INSPIRE and VGI: Species Distribution in Europe

Umberto Di Staso[1], Matteo Mattassoni[1], Raffaele De Amicis[1]

[1] Fondazione Graphitech, via alla Cascata 56/c, 38123 Trento Italy.

Abstract. The growth of the Internet broadband has modified our needs, rising new types of stakeholders, markets and opportunities, especially in the mobile field.

Nowadays, national environmental agencies can share their own data between each other, governments can better monitor their country and civil protection can operate its action plans in a more efficient and quick way.

However, what happens if geographical data have to be used in operational activities, between European Member States? To answer this question, starting from the 2007, the INSPIRE directive tries to address this issue, with the build of 34 spatial data themes for environmental spatial information and, in general, a directive to set up the European Spatial Data Infrastructure. Thanks to this directive, member states have the instruments to exchange and use huge amount of environmental data between each other but, what happens if, following the new social trends of crowdsourcing, we put another variable constituted by normal citizens in the loop?

The creation of new INSPIRE compliant spatial datasets requires specific technical and methodological skills as well as a deep understanding of the directive. On the other hand, nowadays citizens can be seen as mobile sensors (Goodchild, M. F. 2007), and their contribution in the creation of spatial data can be fundamental.

The aim of this paper is to offer a possible solution tuned on the INSPIRE Specie Distribution data theme but applicable on all the others environmental datasets involved in the European directive, to create high quality INSPIRE compliant data, exploiting the users contributions through collaborative information collections (VGI).

Published in "Proceedings of the 1st ICA European Symposium on Cartography", edited by Georg Gartner and Haosheng Huang, EuroCarto 2015, 10-12 November 2015, Vienna, Austria

Keywords. INSPIRE, crowdsourcing, specie distribution, mobile devices, European SDI, eENVPlus, VGI.

1. Introduction

The eENVplus[1] project (Attardo C, Saio G. 2013) aims to unlock huge amounts of environmental data, managed by the involved national and regional environment agencies and other public and private environmental stakeholders, through the integration and harmonisation of existing services. These data are not only collected to answer reporting obligations on the environment to the European Union, but also to support national and local policies and actions.

The project does not design new services but rather, starting from the results of previous European experiences (funded projects, best practices, EU and national and local experiences), it integrates existing infrastructures into an operational framework able to overcome cross-border and language barriers.

eENVplus provides not only the ICT infrastructure but also the description and the support to make this infrastructure operational and profitable through the provision of an organisational model and a tutored training framework.

Figure 1 represents the ICT infrastructure provided within the eENVplus project: how it is possible to see, it is composed by a set of individual components, based mainly on free and open source software, but also new components developed from scratch, that are able to talk each other taking the advantages of the use of standard protocol for data exchange, download and process.

Services involved in the eENVplus infrastructure can be grouped in four main categories:

- Services for data ingestion dissemination [Figure 1 – orange blocks];
- Services for data processing [Figure 1 – purple blocks];
- Services for data cataloguing [Figure 1 – green blocks];
- Services for crowdsourcing [Figure 1 – blue blocks].

In order to unlock the mentioned huge amount of data, the ICT infrastructure provided contains a specific module, designed from scratch, whose role

[1] www.eenvplus.eu

is to acquire, store and disseminate observations collected by users about different environmental topics:

- Specie distributions;
- Protected sites;
- Damage to the environment.

This paper focuses on the topic of species distribution in Europe.

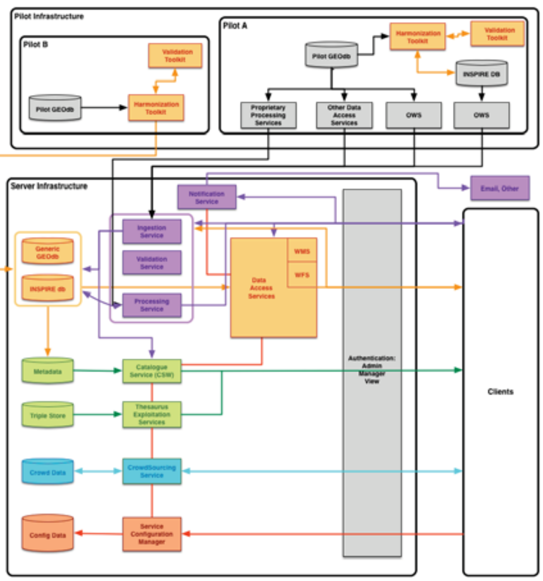

Figure 1 - The eENVplus Spatial Data Infrastructure.

The aim of this paper is to offer a possible solution, tuned on the INSPIRE Specie Distribution data theme but applicable for all the others environmental datasets involved in the European directive, to create high quality INSPIRE compliant data, exploiting the users contributions through collaborative information collections.

2. The Impact of INSPIRE

As (Masser 2015) reports, INSPIRE is *well transposed* by European member states, but a *gap* still exists to put the directive into operational environmental activities.

One of the main implementation barriers that are currently influencing the effectiveness of the directive is the data harmonization process.

The data harmonization process is a very time-consuming operation, directly influenced by the poor availability of personnel who has technical and legislative skills in order to produce INSPIRE-compliant datasets.

For the very specific requirements listed in the previous paragraph, from a study performed in the 2014, only the 20% of the geo-ICT companies have the skills required to work within INSPIRE (Vancauwenberghe, G., Cipriano, P., & Craglia, M. 2014).

The production of harmonized datasets is normally performed starting from existing datasets and using the so-called mapping tools, for exam- ple HALE[2] (Fichtinger, A., Rix, J., Schäffler, U., Michi, I., Gone, M., & Reitz, T. 2011).

As (Otakar Čerba, Karel Charvát, Karel Janečka, Karel Jedlička, Jan Ježek, Tomáš Mildorf 2012) reports, the harmonization process is composed by five actions, where first four of five are manually performed:

1. Theory of spatial data harmonization;
2. Source data understanding;
3. Target data understanding;
4. Definition of the necessary harmonization steps;
5. Practical realization.

The poor availability of professional users with the aforementioned technical skills may influence the production of harmonized datasets.

3. INSPIRE and VGI

The approach described in this paper enables the possibility to generate new harmonized datasets from non-professional users, taking advantages from the enormous resource of volunteered geographic information (VGI) (Flanagin, A.J.; Metzger, M.J 2008) (Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F. and Obersteiner, M. 2009), and the diffusion of INSPIRE into operative activities can be encouraged.

Although previous approach described in the literature (Wieldmann & Bernard 2014) mainly considers crowd sourced data as an improvement of already existing INSPIRE datasets through data fusion techniques, in this

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

paper we focalize our attention in generating completely new INSPIRE compliant datasets starting from the same type of source.

A mobile application, designed and developed from scratch, shown in Figure 2, is the instrument designed in order to acquire and send information from users to the central infrastructure.

Figure 2 - Mobile application for Species Distribution crowdsourcing

The eENVplus mobile application for species distribution allows users to collect, visualize and upload observations through the use of an interactive 2D map. Observations will be displayed as point of interests over publicly available thematic layers.

The data model for Species Distribution collection is represented in the following class diagram, Figure 3. Taking into account the issue related to the amount of information exchanged between mobile devices, the Species Distribution data model was designed in order to be as light as possible.

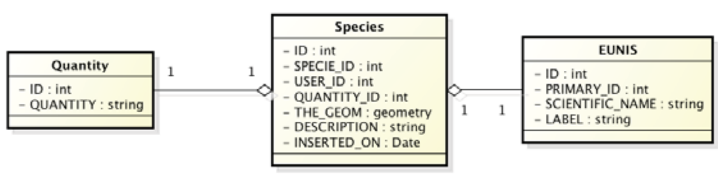

Figure 3 - Species class diagram for crowd-based observations collection

The aforementioned data model was designed considering as main requirement the lightness of the entire package representing each single observation. More in details:

- Species: represents the main table where observations about species are physically stored;
- EUNIS[3]: represents a dump of the EUNIS "Species" database, containing the scientific name of each element and an unique identification number, used also as primary key, for each of them. The EUNIS "Species" database is one of the three databases allowed by the Species Distribution technical guidelines[4];
- Quantity: support table which role is to allow users to specify a quantity of reported elements for each observation.

The observations collected within the use of the mobile application can be assessed also by the use of a HTML 5-based web portal. The aim of this portal, excluding the common point of interests visualization, is to allow the system administrator to approve or reject unpublished observations, enable or disable users and, most importantly, export observations by a comma separated value file (CSV) or within a fully INSPIRE-compliant Species Distribution GML.

---

3 http://eunis.eea.europa.eu

4 http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_SD_v3.0.pdf

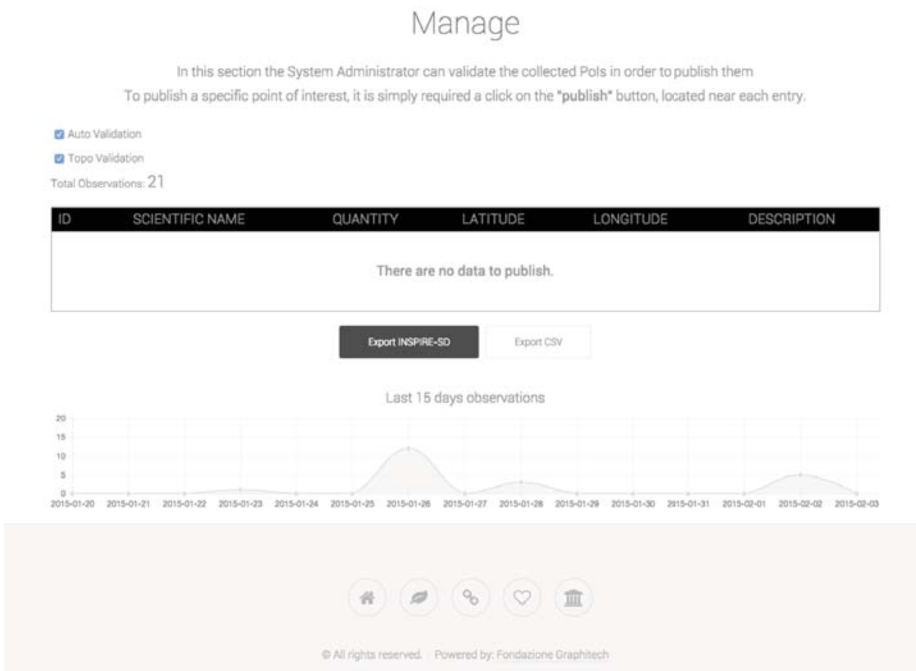

Figure 4 – eENVplus CrowdSourcing Service Administrator management portal

3.1. From CrowdSourced data to INSPIRE SD GML

Taking advantages of the Crowdsourcing service developed within the eENVplus project and described in the previous chapter, our effort has been focused into automatic procedure generation to create, starting from a lightweight data model optimized for mobile devices data exchange, into a complex GML INSPIRE compliant data theme with Species Distribution.

This procedure was designed in order to be applicable on others INSPIRE data themes.

The following figure represents the version 3.0 of the INSPIRE Species Distribution class diagram that has to be replicated from the developed procedure starting from the input data model.

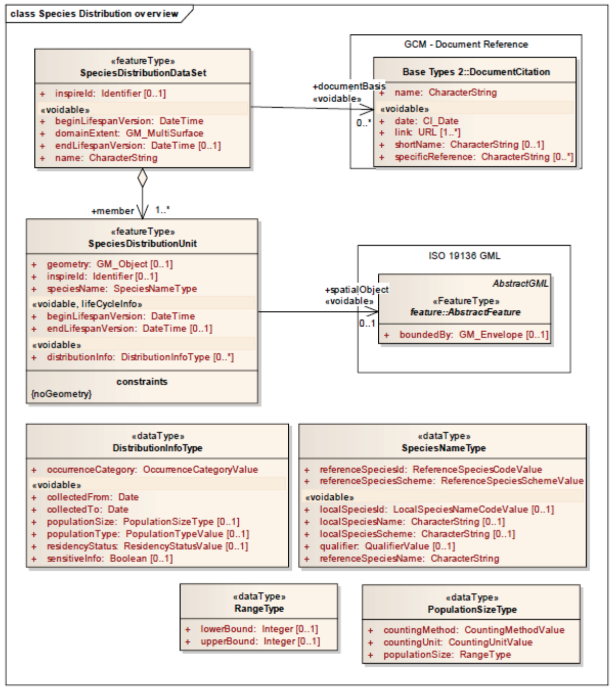

Figure 5 - INSPIRE Species Distribution class diagram v3[5]

Due to the complexity of the schema reported in Figure 5, and the need to minimize the quantity of data collected through a mobile device, our decision was to concentrate our attention mainly, but not only, on the mandatory field of the schema: in this manner we ensured that the compliance with the reference schema is maintained.

Starting from the eENVplus Crowdsourcing service data model for species distribution reported in the previous chapter, the following figure shows the mapping between source and target schemas.

---

[5] http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_SD_v3.0.pdf

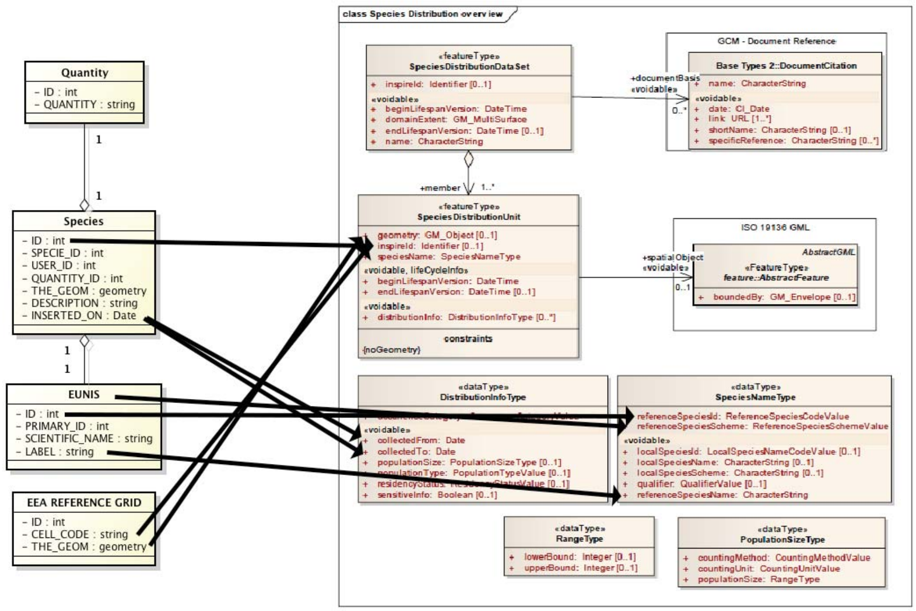

Figure 6 – Mapping between elements in order to generate a SD v3 INSPIRE compliant dataset

In order to generate a fully compliant INSPIRE SD v3 GML a support table is needed in addition to the ones involved in the process of storing data coming from the eENVplus CrowdSourcing Service.

The aforementioned table, represented in Figure 6 named as "EEA reference grid[6]", contains the **1km** grid for the area of interest needed in order to map species distribution according to the data theme technical guidelines.

The following table represents the mapping between source and destination elements in order to obtain a fully compliant INSPIRE SD v3 GML file.

Source	Destination
Species.ID	SpecieDistributionUnit.inspireId
EEA.CELL_CODE	SpecieDistributionUnit.inspireId
EEA.THE_GEOM	SpecieDistributionUnit.geometry
EUNIS.ID	SpeciesNameType.referenceSpecieId
http://inspire.ec.europa.eu/codelist/ReferenceSpeciesSchemeValue/eunis	SpeciesNameType.referenceSpeciesScheme

[6] http://www.eea.europa.eu/data-and-maps/data/eea-reference-grids-1

EUNIS.LABEL	SpeciesNameType.referenceSpeciesName
MIN(Species.INSERTER_ON)	DistributionInfoType.collectedFrom
MAX(Species.INSERTER_ON)	DistributionInfoType.collectedTo

Table 1. Mapping between source and destination data model

Figure 7 represents how the reference grid is composed for the observations collected in France.

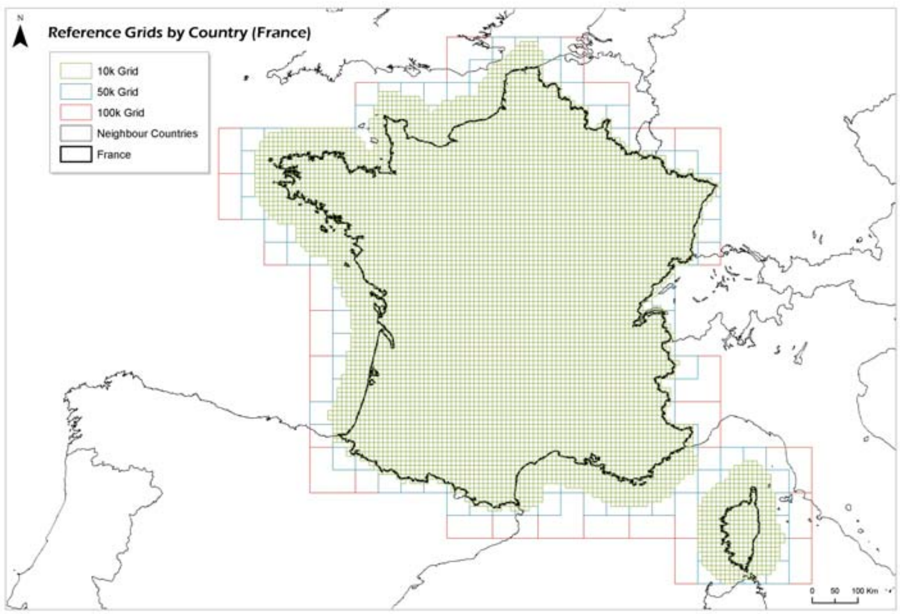

Figure 7 - EEA France[7] reference grid, EPSG:3035

The automatic procedure for data translation is developed by the use of the Java Servlet technology: following the mapping described in Figure 6, the produced output GML is a concatenation of the following INSPIRE Compliant Species Distribution elements, visible in Figure 8.

[7] http://www.eea.europa.eu/data-and-maps/data/eea-reference-grids-2#tab-gis-data

```xml
<base:member>
    <sd:SpeciesDistributionUnit gml:id="eenvplus_SD_export_0_53">
        <sd:inspireId>
            <base:Identifier>
                <base:localId>0</base:localId>
                <base:namespace>SD.EENVPLUS.Unit</base:namespace>
                <base:versionId>1.0</base:versionId>
            </base:Identifier>
        </sd:inspireId>
        <sd:geometry>
            <gml:Polygon gml:id="0_53" srsName="urn:ogc:def:crs:EPSG::3035">
                <gml:exterior>
                    <gml:LinearRing>
                        <gml:posList>2888000.000125 3760999.99984 2888000.000125 3760000 2888999.999965 [...]</gml:posList>
                    </gml:LinearRing>
                </gml:exterior>
            </gml:Polygon>
        </sd:geometry>
        <sd:speciesName>
            <sd:SpeciesNameType>
                <sd:referenceSpeciesId xlink:href="http://eunis.eea.europa.eu/species/53"/>
                <sd:referenceSpeciesScheme xlink:href="http://inspire.ec.europa.eu/codelist/ReferenceSpeciesSchemeValue/eunis"/>
                <sd:referenceSpeciesName>Callimorpha hera Linnaeus, 1767</sd:referenceSpeciesName>
            </sd:SpeciesNameType>
        </sd:speciesName>
        <sd:distributionInfo>
            <sd:DistributionInfoType>
                <sd:occurrenceCategory xlink:href="http://inspire.ec.europa.eu/codelist/OccurrenceCategoryValue/present"/>
                <sd:populationSize>
                    <sd:PopulationSizeType>
                        <sd:countingMethod xlink:href="http://inspire.ec.europa.eu/codelist/CountingMethodValue/counted"/>
                        <sd:countingUnit xlink:href="individuals"/>
                        <sd:populationSize>
                            <sd:RangeType>
                                <sd:upperBound>1</sd:upperBound>
                                <sd:lowerBound>1</sd:lowerBound>
                            </sd:RangeType>
                        </sd:populationSize>
                    </sd:PopulationSizeType>
                </sd:populationSize>
                    <sd:sensitiveInfo>false</sd:sensitiveInfo>
                    <sd:collectedFrom>2015-05-06</sd:collectedFrom>
                    <sd:collectedTo>2015-05-06</sd:collectedTo>
            </sd:DistributionInfoType>
        </sd:distributionInfo>
        <sd:beginLifespanVersion xsi:nil="true"/>
    </sd:SpeciesDistributionUnit>
</base:member>
```


Figure 8 - Portion of the generated GML file SD v3 INSPIRE Compliant

In order to have a visual validation of the produced GML file, the following figure represents:

- On the left, the source dataset, composed by a set of point of interest. Each POIs contains the information acquired by users related to a specific specie. The reference data model for each point of interest is the one reported in Figure 3;
- On the right side of the figure, the generated INSPIRE compliant Species Distribution GML file. The GML file is obtained applying automatically the mapping described in **Figure 6** from the source data model to the target data model.

Figure 9 - Comparison with the source dataset and final GML file

3.2. Publish INSPIRE SD through INSPIRE network services

Although in the previous section the obtained output was a fully INSPIRE-compliant GML datasets according with the Species Distribution technical guidelines, the main limitation at this stage is due to the fact that no standard services to exploit the produced dataset was part of the loop.

The eENVplus infrastructure, mentioned in the first chapter of this paper, has a specific module to ingest, process and distribute INSPIRE- compliant datasets through the use of a network service based on the Web Feature Service (WFS) protocol[9] (Vretanos, P. A. 2005).

The aforementioned component was not built from scratch, but constructed on top of an open source software for spatial data infrastructure (Groot, R., & McLaughlin, J. D. 2000) and geospatial web. This software, called deegree[10] (Fitzke, J., Greve, K., Müller, M., & Poth, A. 2004), includes a specific module to create, starting from any XML Schema Definition, XSD, file, specific mapping between relational databases that will contain the representation of the XSD file taken as input, and a WFS service.

In this way the generation of the INSPIRE-like database to allow the storage of data coming from INSPIRE-compliant datasets in accordance with Annex I, II and III, can be easily managed.

The adoption of degree introduces many advantages summarized below:

- The use of a standard protocol for geographical data exchange: the use of a standard protocol for data exchange, the Web Feature Service in this context, allows the possibility to exploit the result dataset in other Spatial data infrastructure and in other applications.

- The possibility to store data in a relational database: files will be moved from a GML-based format into a more flexible relational database.
- The reasoning enabled through the use of the supported ISO 19143[8] Filter Encoding (Vretanos, P. A. 2005) query language: WFS 2.0 allows users to perform filters over the result dataset. For example, it is possible to ask for a element that contains a specific attribute value, or it is possible to perform spatial and logical operations:
    - **Logical** operations: *and, or and not*;
    - **Comparison** operations: *equal to, not equal to, less than, less than or equal to, greater than, greater than or equal to, like, is null and between*;
    - **Spatial** operations: *equal, disjoint, touches, within, overlaps, crosses, intersects, contains, within a specified distance, beyond a specified distance and BBOX*;
    - **Temporal** operations: *after, before, begins, begun by, contains, during, ends, equals, meets, met by, overlaps and overlapped by*;
    - **Equal** operation: to test whether the identifier of an object *matches* the specified value.

In this regard, the publication action is completed performing a transactional WFS call from the eENVplus Crowdsourcing Service to the spatial data infrastructure able to manage these commands.

The following sequence diagram shows how the system reacts to a publication request performed by the moderator.

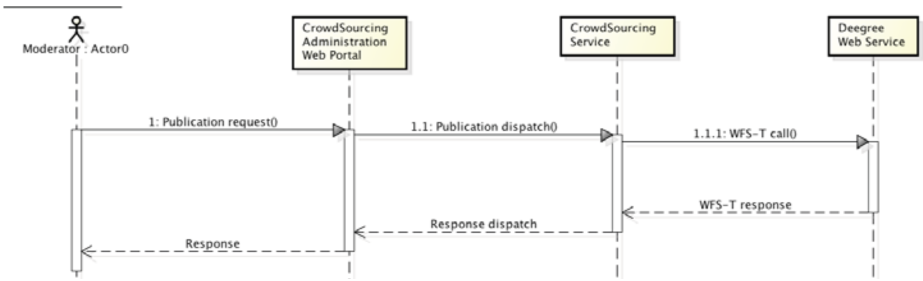

Figure 10 - GML publication Sequence Diagram

The entire overview of the workflow is visible in the following schema.

[8] https://portal.opengeospatial.org/files/?artifact_id=39968

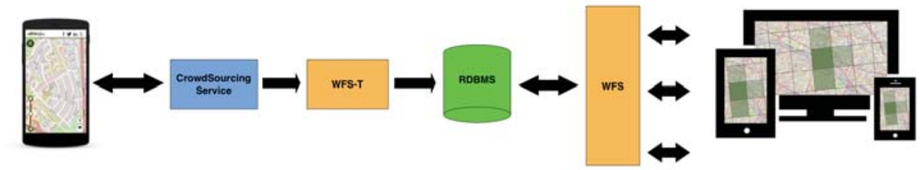

Figure 11 - Data generation, ingestion, harmonization and publication

4. Future Work

There are three main areas that will be influenced by the future work plan:

1. **Replacement** of the current data collection method, developed from scratch, taking the advantages of the WFS-T functionality provided by the deegree web service: The choice to develop the eENVplus CrowdSourcing service from scratch was due to the fact that the aforementioned service was designed to acquire different data models, sometimes containing pictures. The focus of this paper was related to a small part of the functionalities developed within the context of the eENVplus CrowdSourcing service. The use of the WFS-T for data ingestion also enables the use of the WFS protocol for data exchange between mobile apps.
2. **Increasing** the number of INSPIRE data theme produced by the system. Taking advantage of the WFS-T based version of the eENVplus CrowdSourcing Service, the objective in this context is to increase the number of INSPIRE data themes supported by the application. In this way, future version of the developed service can produce different INSPIRE-compliant datasets according with different data theme specifications.
3. **Updating** the procedure for INSPIRE-compliant Species Distribution GML creation and publication, producing a WPS (Schut, P., & Whiteside, A. 2007) based version of the tools. The process of "standardization" of the procedure aimed at data collection (the use of the WFS-T for data ingestion, and WFS for data dissemination) will influence also the automatic procedure for data harmonization. By the use of the WPS protocol, the developed processing services (one for INSPIRE data theme compatible with the system) will be also re-usable in other Spatial Data Infrastructures.

5. Conclusions

The aim of this paper is to offer a possible solution, different from the previous approach, in order to put non-professional users in the loop of generating spatial data compliant within the INSPIRE directive.

For the first version of the system we concentrate our attention on the use case of the species distribution in Europe, designing and developing a crowdsourcing mobile application to collect data from users according to a lightweight data model and, on the server side, a set of APIs to manage the incoming data, harmonizing, on the fly, information from the source to the destination (INSPIRE SD v3) data theme.

Through the use of a spatial data infrastructure, based on deegree web service, able to ingest, process and disseminate INSPIRE-compliant GMLs, a specific Crowdsourcing Service module was designed in order to automatically perform transactional WFS calls in order to store harmonized information in a more complex SDI.

Information stored and disseminated by deegree can now be accessed through external SDI, encouraging the diffusion of INSPIRE into operative activities of Member States.

Acknowledgments

The research leading to these results has received funding from the European Community's Competitiveness and Innovation framework Programme (CIP/2007-2013) under the Grant Agreement n. 325232. This publication reflects the authors view, and the European Commission is not responsible for any use which may be made of the information contained therein.

References

Attardo, C., & Saio, G. (2013). eENVplus: a framework to support eEnvironmental services and applications. In *EnviroInfo* (pp. 684-692).

Directive, I. N. S. P. I. R. E. (2007). Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). *Published in the official Journal on the 25th April*.

European Environment Agency (2014). Mid-term evaluation report on INSPIRE implementation, Joint EEA-JRC report, EEA Technical report 17/2014 http://inspire.ec.europa.eu

Fichtinger, A., Rix, J., Schäffler, U., Michi, I., Gone, M., & Reitz, T. (2011). Data Harmonisation Put into Practice by the HUMBOLDT Project. *International Journal of Spatial Data Infrastructures Research*, *6*, 234-260.

Fitzke, J., Greve, K., Müller, M., & Poth, A. (2004). Building SDIs with Free Software-the deegree project. In *Proceedings of GSDI-7, Bangalore, India*.

Flanagin, A.J.; Metzger, M.J. (2008). The credibility of volunteered geographic information. GeoJournal, 72, 137-148.

Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F., ... & Obersteiner, M. (2009). Geo-Wiki. Org: The use of crowdsourcing to improve global land cover. *Remote Sensing*, *1*(3), 345-354.

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, *69*(4), 211-221.

Groot, R., & McLaughlin, J. D. (Eds.). (2000). *Geospatial data infrastructure: concepts, cases, and good practice* (pp. 1-11). Oxford: Oxford university press.

Masser, I. (2007). Building European spatial data infrastructures. Esri Press.

Masser, I. (2015). An Autobiography? The first seven years of INSPIRE implementation. Published in Gis Professional 62, 20-22, 2015

Otakar Čerba, Karel Charvát, Karel Janečka, Karel Jedlička, Jan Ježek, Tomáš Mildorf (2012). THE OVERVIEW OF SPATIAL DATA HARMONISATION APPROACHES AND TOOLS. 4th International Conference on Cartography and GIS.

Schut, P., & Whiteside, A. (2007). OpenGIS web processing service. *OpenGIS Standard, Version 1.0. 0, OGC 05-007r7*.

Vancauwenberghe, G., Cipriano, P., & Craglia, M. (2014). Exploring the market potential for geo-ICT companies in relation to INSPIRE.

Vretanos, P. A. (2005). Web feature service implementation specification. *Open Geospatial Consortium Specification*, 04-094.

Vretanos, P. A. (2005). OpenGIS (R) Filter Encoding Implementation Specification.

Wieldmann and Bernard (2014). Linking crowdsourced observations with INSPIRE. Proceedings of the AGILE' 2014 International Conference on Geographic Information Science, Castellón, June, 3-6, 2014. ISBN: 978-90-816960-4-3