Reference Data and Services Infrastructure for the Danube Region

Alexander Kotsev, PhD,
Joint Research Centre, European Commission, Ispra, Italy
Institute for Environment and Sustainability,
HO6 - Digital Earth and Reference Data Unit - TP 263,
Via Enrico Fermi, 2749, I-21027 Ispra (VA), Italy
E-mail: alexander.kotsev@jrc.ec.europa.eu

Jean Dusart,
Joint Research Centre, European Commission, Ispra, Italy
Institute for Environment and Sustainability,

Robin S. Smith, PhD
Joint Research Centre, European Commission, Ispra, Italy
Institute for Environment and Sustainability

Abstract: *The overall objective of the manuscript is to provide a brief overview of the main achievements of the Danube Reference Data and Services Infrastructure (DRDSI) project with emphasis on the value it adds for stakeholders in the Danube region. The scope of activities is provided, together with a technical overview of the architecture of the open data platform, established to support decision makers in the Danube region.*

Key words: *Open data, EU Strategy for the Danube Region, web services, data infrastructures, INSPIRE*

JEL Classification: *O19, O52, Q01, P41*

I. Introduction

The Danube Region is home to more than 115 million people, a fifth of the EU's population. The countries are different in terms of economic strength but the Region is strongly interlinked, with potential for further integration and growth. Heterogeneity is also a characteristic of the data infrastructures in the region. Within this context, the Joint Research Centre of the European Commission (JRC) is developing the Danube Reference Data and Service Infrastructure (DRDSI) in support to the European Union Strategy for the Danube region (EUSDR) (Scientific Support to the Danube Strategy, 2015). The region, as defined within the Strategy, covers 14 countries, nine of which are EU Member States[1]. Within this context the overall idea of the DRDSI project is to 'unlock' and harmonise data which are relevant for the so-called 'macro-regional' development

of the Danube region. Priority is given to geospatial data within the scope of the EUSDR, including research projects of the JRC. Moreover, the establishment of the DRDSI will benefit from the lessons being learned, and the capacity and resources being developed, in the creation of a pan-European Spatial Data Infrastructure as a result of the implementation of the European INSPIRE Directive (2007/2/EC, 2007). The project is not a stand-alone act but, from the very beginning, attempts to align the scope of the activities with other similar initiatives. For example, a technical report (Smith, R., A. Kotsev, H. Gerlach; J. Dusart, 2015) has investigated existing project databases and similar resources, as well as how this information may be presented and reused by DRDSI.

The overall objective of this paper is to provide a brief overview of the main achievements of the project with an emphasis on the value being added for stakeholders in the Danube region. Section 2 focuses on the components of the DRDSI, and describes the (i) open source platform which acts as an entry point to well

---

1 Austria, Bosnia and Herzegovina, Bulgaria, Hungary, Germany (Baden-Württemberg and Bavaria), Slovakia, Slovenia, Ukraine (Zakarpatska, Ivano-Frankivska, Chernivetska and Odeska oblast), Moldova, Montenegro, Croatia, Czech Republic, Serbia and Romania.

documented data for the region, (ii) pilot activities which describe the value being added from the establishment of an open data infrastructure, and (iii) the network of experts, which is established within the project to help unlock existing data relevant to macro-regional development. The overview is followed by an outline of the activities which are foreseen for 2016.

II. Components of the Danube Reference Data and Services Infrastructure

1. Open Data Platform

Open data plays an increasingly important role as a means for public sector authorities to address the needs of both citizens and policy-makers in a better way. Within this context, an increasing number of open data platforms are put in place with the overall aim to provide an easy to use means of access to existing public sector data, helping to achieve some of the ob-

jectives of the Public Sector Information Directive (2013/37/EU). The first phase of the DRDSI involved setting-up an open data platform, providing access to existing resources (including data, metadata and web services) related to the Danube region (Figure 1). The platform currently provides access to more than 6,700 datasets (as of December 2015) which can act as a solid foundation for the integration of scientific knowledge into policy-making process on different levels (local, regional and international). From the perspective of macro-regional strategies, this would only be possible if data can be used across borders and domains, and put in the right context.

Within the platform the available datasets can be discovered together with metadata. Furthermore, the information can be visualized interactively online or, whenever available, be loaded into an external Geographic Information System (GIS).

Figure 1. DRDSI platform, available at: http://drdsi.jrc.ec.europa.eu

1.1. Platform architecture

As described above, the overall objective of the DRDSI platform is to provide access to distributed resources, made available within a service oriented architecture (SOA) (Thomas, E., 2008). This task is challenging for at least the following reasons: (i) lack of metadata, (ii) poor quality of metadata whenever available, (iii) maintenance may not be guaranteed and changes not updated), (iv) use of different standards, (v) lack of services to readily access data. Given those bounding conditions, in the attempt to maximize the use of available resources, the project has elaborated a workflow which is adapted to the needs of such a regional data infrastructure. During the design stage of the platform, open source technologies were chosen in order to take advantage of the fast release cycle, the minimum initial cost for deployment, the open nature of the selected products and the potential reuse of developments by others. Furthermore, we fully share the vision defined by Bouras, C., Filopoulos, A., Kokkinos, V., Michalopoulos, S., Papadopoulos, D., Tseliou, G. (2014), that open source can provide new opportunities, minimize research and development costs, and contribute to the development of ICT skills and e-inclusion in Europe if its full potential is reached.

The workflow for the generation of data for the DRDSI platform is metadata-driven. It is, in general, based on (i) harvesting heterogeneous endpoints, (ii) backend processing of metadata, and (iii) publishing the metadata. In terms of process, the workflow includes five consequent steps (Figure 1.) through which data is harvested, converted through an extract-transform-load (ETL) process and stored in a local database. Metadata are then made available through a frontend after a background quality check.
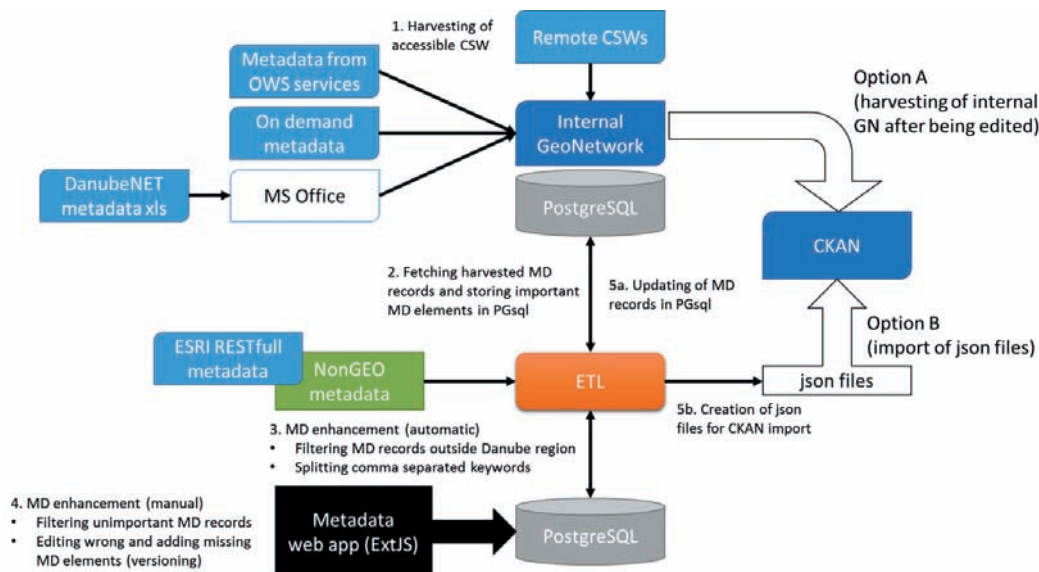

Figure 2. DRDSI platform architecture (Author: D. Divjak)

The core open source products which ensure the functionality of the platform are provided below (see Table 1).

Table 1.

Product	Description
GeoNetwork [8]	Geospatial metadata harvesting and management
PostgreSQL [9]	Data storage in a relational database
CKAN [10]	Open source data management platform
Metadata web app	Tool for metadata editing, based on the ExtJS JavaScript library

1.2. Content

A total of 6,712 datasets are available within the DRDSI platform as of 25 January 2016. The major data contributors to the platforms are provided in the table below:

Table 2.

Data source	Number of datasets	Relative share (%)
National INSPIRE Geoportals	3 712	55.3
Danube_Net data inventory[2]	1 436	21.4
Open Data Portals	865	12.9
Projects	565	8.4
Pan-European institutions	134	2.0

A series of pilot projects, based on data from the DRDSI catalogue are being implemented in 2016. They target both the creation of value-added applications, as well as filling existing gaps in the data infrastructure of the Danube region. Additional value from the implementation of these pilots include educational aspects and synergies with, for example, the implementation of INSPIRE in EU Member States and open data initiatives across the region. Further information regarding the scope of the pilot activities is provided by Smith R., A. Kotsev, J. Dusart (2015).

2. Danube_Net network of experts

The contribution of stakeholders of the Danube Region is essential for the overall success of the JRC support to the EUSDR. Within this context, in 2014, the JRC established a network of experts, entitled DANUBE_NET, covering the 14 EUSDR participating countries. Members of the network have been acting as 'ambassadors' of the DRDSI in support of data for macro-regional development, thus bridging the existing gap between data providers and the region's users. Lisjak, J., & Cetl, V. (2014). Provide additional information regarding the scope of activities of the network.

Specifically, DANUBE_NET members have been engaged in creating all the necessary means for 'unlocking' data, which is of relevance to the EUSDR, including the promotion of open data principles. To date, the network has created the following products in support of the establishment of a macro-regional open data infrastructure:

- State-of-play of the technical and policy context of data-sharing in Danube region countries (http://drdsi.jrc.ec.europa.eu/state-of-play)

- Inventory and documentation of new and potential data sources for publication on the DRDSI Open Data Platform

- Assessment of data access conditions and licensing constraints

- Outreach activities in the Danube region on different levels (local, regional and international) to increase the visibility of the DRDSI project, including the principles promoted by INSPIRE and Open Data initiatives.

III. Way ahead

The DRDSI's project pilots will (i) fill existing gaps through the establishment of missing infrastructure components, such as discovery and view services, (ii) harmonise datasets for cross-border and cross-domains, and (iii) demonstrate the benefit of scientific data for policy support through the creation of value-added products. The lessons learnt in these pilots can be further built upon by others in the region through their documentation in the platform. Furthermore, we consider the project as innovative as the majority of countries in the Danube region do not have a long tradition in data sharing, and open data has just recently been endorsed on a political level.

---

2   See section 2

Work is also underway to create a set of collaborative tools to aid data-sharing and the harmonisation of sources relevant to both JRC activities and data providers active in the region, alongside the further development of the DRDSI platform as a key entry point to the data resources uncovered. These online resources will allow the Danube region's stakeholders (policy makers, research organisations, non-governmental organisations, etc.) to access high quality data and information through common entry-points and collaborate to ensure new data investments are better aligned.

Reference / Referenzen

Bouras, C., Filopoulos, A., Kokkinos, V., Michalopoulos, S., Papadopoulos, D., Tseliou, G. (2014). Policy recommendations for public administrators on free and open source software usage, Telematics and Informatics, Volume 31, Issue 2, May 2014, pp 237-252. DOI: 10.1016/j.tele.2013.06.003

Lisjak, J., & Cetl, V. (2014). Building Danube Reference Data and Service Infrastructure for Danube Region Strategy. Zbornik radova dani iPP-a 2014 s međunarodnim sudjelovanjem, 37.

Smith, R. S., A. Kotsev, H. Gerlach; J. Dusart. Danube Region data projects. 2015. JRC Technical Report. DOI 10.2788/306239. http://publications.jrc.ec.europa.eu/repository/handle/JRC99268

Smith R, A. Kotsev, J. Dusart. (2015). Cross border Data-sharing Pilots .2015. JRC Technical Report. DOI 10.2788/337020. https://ec.europa.eu/jrc/en/publication/cross-border-data-sharing-pilots

Erl, Thomas. (2008). Soa: Principles of service design. Vol.1. Upper Saddle River: Prentice Hall.

Danube Reference Data and Services Infrastructure (DRDSI) project fiche http://ec.europa.eu/dgs/jrc/downloads/events/20130516-danube-strategy/jrc-20130516-danube-strategy-drdsi-flagship_cluster.pdf

Scientific Support to the Danube Strategy, Concept paper, http://ec.europa.eu/dgs/jrc/downloads/events/20130516-danube-strategy/jrc-20130516-danube-strategy-concept-paper.pdf

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) (OJ L 108, 25.4.2007, p. 1–14), http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32007L0002:EN:NOT

Directive on the re-use of public sector information (2003/98/EC), revised by Directive 2013/37/EU, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:EN:PDF

GeoNetwork opensource website. www.geonetwork-opensource.org, Accessed on 22.01.2016.

PostgreSQL website. www.postgresql.org. Accessed on 22.01.2016.

CKAN website. http://ckan.org/. Accessed on 22.01.2016.