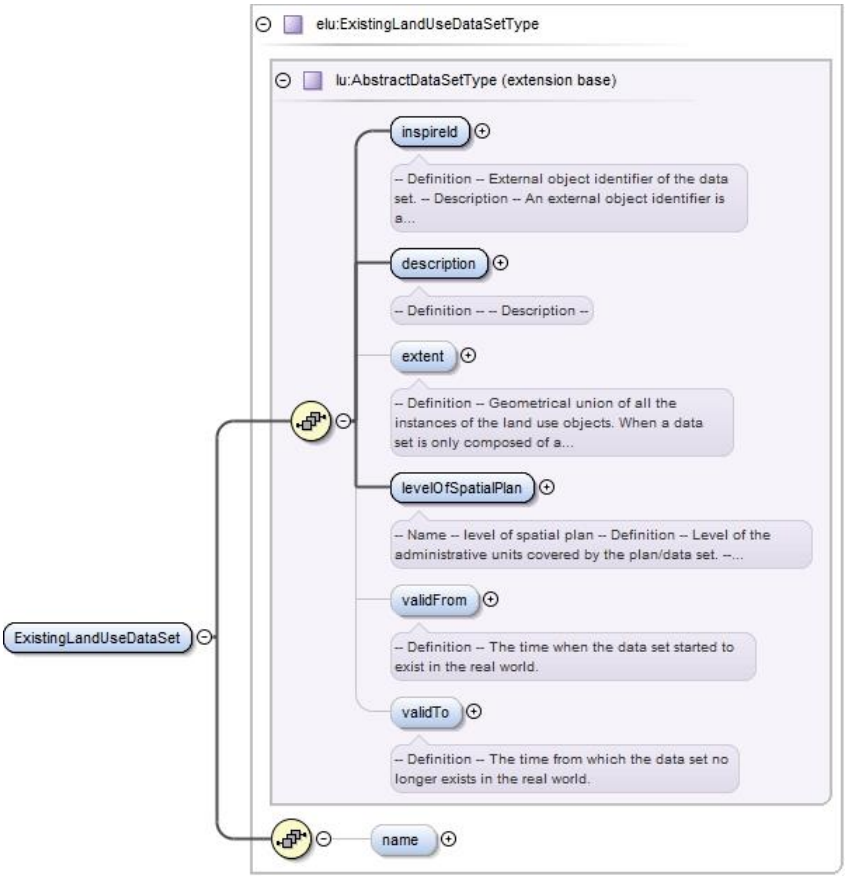

Figure 5 Representation of existing land use data set in the plan4business intermediate model

[7] Comma separated values (.csv), e.g. originating from spreadsheets or database export with support for OGC WKT expressions

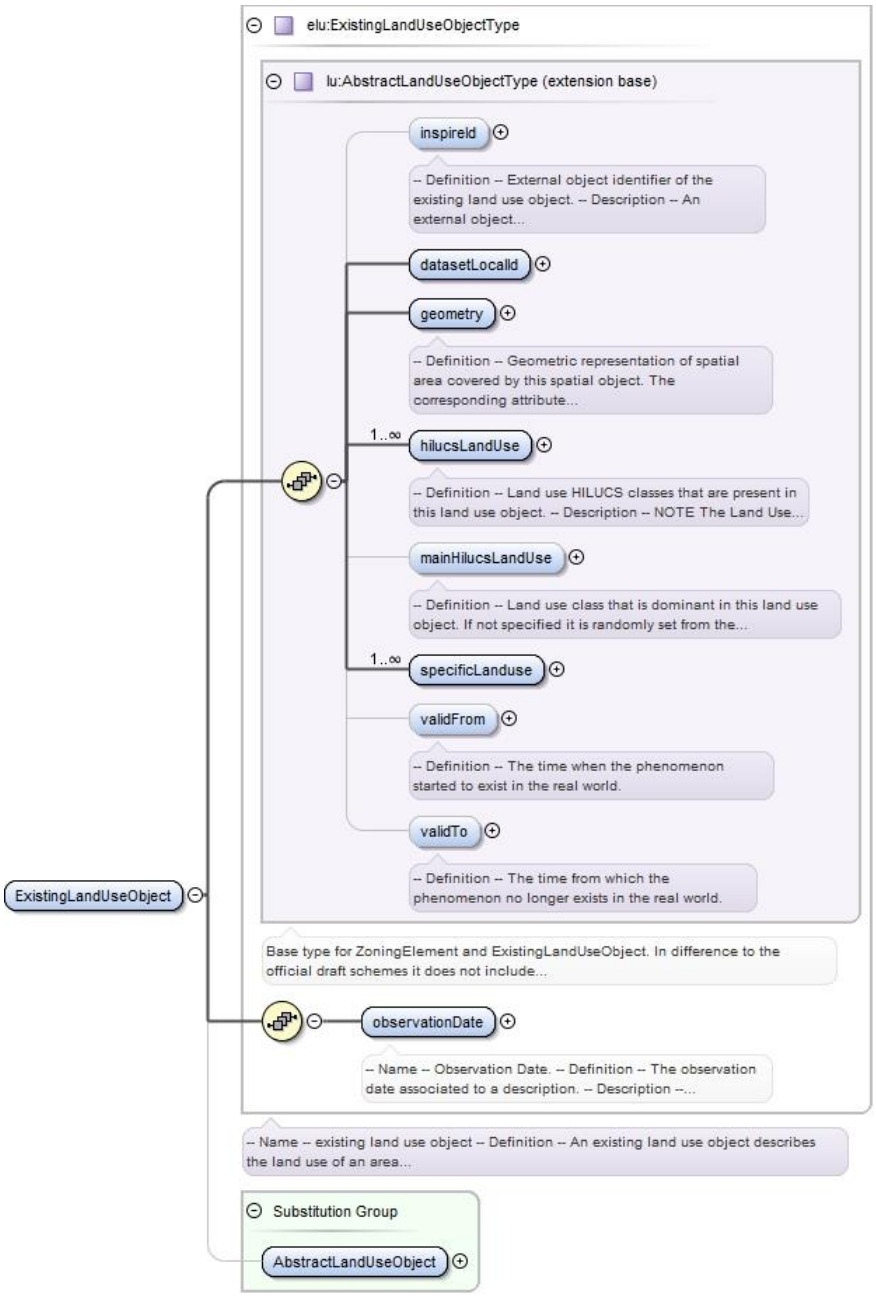

Figure 6 Representation of existing land use objects in the plan4business intermediate model

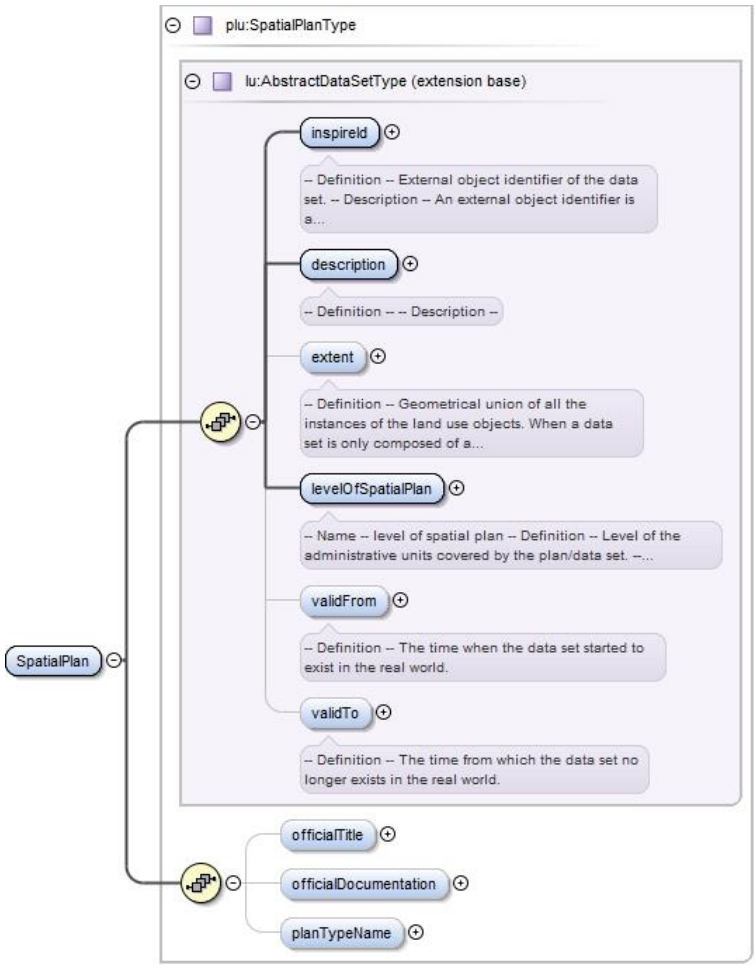

Figure 7 Representation of a planned land use data set in the plan4business intermediate model

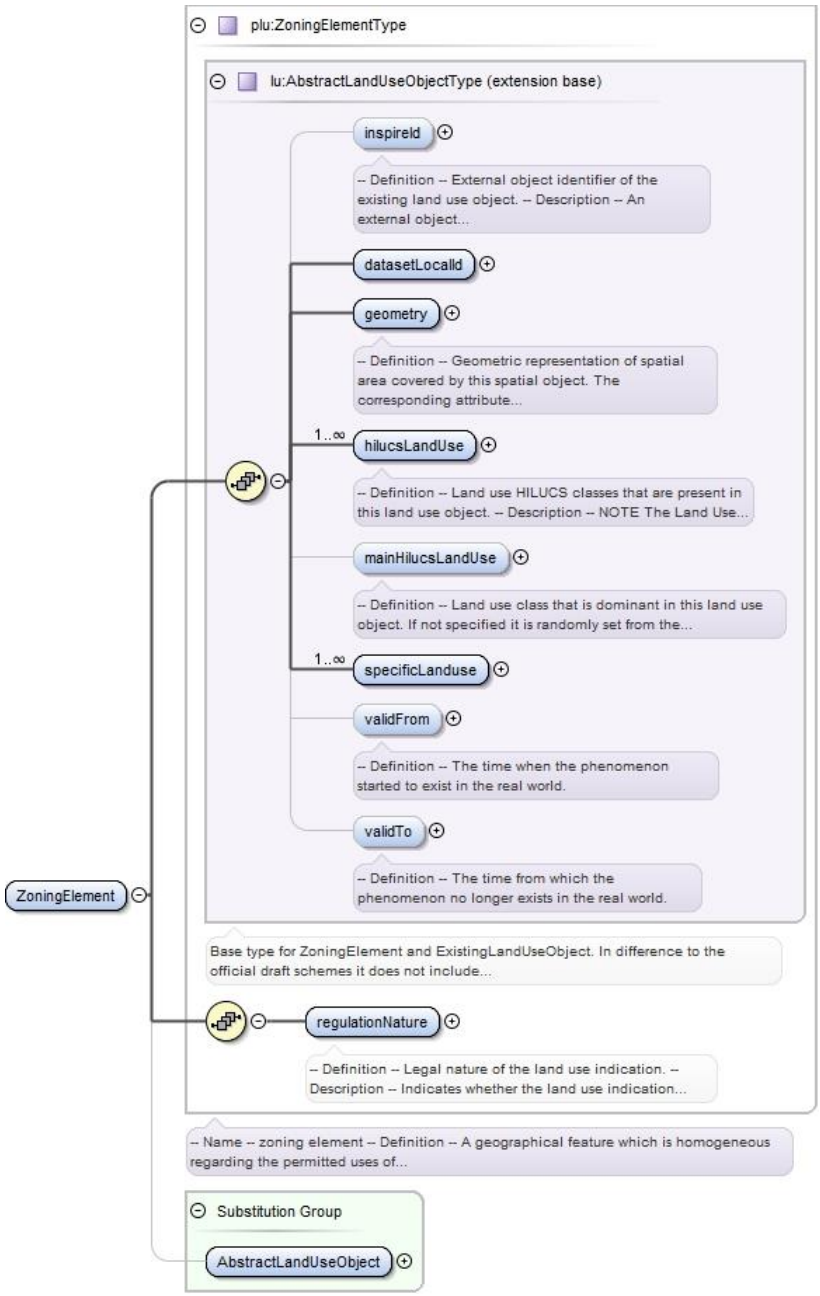

Figure 8 Representation of a zoning element of a planned land use data set

4.3 Data integration

Data sources that are uploaded using the Plan Integrator are to be integrated into the primary data pool. For this a schema mapping to the plan4business intermediate model has to be provided. This mapping can be created using the desktop tool HALE. The Plan Integrator allows uploading a mapping project and associating it with a data source. The transformation service of the Integration Engine transforms the source data to the intermediate model based on the given mapping. Once the data is represented in the intermediate model it is analysed and validated before it is stored in the primary data pool. The transformation from the intermediate model to the primary data pool is independent of the source data and relatively trivial as both models are very similar.

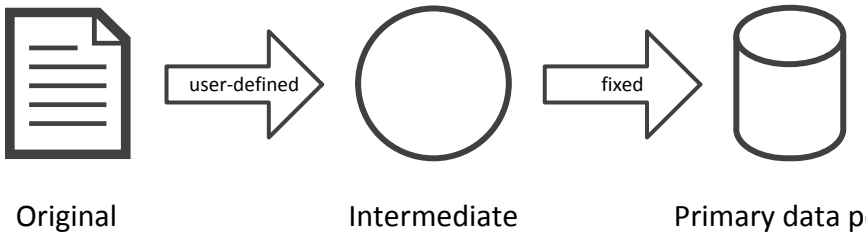

Figure 9 Transformation between source data, intermediate and primary data pool models

4.3.1 Schema Mapping

4.3.1.1 Template projects

In the Plan Integrator a user can download a template project for the schema mapping for each of his data sources. Each file in a data source is analysed and classified. Files with supported vector formats are automatically included in the mapping project, as source schema and/or data. The template project also includes the plan4business intermediate model as target schema and a style for visualising the land use objects according to their main HILUCS classification. The template project serves as a starting point for the schema mapping. It provides all necessary prerequisites and the user only needs to define the relations between the source and target schema entities and upload the completed mapping again to the Plan Integrator.

4.3.1.2 HUMBOLDT Alignment Editor

The HUMBOLDT Alignment Editor (HALE)[8] is free and Open Source desktop software for creating and evaluating schema mappings that is used to define mappings from the heterogeneous models of planning data to the plan4business intermediate model. Creating a mapping usually is done based on the template project provided by the Plan Integrator. HALE allows to browse source and target schemas, and to define relations between source and target classes and properties. The declarative approach used in HALE for the mapping allows an easier understanding of the mapping itself, and the interactive approach provides direct feedback: While creating the mapping, sample data is transformed inside HALE based on the current

[8] http://www.dhpanel.eu/humboldt-framework/hale.html

mapping. Thus the user gets feedback on the state of his mapping and the potential results even before he uploads the data and the actual transformation in the Integration Engine takes place (see also Figure 12). Constraints defined in the intermediate schema serve to validate the transformation samples to point the user to possible problems or mismatches.

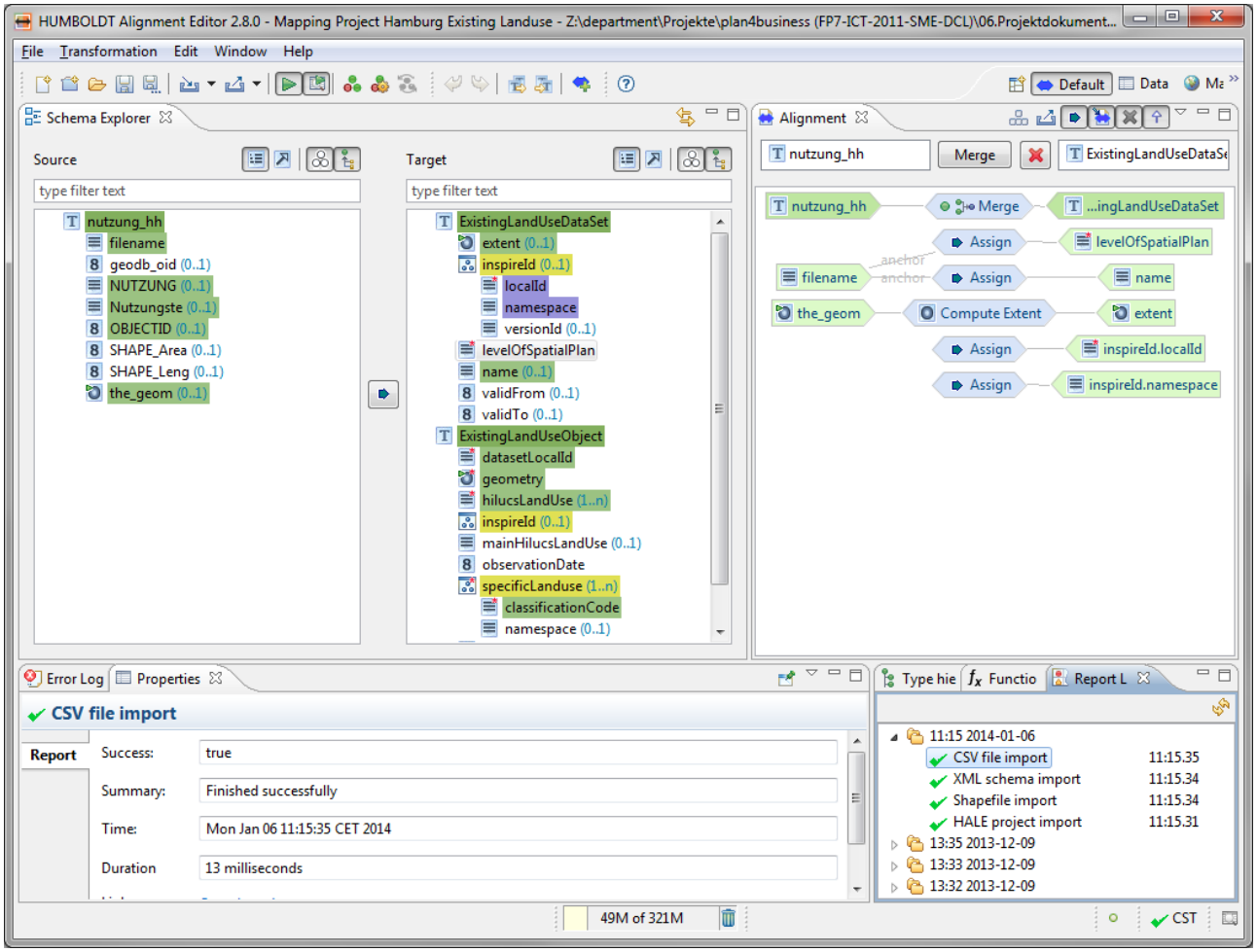

Figure 10 Default mapping view in HALE

4.3.1.3 HALE plan4business Edition

To guide the user through the creation of the mapping, an adapted version of HALE was created – the *HALE plan4business Edition*. It is provided through the plan4business platform and apart from providing the intermediate model for offline usage it mainly adds a guide on how to proceed with the mapping and what information should be provided. The guide is provided in form of a so called cheat sheet where the user is asked to perform a set of tasks (see Figure 11). Most tasks offer assistance to the user by semi-automatically performing the task – this can be triggered through the *Click to perform* option. As an example, tasks that relate to mapping a certain target class or property set it as target for the mapping and may even provide a suggestion for the corresponding source schema entity to select.

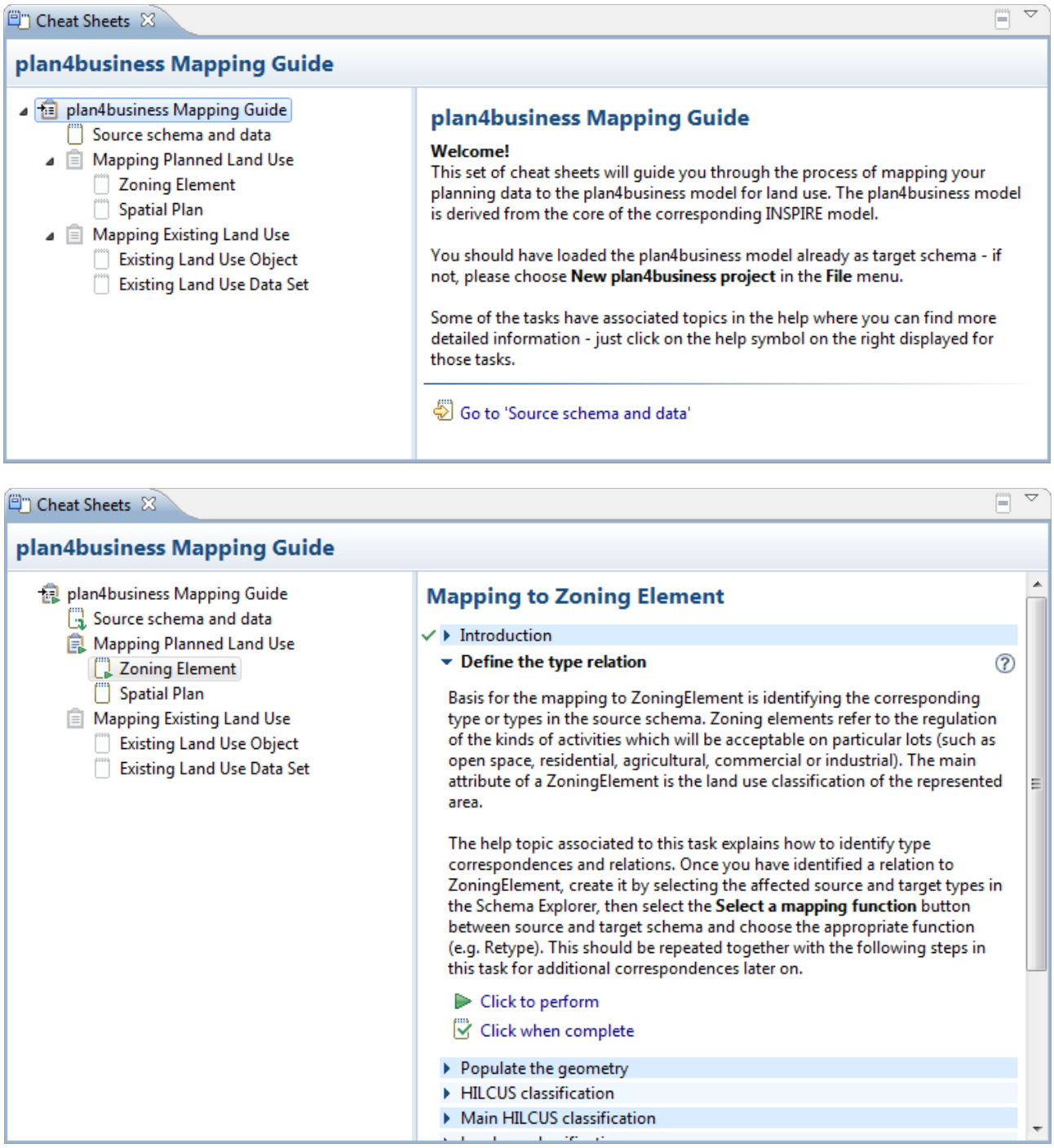

Figure 11 Mapping Guide for the intermediate model in HALE plan4business Edition

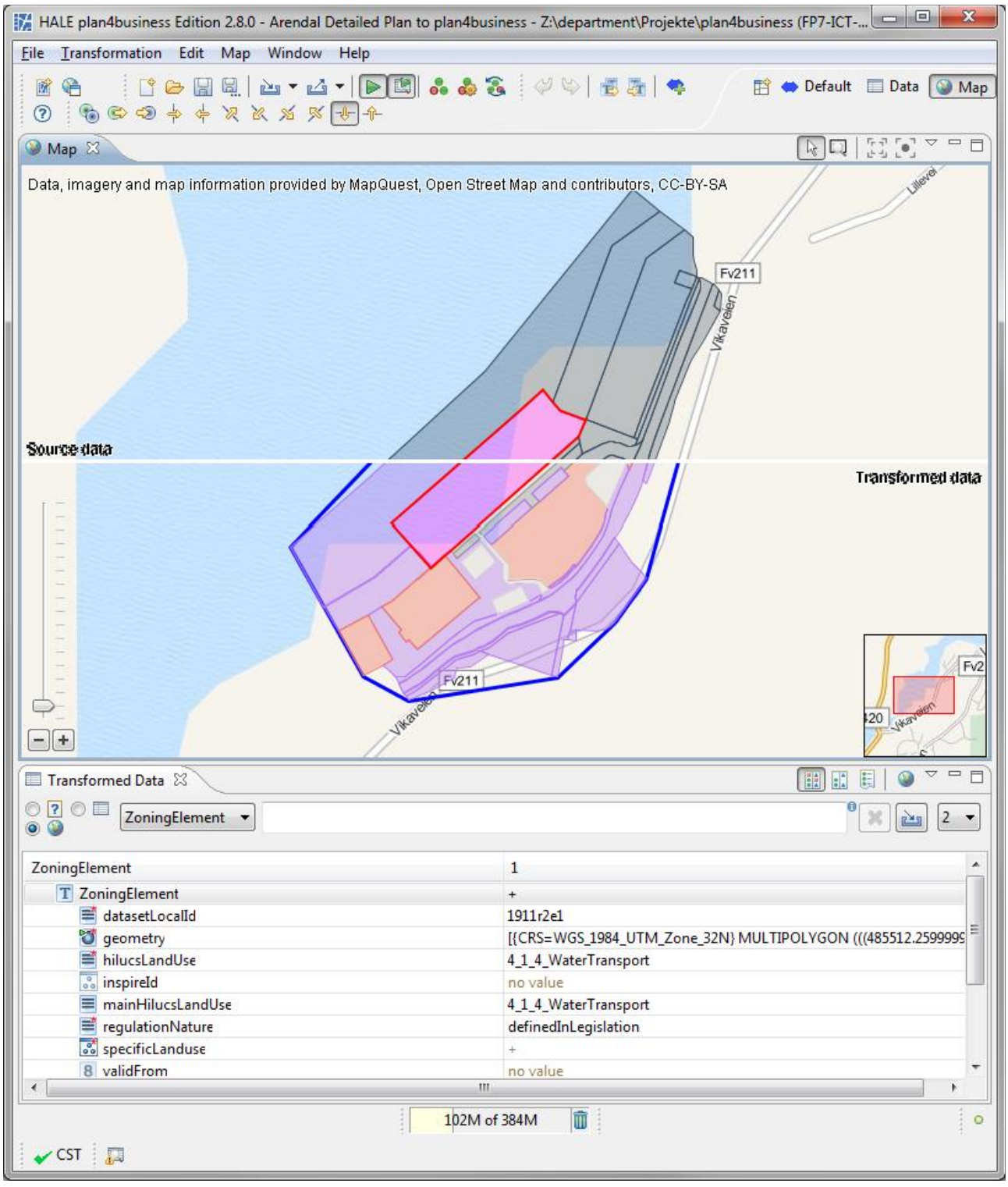

Figure 12 Transformation analysis based on sample data – the map shows original and transformed data

The basic steps for creating the mapping are:

1. Identify correspondences between source and target types/classes.
   Following the mapping guide the user goes through the different target types for planned and existing land use.
2. For each type correspondence
   a. define the corresponding condition and relation and
   b. identify the property correspondences.
   c. For each property correspondence define the corresponding condition and relation.

Figure 13 shows an exemplary mapping to *ExistingLandUseObject*. The source type *nutzung_hh* is mapped to *ExistingLandUseObject* with a 1:1 relation (*Retype*). In context of the type relation several relations between source and target properties are defined; e.g. a reclassification of the original code in *NUTZUNG* to the HILCUS classification in the *hilucsLandUse* property.

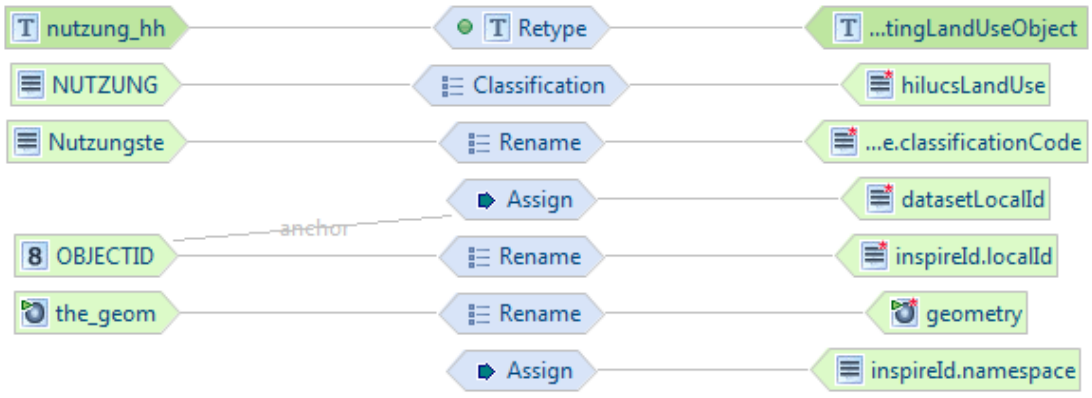

Figure 13 Example of a mapping to ExistingLandUseObject

4.3.2 Data Transformation

Based on a created schema mapping a data source is transformed to the intermediate model. The headless transformation API of HALE/CST is used for this purpose. It allows the execution of multiple transformations independently of each other, based on the respective schema mapping. Results of this transformation are HALE *Instances*[9] for each object, with an associated schema type, e.g. *ZoningElement*.

For the access to the primary data pool the Integration Engine uses the Java Persistence API (JPA). Thus the *Instances* resulting from the transformation are converted to the corresponding JPA Entities that represent the primary data pool model. During this process, the objects are analysed and validated.

The validation is performed by checking a transformed collection of data sets for completeness of attributes and consistency. Therefore, the instances are traversed and detected whether the mandatory attributes are

[9] http://hale.igd.fraunhofer.de/jenkins/job/hale-release-nightly/javadoc/eu/esdihumboldt/hale/common/instance/model/Instance.html

defined. The attribute values are checked also for consistency, e.g. conformity to the specified code lists. If any inconsistencies are detected, the process fails and returns a report including the inconsistent entities. Only if the mandatory properties are populated and the links between land use objects and data sets can be resolved the data will be saved to the primary data pool.

4.4 Data Publishing

In the Plan Integrator the user configures for each data source how the corresponding data sets should be published. Metadata for the integrated data sets is published to the Metadata Catalogue in any case; the user can choose if the data sets should be additionally made available as WMS and/or INSPIRE compliant GML.

4.4.1 Metadata

The metadata for each integrated data set is derived from the information the user provided for the uploaded data source and also information contained in the data set itself, e.g. the extent of the data set. Depending on the publishing options the metadata also links to the WMS layer and the GML download.

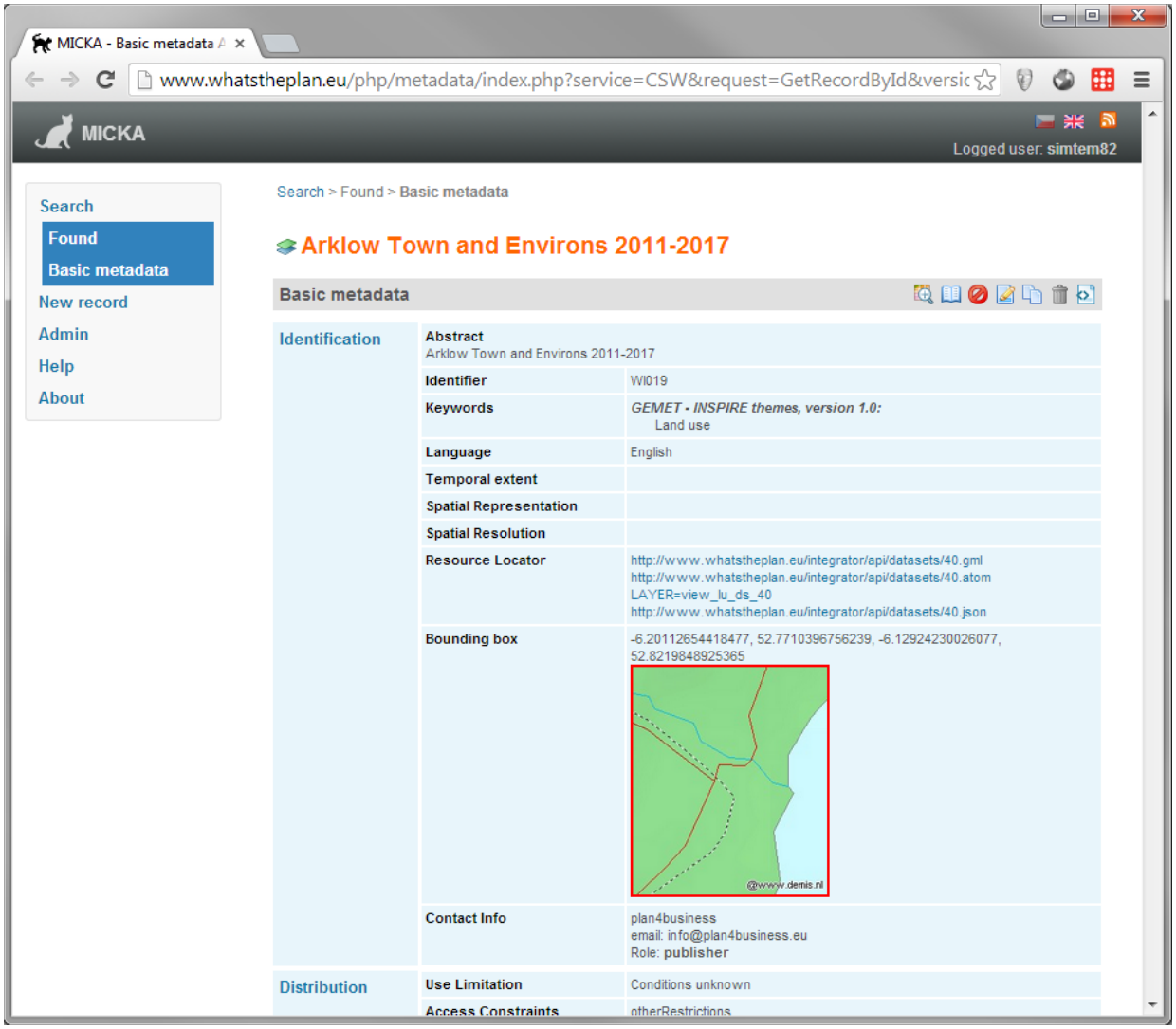

Figure 14 Metadata for a spatial plan published in MICKA

4.4.2 WMS

For viewing the integrated data the data sets can be published as OGC Web Map Service (WMS). The WMS is provided through the GeoServer part of the plan4business platform. The Layer Manager API (see section 4.5.2) is used by the Integration Engine to publish existing and planned land use data sets as individual layers. If a data set is published as WMS, the corresponding reference to the service is included in the metadata.

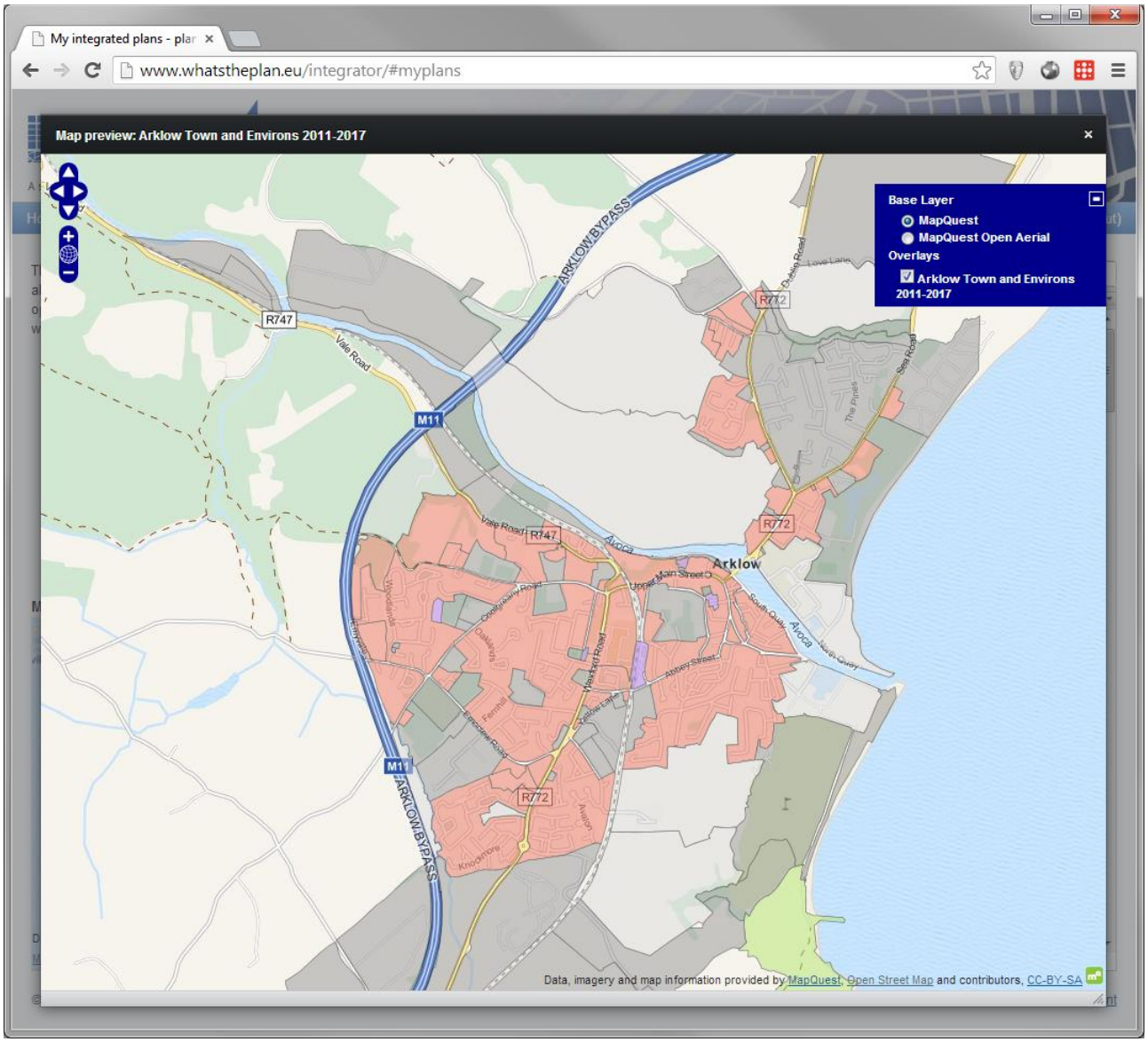

Figure 15 Map preview for a spatial plan in the Plan Integrator based on the corresponding published WMS

4.4.3 INSPIRE compliant GML

The Integration Engine also features a conversion from the primary data pool model to the existing and planned land use INSPIRE application schemas (Version 3.0rc3). Thus integrated data sets can be made available as INSPIRE compliant GML. It is available for all data sets where the owner enabled the publication for INSPIRE. The data is also offered for download in an Atom based pre-defined data set download service – see section 4.5.1 for more details.

```
5120    </ns1:member>
5121    <ns1:member>
5122      <plu:ZoningElement gml:id="Zoning_142">
5123        <plu:inspireId>
5124          <base:Identifier>
5125            <base:localId>WI019-26513</base:localId>
5126            <base:namespace>eu.plan4business:plu:ireland</base:namespace>
5127            <base:versionId nilReason="unknown" xsi:nil="true"></base:versionId>
5128          </base:Identifier>
5129        </plu:inspireId>
5130        <plu:geometry>
5131          <gml:MultiSurface gml:id="_34dd758e-f16e-47f8-8fd5-09f3239970a4" srsName="EPSG:3857">
5132            <gml:surfaceMember>
5133              <gml:Polygon gml:id="_15e7b4cf-300d-486c-b1e1-aa63d77c369a" srsName="EPSG:3857">
5134                <gml:exterior>
5135                  <gml:LinearRing>
5136                    <gml:posList>-686379.1133456456 6946010.761658652 -686370.4503359767 6946004.4998816205 -686364.6465787104 6946006.254444164 -68635
5137                  </gml:LinearRing>
5138                </gml:exterior>
5139              </gml:Polygon>
5140            </gml:surfaceMember>
5141          </gml:MultiSurface>
5142        </plu:geometry>
5143        <plu:validFrom>2011-04-05</plu:validFrom>
5144        <plu:validTo>2017-03-06</plu:validTo>
5145        <plu:hilucsLandUse xlink:type="simple" xlink:href="http://inspire.ec.europa.eu/codelist/HILUCSValue/5_ResidentialUse"></plu:hilucsLandUse>
5146        <plu:beginLifespanVersion xsi:nil="true"/>
5147        <plu:hilucsPresence xsi:nil="true"/>
5148        <plu:specificLandUse xsi:nil="true"/>
5149        <plu:specificPresence xsi:nil="true"/>
5150        <plu:regulationNature xlink:type="simple" xlink:href="http://inspire.ec.europa.eu/codelist/RegulationNatureValue/definedInLegislation"></plu:re
5151        <plu:processStepGeneral xsi:nil="true"/>
5152        <plu:backgroundMap xsi:nil="true"/>
5153        <plu:officialDocument xsi:nil="true"/>
5154        <plu:plan xlink:type="simple" xlink:href="#SpatialPlan"></plu:plan>
5155      </plu:ZoningElement>
5156    </ns1:member>
5157    <ns1:member>
5158      <plu:SpatialPlan gml:id="SpatialPlan">
5159        <plu:inspireId>
5160          <base:Identifier>
5161            <base:localId>WI019</base:localId>
5162            <base:namespace>eu.plan4business:plu:ireland</base:namespace>
```

Figure 16 Excerpt from a INSPIRE Planned Land Use GML file created from an integrated data set

Currently there are some limitations regarding the INSPIRE compliant GML:

- The GML is based on the INSPIRE Applications Schemas in version 3.0rc3, as the final versions were not available yet.
- Due to an error in the release candidate of the INSPIRE Base Types schema existing land use data sets cannot be served with a *SpatialDataSet* container, instead a GML *FeatureCollection* is used as container.
- Some properties cannot be properly populated yet, as the appropriate national and data provider specific code lists and registries are mostly not available yet (e.g. for the *planTypeName* property in *SpatialPlan* national code lists are foreseen).

4.5 REST API

4.5.1 Integration Engine API

The Integration Engine offers a REST API for access to data and information on uploaded data sources and integrated data sets. The API provides a description of the API in the Web Application Description Language (WADL[10]) and generated HTML documentation on all its resources and methods.

Table 1 Integration Engine API documentation locations in the plan4business platform

WADL description	http://www.whatstheplan.eu/integrator/api/application.wadl
HTML documentation	http://www.whatstheplan.eu/integrator/api/

Access control for the API is decoupled from the portal access control as the authentication is not performed through the Liferay portal (see D4.1.2). Instead the user can activate the API access by generating a specific API password in the Plan Integrator (see Figure 17). Authentication to the API is performed through HTTP Basic Authentication. Through the separate authentication token it is ensured that the API password cannot be used to login to the portal and change the account details or settings. Also, the user has control if the API access should be enabled at all.

Only the data sources related resources (see section 4.5.1.1) are currently protected through the authentication mechanism – all resources that are available for the integrated data sets are available publicly with anonymous access, with eventual limitations based on the corresponding publishing settings.

Figure 17 API password generation in the Plan Integrator

[10] http://www.w3.org/Submission/wadl/

4.5.1.1 Data sources and mapping

Through the Integration Engine API the user has access to his personal uploaded data sources and the associated mapping projects. It also allows the creation of new data sources and uploading schema mappings. For each data source the template project for creating the mapping with HALE is available as well. The following sections give an overview of the API resources.

4.5.1.1.1 User data sources

http://www.whatstheplan.eu/integrator/api/user/sources

This resource represents all data sources created by the authenticated user. The user can list his data sources or create a new data source.

Create a new data source (POST)

The user can upload files that form a new data source. Name and description of the data source can be provided as well.

Request query parameters:

parameter		description
name	(required)	The data source name for you to be able to identify it easily.
description		An optional data source description.
file	(required)	Files to upload as part of the data source.

Acceptable request representations:

- multipart/form-data

Available response representations:

- application/json

List user data sources (GET)

Lists all data sources created by the authenticated user. Each data source has its own resource where additional information can be retrieved (see next section).

Available response representations:

- application/json

4.5.1.1.2 Data source

http://www.whatstheplan.eu/integrator/api/user/sources/{id}

This resource represents a specific data source created by the authenticated user. The template parameter *id* is the system identifier of the data source.

Data source information (GET)

Gives detailed information on an individual data source. It also links to the associated mapping and integrated data sets.

Available response representations:

- application/json

4.5.1.1.3 Mapping project

http://www.whatstheplan.eu/integrator/api/user/sources/{id}/mapping

This resource represents the mapping project associated to a user's data source. The template parameter *id* is the system identifier of the data source.

Mapping project download (GET)

Download the mapping project archive associated to the specified data source.

Available response representations:

- application/zip

4.5.1.1.4 Mapping project template

http://www.whatstheplan.eu/integrator/api/user/sources/{id}/mapping/template

This resource represents the mapping project template that can be used as a starting point to create a mapping to integrate the data source into the plan4business platform. The template parameter *id* is the system identifier of the data source.

Download mapping template (GET)

Generates a mapping project archive for the specified data source. The project can be used to define a mapping for harmonising the data with the plan4business model.

Available response representations:

- application/zip

4.5.1.2 Integrated data sets

The Integration Engine API provides access to information on the platform's integrated land use data sets. If enabled, access to a data set as INSPIRE compliant GML through an INSPIRE pre-defined data set download service is provided as well. The following sections give an overview on the data set related resources in the API. Access to these resources is currently granted without authentication.

4.5.1.2.1 Data sets

http://www.whatstheplan.eu/integrator/api/datasets

This resource represents all data sets integrated into the plan4business platform. Depending on the accepted content type a different representation of the data sets is returned. Different representations can also be explicitly retrieved by adding an extension, e.g. **.atom** for the download service feed.

Basic data sets information (GET)

Lists all data sets with basic information and provides a link to a specific resource for each data set.

Available response representations:

- application/json (.json)

INSPIRE download service feed (GET)

Atom feed of the INSPIRE download service providing pre-defined data sets as INSPIRE compliant GML.
In the download service only data sets are available, for which it has been explicitly enabled by the data set owner.

Available response representations:

- application/atom+xml (.atom)

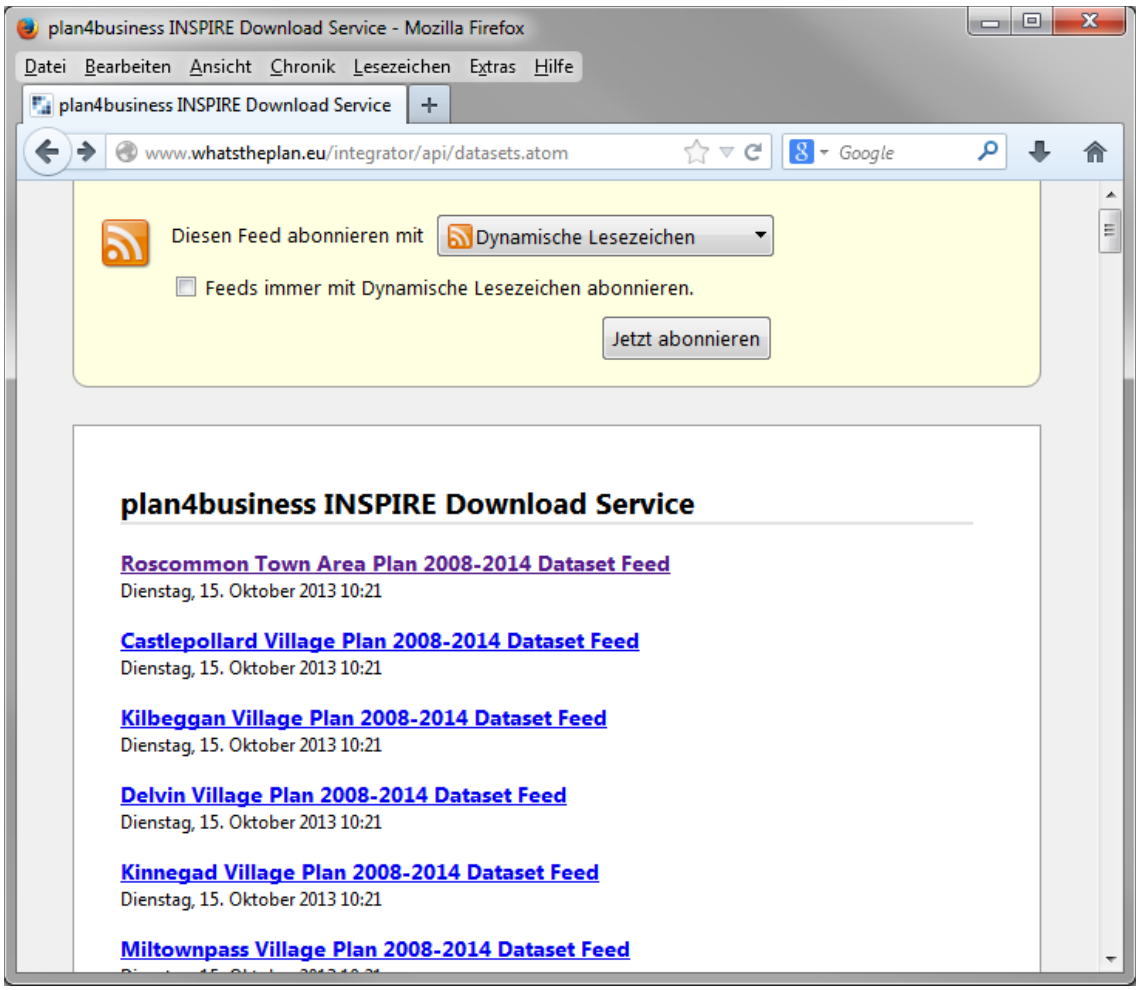

Figure 18 plan4business INSPIRE download service feed accessed with Mozilla Firefox

4.5.1.2.2 Data set

http://www.whatstheplan.eu/integrator/api/datasets/{id}

This resource represents a specific integrated data set in the plan4business platform. The template parameter *id* is the system identifier of the data set. Depending on the accepted content type a different representation of the data set is returned. Different representations can also be explicitly retrieved by adding an extension, e.g. **.atom** for the data set feed.

INSPIRE download service data set feed (GET)

Atom feed describing the pre-defined data set part of the plan4business platform INSPIRE download service. *Only available if the data set owner has enabled it.*

Available response representations:

- application/atom+xml (.atom)

Basic data set information (GET)

Provides basic information on the data set, e.g. the feature count.

Available response representations:

- application/json (.json)

INSPIRE pre-defined data set (GET)

Produces an INSPIRE compliant GML representation of the data set.
Only available if the data set owner has enabled it.

Available response representations:

- application/gml+xml (.gml)

4.5.1.2.3 Data set metadata

http://www.whatstheplan.eu/integrator/api/datasets/{id}/metadata

References the metadata of a specific data set in the plan4business platform. The template parameter *id* is the system identifier of the data set.

Retrieve metadata (GET)

Get the metadata document associated to the data set.

Available response representations:

- application/xml;charset=UTF-8

4.5.1.2.4 OpenSearchDescription

http://www.whatstheplan.eu/integrator/api/datasets/opensearchdescription.xml

This resource is the OpenSearch Description required for the plan4business INSPIRE download service feed.

Get the OpenSearchDescription (GET)

Available response representations:

- application/opensearchdescription+xml

4.5.1.2.5 Data set query

http://www.whatstheplan.eu/integrator/api/datasets/search

Executes search queries on datasets in the plan4business platform and lists the results. This is the backend for the basic search referred to in the OpenSearchDescription.

Query data set (GET)

Request query parameters

parameter		description
q	(required)	Search terms.

Available response representations:

- text/html

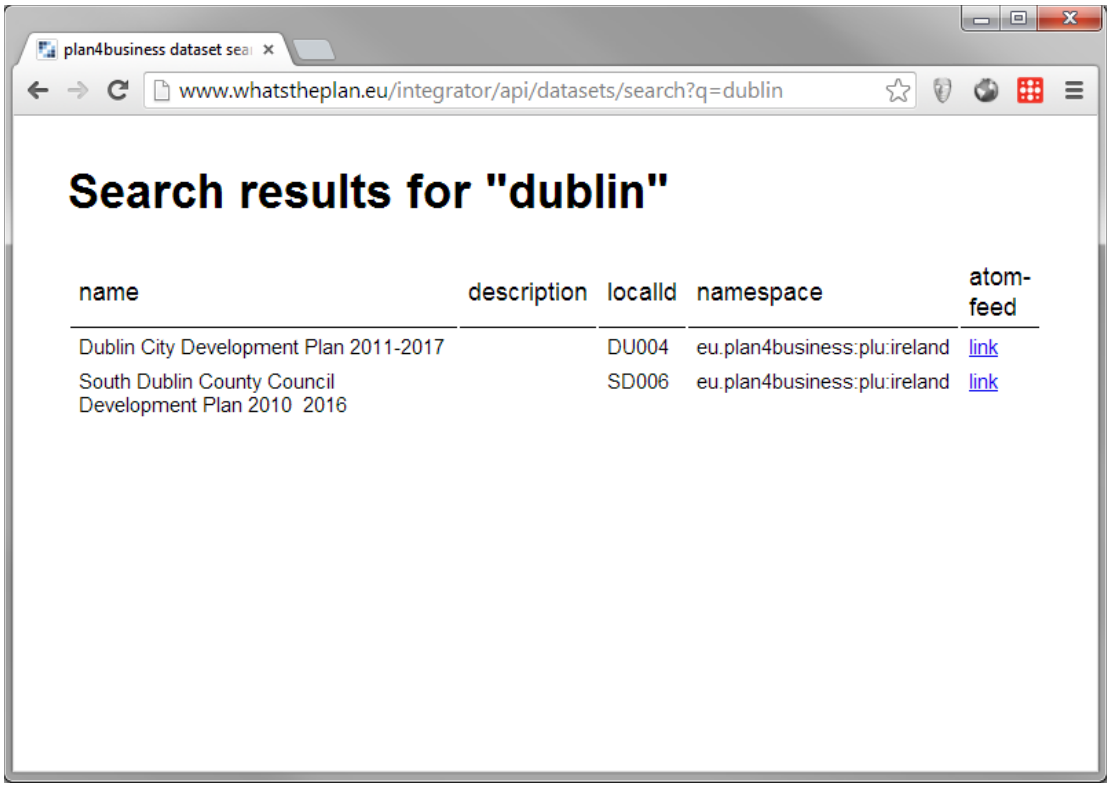

Figure 19 Example search results of a data set query

4.5.1.2.6 Data set feed query

http://www.whatstheplan.eu/integrator/api/datasets/search/atom

Looks up a specific data set feed in the plan4business platform. This is the backend for the identifier based feed search referred to in the OpenSearchDescription.

Query data set feed (GET)

Request query parameters:

parameter		description
spatial_dataset_identifier_code	(required)	Dataset identifier code.

spatial_dataset_identifier_namespace		Dataset identifier namespace.
language	(required) Default: *	Requested language of the feed.

Available response representations:

- application/atom+xml

4.5.1.2.7 Data set download query

http://www.whatstheplan.eu/integrator/api/datasets/search/download

Downloads a specific data set feed in the plan4business platform. This is the backend for the identifier based data set search referred to in the OpenSearchDescription.

Download data set (GET)

Request query parameters:

parameter	value	description
spatial_dataset_identifier_code	(required)	Dataset identifier code.
spatial_dataset_identifier_namespace		Dataset identifier namespace.
language	(required) Default: *	Requested language of the download.
crs	(required) Default[11]	Requested CRS of the download.

Available response representations:

- application/gml+xml

4.5.2 Layer Manager API

The Layer Manager is responsible for the management of data in context of the GeoServer installation that is part of the plan4business platform. The Layer Manager provides a REST API for data management and a web interface that is described in detail in D4.1.2. The Integration Engine makes use of the Layer Manager API to publish integrated land use data sets through GeoServer (see section 4.4). Following is an overview on the resources part of the Layer Manager API:

File Manager methods allow management of files in the secured file system. Every user has his/her private upload directory (Table 2).

[11] Default value is http://www.opengis.net/def/crs/EPSG/0/3857

/fileman

Method	Action	Return Code	Formats
GET	List the user directory	200	JSON
POST	Upload the file. If the file already exists, DO NOT overwrite	201 if created, 409 if the file already exists	
PUT	X	405	
DELETE	Delete all the files.	200	

/fileman/<file>

Method	Action	Return Code	Formats
GET	Get the file	200	JSON
POST	X	405	
PUT	Update the file. If the file does not exist, create it.	200	
DELETE	Delete the file	200	

/fileman/detail/<file>

Method	Action	Return Code	Formats
GET	Get the file details	200	JSON
POST	X	405*	
PUT	X	405*	
DELETE	X	405*	

* we may allow editing the details in future - e.g. set the file projection or add some attributes

Table 2 File Manager

The **Layer Editor** methods allow publishing and managing layers in PostGIS and GeoServer. Every group has its own publishing directory in the file system, database schema and GeoServer workspace (see Table 3).

/layed

Method	Action	Return Code	Formats	Parameters
GET	Get the list of layers	200	JSON	group
POST	Import corresponding file to database and publish in GS	201		group
PUT	X	405		
DELETE	X	405		

/layed/<layer>

Method	Action	Return Code	Formats	Parameters
GET	In future, this should return KML or WMS capabilities	200		
POST	X	405		
PUT	Update / overwrite the layer	200		group
DELETE	Delete layer	200		group

/layed/config/<layer>

Method	Action	Return Code	Formats	Parameters
GET	Get layer parameters	200	JSON	group
POST	X	405		
PUT	Set layer parameters	200	JSON	group
DELETE	X	405		

Table 3 Layer Editor

5 Conclusion

The Integration Engine provides the basic functionality needed to enable the integration of land use data sets into the plan4business platform. The INSPIRE based data model allows analysis across data sets and the provision of the data in a INSPIRE compliant way. The Integration Engine is closely linked to other components in the platform, foremost the Plan Integrator (Task 4.1), which serves as a User Interface for the configuration of the Integration Engine, and the Storage Engine (Task 5.3), where the harmonised data is stored. The deliverable on the Storage Engine (D5.4) gives an overview on which data has already been integrated into the platform.

Even though Task 5.1 is completed with project month 21, further development will take place for Integration Engine maintenance, for instance the update of the data model and INSPIRE GML generation to the final version (3.0) of the INSPIRE data specification on land use. Also, as Tasks 4.1 (Plan Integrator) and 5.3 (Storage Engine) are still ongoing, further development on the Integration Engine may be needed in context of these tasks, for instance an extension of the API and tighter integration with HALE.

We want to follow up on the Integration Engine development even after the project and further develop the concepts for the data integration, especially on aspects regarding the usability and maintainability. To better maintain schema mappings we want to explore versioning of schema mappings and management, detection and comparison of schemata. Another aspect we want to look into is the mechanism for updating data from external sources in a centralised platform like the plan4business platform.