DATA INTEGRATION 2

MARCH 2017

DELIVERABLE

Project Acronym: **SDI4Apps**

Grant Agreement number: **621129**

Project Full Title: **Uptake of Open Geographic Information Through Innovative Services Based on Linked Data**

D5.3.2 DATA INTEGRATION 2

Revision no. 03

Authors: Otakar Čerba (UWB)

Jáchym Kellar (UWB)

Dmitrii Kožuch (HSRS)

John O'Flaherty (MAC)

Martin Tuchyna (SAZP)

REVISION HISTORY

Revision	Date	Author	Organisation	Description
01	20/03/2017	Otakar Čerba	UWB	Initial draft
		Jáchym Kellar	UWB	
		Dmitrii Kožuch	HSRS	
		John O'Flaherty	MAC	
		Martin Tuchyna	SAZP	
02	25/03/2017	John O'Flaherty	MAC	Internal review
		Karel Charvát	CCSS	
03	31/03/2017	Otakar Čerba	UWB	Final Version

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Disclaimer:

Views expressed in this document are those of the individuals, partners or the consortium and do not represent the opinion of the Community.

TABLE OF CONTENTS

LIST OF TABLES

EXECUTIVE SUMMARY

This report describes particular data integration issues in the SDI4Apps project. The particular activities of data integration and harmonization are related with a requirement of combining data from heterogeneous resources, re-using of existing data or publishing data as open data or Linked data. Due the fact, that particular data integration activities vary in particular pilots, the sections of this document are connected to SDI4Apps pilots – Easy data access, Open Smart Tourist Data, Open Sensors Network, Open Land Use Map Through VGI, Open INSPIRE4Youth and Ecosystem Services Evaluation.

1 INTRODUCTION

The potential of geographic information (GI) and geographic (or spatial) data collected by various actors ranging from public administrations to voluntary initiatives of citizens is not fully exploited. The advancements of information and communication technologies and shift towards Linked Open Data (LOD) give an excellent foundation for innovation based on reuse of geographic information and data. The establishment of spatial data infrastructures has largely been driven by the "traditional" GI community and the national and European policies governing this sector. However, GI is no longer a separate information space but finds itself part of a larger European information space where the ultimate objective is the creation of value-added services based on reuse of public sector information as defined by the PSI and INSPIRE directives rather than exchange of "layers" between different GI software.

Establishing an infrastructure to meet this new and wider objective puts greater strain on local authorities and institutions that traditionally were users of GI but now find themselves in an environment where they are also expected to be data and service providers, a role that is far more demanding in terms of technical knowledge and resources.

The main target of SDI4Apps is to build a cloud based framework that will bridge the gap between:

1. the top-down managed world of INSPIRE, Copernicus and GEOSS, built by SDI experts, and

2. the bottom-up mobile world of voluntary initiatives and thousands of micro SMEs and individuals developing applications (apps) based on GI.

SDI4Apps is adapting and integrating experience from previous projects and initiatives such as HABITATS, Plan4business and EnviroGrids, to build its cloud based platform with an open API for data integration, easy access and provision for further reuse. The solution has been validated through six pilot applications focused on easy access to data, tourism, sensor networks, land use mapping, education and ecosystem services evaluation.

This deliverable describes particular data integration issues in the SDI4Apps project. The activities of data integration and harmonization are related with a requirement of combining data from heterogeneous resources, re-using of existing data or publishing data as open data or Linked data.

Due the fact, that particular data integration activities vary in particular pilots, the sections of this document are connected to SDI4Apps pilots – Easy data access, Open Smart Tourist Data, Open Sensors Network, Open Land Use Map Through VGI, Open INSPIRE4Youth and Ecosystem Services Evaluation. This is conducted in cooperation with dissemination activities, organised hackathons and stakeholder management group.

2 DATA INTEGRATION 1 (SHORT SUMMARY)

This deliverable follows the document D5.3.1 Data Integration 1 published in March 2016. According to the Description of Work of the SDI4Apps project the second version of the data integration report should be "an updated list of Linked Open Data in the SDI4Apps cloud".

The previous report was divided into six parts corresponding to particular pilots:

1. Easy Data Access

2. Open Smart Tourist Data

3. Open Sensors Network

4. Open Land Use Map through VGI

5. Open INSPIRE4Youth

6. Ecosystem Services Evaluation

Each sub-chapter contained examples of data integration and harmonization activities which were realized for pilot purposes. They represent cases of spatial data harmonization and integration processes, which could be re-used (probably with minor changes) for similar task combining data and information from heterogeneous resources, including Linked data.

The Table 1 describes the data and harmonization processes published the previous report D5.3.1.

Data	Harmonization process	Pilot
ETIS (European Tourism Indicator System) for the Sustainable Management of Destinations to monitor and measure performance of destinations in Europe	• Transformation from Excel/CSV to RDF format (fitting for LOD purposes)	Easy Data Access
Potential Monuments Voluntary Geographic Information Dataset – dataset to record Voluntary Geographic Information (VGI) reports from professionals, visitors and people interested in their local heritage, to seek out and groundtruth potential Monument sites in the Burren and beyond.	-	Easy Data Access
Data inputs for Smart Points of Interest – OpenStreetMap, GeoNames.org, Citadel on the Move, Natural Earth, OpenTrasportNet (data from Antwerpen and Issy), experimental ontologies, local data from Sicily or Posumavi region	• Transformation to SPOI data model • Filtering • Format conversion • Re-classification • Transcription to a structured form	Open Smart Tourist Data
Open Land Use Map – a land use map with as much detail as possible, and covering the whole EU territory	• Geometry composing • Re-classification	Open Land Use Map through VGI

	• Transformation to OLUM data model	
Datasets related to environmental protection – CORINE land cover, Protected sites, Urban Atlas, Open Street Map, relevant data from Danube reference data and service infrastructure (DRDSI) catalogue, Land Use and Coverage Area frame Survey (LUCAS), statistical data.	• Development of maps and data compositions • Filtering and re-classification	Ecosystem Services Evaluation

Table 1: Spatial data and relevant harmonization processes (summary of D5.3.1 report).

3 UPDATED LIST OF INTEGRATED DATASETS

During the last year (March 2016 – March 2017) there were several new data sets integrated to particular pilot solutions (Table 2):

Data	Harmonization process	Pilot
Irish National Monuments Dataset	• Implementation of data with using the standardized INSPIRE Protected Site Theme – PS v3.2 data model	Easy Data Access
External Open Datasets (Logainm Placenames Database of Ireland, Irish Open Government Data Portal, National Biodiversity Data Centre Ireland, EU Open Data Portal, Eurostat Linked Data, Open Street Map, GeoNames.org)	• Automatically used (as they are provided; without any spacial harmonization steps)	Easy Data Access

Data from Belluno (partner of the Foodie project)	• Original data: Shapefile
	• Transformation process:
	  ◦ Transformation of Shapefile to CSV table
	  ◦ Modification of the content (for example replacing empty lines etc.)
	  ◦ Filtering of data and transformation to RDF (SPOI data model)
	• Transformation tools: Bash, QGIS, spreadsheet software (Libre Office Calc), text editor (gedit)
		Open Smart Tourist Data (SPOI)

Wrecks and Obstructions Database	• Original data: KML/KMZ.
	• Transformation process:
	  ◦ Downloading and unzipping original data.
	  ◦ Transformation to RDF (SPOI data model) using PHP.
	• Transformation tools: PHP, text editor.
		Open Smart Tourist Data (SPOI)

Open-Data-Portal of Deutsche Bahn AG	• Original data: compressed GeoJSON • Transformation process: ◦ Downloading and unzipping original data.	Open Smart Tourist Data (SPOI)

	○ Transformation to RDF (SPOI data model) using PHP. • Transformation tools: PHP, text editor	
Wikidata (because of changes of download services, the current version of SPOI does not contain Wikidata features; the new harmonization process is under development, now.	• Original data: RDF (the list of countries and types of Wikidata object to download are stored in CSV tables, the result of query combined country and type of the object is provided as JSON file). • Transformation process: ○ Development of lists of countries and types of object to be processed. ○ Creation of relevant query for each combination of a country and a type of the object. ○ Downloading JSON file with IDs of relevant objects. ○ Transformation of JSON file to temporal text file. ○ Downloading of RDF file for each object. ○ Transformation to RDF (SPOI data model) using XSLT template. • Transformation tools: Bash, Java, Saxon 9 HE, text editor (gedit), spreadsheet software (Libre Office Calc), wget.	Open Smart Tourist Data (SPOI)
• Corine Land Cover 2012 • Urban Atlas • GRBGis – Flemmish national geodatabase • Czech digital cadastre • Austrian LPIS • Czech LPIS • Birmingham Historical Land Use Survey • Eurostat – municipality borders • Austrian municipality borders	• Original data: SHP, GML • Transformation process: ○ Downloading and/or unzipping data ○ Import data to PostgreSQL database ○ Transform data attributes and geometry – SQL script ○ Calculate Open Land Use Map – SQL script • Transformation tools: Python, PostgreSQL with extension PostGIS	Open Land Use Map Through VGI (Open Land Use Map)
Sensor catalogue IoT Discovery (Only sensor catalogue was prepared via sensor metadata importing to IoT Discovery component. This IoT Discovery component provides storage of	• Original data: CSV, Excel Sheets • Transformation process: ○ Downloading and/or unzipping data	Open Sensor Network (IoT Discovery, IoT Discovery View)

semantic description of sensors available by described endpoints.)	○ Transforming to RDF (IoT-A model) by XSLT ○ Importing to IoT Discovery by REST API - Transformation tools: Text editor, XSLT processor, REST client	
- Corine land cover - Open Street Map - Administrative units - LUCAS	- Original data: CSV, ESRI SHP, ESRI, Geodatabase. - Transformation process: loading original data into predefined tables in PostgreSQL database, executing land use calculation function, mapping to INSPIRE Existing Land Use data model, publishing via WFS API. - Transformation tools: shp2pgsql, SQL Editor, GeoServer APP Schema	Ecosystem Services Evaluation (Open Land Use Map encoded in INSPIRE ELU application schema for Slovakia, Hungary, Serbia and Croatia
- Corine land cover - Protected sites - Administrative boundaries	- Original data: WFS, GML, SHP - Transformation process: In order to create maps for selected ESS, including their transformation into the map of the overall assessment of the ESS set of processing routines documented via ESS process model, had to be defined, tested and implemented. Each map was prepared based on input data following the rules defined in a set of tools has been applied, resulting in the layer with coverage spatial representation (ESRI Grid) layer. Each target layer expresses volume of the benefits provided by a given ecosystem service in physical units as they were evaluated. E.g. m3 of wood (paper pulp) or m3 of carbon in the layer of carbon sequestration in different types of landscape, therefore these layers have been identified as basic ones. - Transformation tools: GIS ArcGIS suite	Ecosystem Services Evaluation

Table 2: Spatial data and relevant harmonization processes (new datasets not included to D5.3.1 report).

4 CONCLUSION

This deliverable is the final document focused on data integration and harmonization activities in the SDI4Apps project and its pilots. In the first part there are summarized results (data and harmonization processes) published in the previous report D5.3.1 Data Integration 1. Then the updated information divided into parts corresponding to the project pilots follows. The particular sections representing pilot activities of the project seems to be a bit unbalanced, but it is necessary to mention that this fact reflects the complexity of the data integration issues on the one hand and on the second hand the pilots are very different from the view of progress as well as dealing with the data and need of integration.