

Paper number ITS-2567

OTN data hubs: stimulating innovation by improving access to open geographic information

**Susie Ruston^{1*}, Eric Legale^{2*}, Tomas Mildorf^{3*}, Indulis Makins^{4*}, Karel Charvat^{5*},
Dimitri Kozuch^{6*}**

1. 21c Consultancy, United Kingdom, +442072297316, susie@21cconsultancy.com

2. Issy Media, France, +33141238260, eric.legale@ville-issy.fr

3. University of West Bohemia, Czech Republic, +420737679838, mildorf@centrum.cz

4. Exigen Services, Latvia, +37129188251, indulis.makens@exigenservices.com

5. HSRS, Czech Republic, +42060461732, charvat@ccss.cz

6. HSRS, Czech Republic, +420317724620, dmitrii@seznam.cz

Abstract

The increasing availability of open Geographic Information (GI) presents a new opportunity for European Private and Public stakeholders, especially SMEs, to extract extra value from Open Data due to the fact that a vast amount of information has direct or indirect spatial references that open up new ways of interpreting it. OpenTransportNet (OTN), a European funded project, unlocks value from this data by creating a European network of data hubs that aggregate transport-related data and spatial information to drive the rapid creation of innovative and collaborative new ITS applications and services. During the first year of development, OTN worked with four Cities to create an INSPIRE (European Standard) based data model for transport networks that can be used with existing geo-spatial middleware technologies to harmonize and expose data in the form of easy-to-understand visual interpretations.

Keywords:

harmonisation, innovation, metadata

Challenges using geographic data

The use of Open Data (OD) for innovation is a relatively new phenomenon that has the potential to generate significant value for both the Public and Private Sector. For the Private Sector, Open Data offers an opportunity to gain competitive advantage by creating new services to generate income. For the Public Sector, it offers the means to create innovation and efficiencies that enhance service delivery to citizens. The increasing availability of open Geographic Information (GI) data presents even more opportunities for stakeholders to extract extra value from OD due to the fact that a vast amount of information has direct or indirect spatial references that open up exiting new ways of interpreting it.

Currently, there are two main EU initiatives dealing with GI data - INSPIRE¹ and Copernicus (former GMES²). The INSPIRE Directive defines standards and rules for access to public GI data. Copernicus focuses on Earth monitoring and data collection. These programmes represent important milestones in the effort to unlock experimentation with GI data. Nevertheless, the true innovative potential of GI data in Europe is hindered by the fact that access to GI data remains limited in many countries.

SMEs, for instance, are one of the strongest potential sources of GI Data innovation and yet often lack the time and resources to collect their own GI data let alone find this data amidst the massive variety of existing sources. This situation leads to interest in volunteered geographic information (VGI) such as OpenStreetMap (OSM)³, where volunteers across the world are collecting topographic information. OSM data is widely used and the coverage and quality is improving. Nevertheless, the OSM data model itself is complicated and not easily usable for non-GI experts. (Charvát, K. et al., 2013. SDI, INSPIRE and other initiatives. In INSPIRE and Social Empowerment for Environmental Sustainability, Results from the HABITATS project. Spain: TRAGSA.) At the same time, standards remain a second critical limitation for access to geospatial data. There exists large sets of geospatial standards introduced by the Open Geospatial Consortium (OGC)⁴, but again these standards are relatively complicated and not extensively used outside the geospatial community. The INSPIRE principles for networking, discovery and view of services are broadly adopted but other services such as web processing services, download services and defined data models are still in testing phase. Moreover, the INSPIRE data models need to be enriched further to make real progress in aligning different policies and requirements.

The above limitations mean many of the applications currently developed by non-GI experts are predominantly based on relatively easy Application Programming Interfaces (APIs) and accessible data sources from large private technology providers such as Google. This trend undermines the laudable efforts of the EU and Members States who have together invested billions of EUR to build Spatial Data Infrastructures that actually offer a larger diversity of data from different domains than Google. As a result, Europe is confronted with a potential monopoly of a few large commercial vendors at the expense of its most valuable source of home grown innovation – its SMEs⁵.

Three major sub-challenges must be addressed to unlock the full innovative potential of Open GI:

1. **Data:** Data sets (public, private, VGI) are currently **disharmonized**, scattered across different domains, difficult to access, and, in the case of VGI, not validated.
2. **Technology:** Spatial and non-spatial data needs to be combined/ linked in order to extract value and increase accuracy. Technology needs to ensure a) the easy integration of data streams from sensors and mobile objects including humans, b) the clear **visualization/presentation** of data c) the security and privacy of all users, particularly with regard to VGI
3. **Innovation:** Businesses and other stakeholders need easier APIs and GUI for accessing and analyzing data and new services (web processing, download services, defined data models) to more easily derive accurate insights and apply knowledge in a real world context.

OpenTransportNet (OTN) is a European funded project designed to extend beyond today's state of the art by specifically setting forth to stimulate the creation of innovative new applications and services in an area that touches upon all aspects of European society – transport. Although a variety of previous GI projects in Europe - such as Plan4business⁶, EnviroGRIDS⁷, Habitats⁸, Humbolt⁹, Briseide¹⁰, E.L.F¹¹ and Plan4al¹² - have furthered the GI state-of-the-art, they have generally tended to focus more on research and standardisation than

OTN data hubs: stimulating innovation by improving access to open geographic information

innovation. Additionally, they have likewise tended to focus on more narrowly defined target areas such as planning and crisis management rather than wider societal challenges such as transport that has been identified by Horizon 2020 as a key area of innovation need¹³.

OpenTransportNet Solution

Building upon the INSPIRE Directive as well as results and lessons learned from the previous GI initiatives, listed above, OpenTransportNet drives the rapid creation of innovative new applications and services by creating virtual service hubs that aggregate, **harmonise and visualise** Open Transport-related Data from a wide variety of sources.

The OpenTransportNet Service delivery hub uses an automated flexible dataset aggregator to integrate and harmonise transport related data. OpenTransportNet Hubs combine spatial (GI), dynamic data streams (sensors and monitoring of mobile objects) and non-spatial (OD) data using techniques such as RDF and data analysis based on ontology's, and derive insights from the data through visualisation tools and pattern detection algorithms. OTN Hubs will go beyond the State-of-the-Art by improving the accuracy of data insights by a) enhancing knowledge with Volunteered Geographic Information (VGI), and b) deploying a sophisticated Access Control and Identity Management system (ACM) which will manage privacy controls for personal and/or sensitive data. Finally, a 'data usage' approach to structuring Hub content and services will enable visitors to the platform to be directed to the innovation tools that best suit their needs. Users can play with data mash-ups, share insights with the OTN community, post service needs and create new business applications all for free.

The OpenTransportNet's service delivery hubs overcome the key challenges that currently hinder the use of GI data to solve real world transport-related problems and create new market opportunities:

1. **Data:** OpenTransportNet brings together data sets from a range of sources that can be used to improve transport and make them easier for innovators to access by:
 - Combining spatial (GI), dynamic data streams (sensors & monitoring of mobile objects) and non-spatial (OD) data using techniques such as RDF and data analysis based on ontology's
 - Using an automated flexible dataset aggregator to integrate and harmonise transport related data
2. **Technology:** OpenTransportNet combines spatial and non-spatial data to ensure the easy integration of data streams from sensors and mobile objects by:
 - Deploying visualisation tools and pattern detection algorithms to help users understand data
 - Deploying a sophisticated Access Control and Identity Management system (ACM) which will manage privacy controls for personal and/or sensitive data
3. **Innovation:** OpenTransportNet makes it easier to access and analyse data and derive more accurate insights for service development in a real world context by:
 - Using a data usage' approach to structuring Hub content and services to direct visitors to the innovation tools that best suit their needs
 - Enhancing knowledge with Volunteered Geographic Information (VGI)

Year one: overcoming data challenges

The OTN challenges are being tackled over a three-year period, by a range of partners, including city administrations, data experts, academics and transport SMEs. Currently in the first year of development, major focus has been on creating a strong, resilient, data model that can manage the complex range of formats and structures of the various transport sources. The rest of this paper outlines the OTN solution for a) harmonising metadata, b) using INSPIRE standards and c) exposing data for visual analysis.

Metadata: harmonising various metadata

One of the most important solution components in the OTN Hubs is its **metadata catalogue**. The catalogue helps users to search data and evaluate its fitness for intended use. OTN has chosen to use the MICKA metadata catalogue developed by Help Service Remote Sensing from the Czech Republic, due to its user-friendly interface making it a tool that non-GI experts can quickly adopt and use.

MICKA, as part of the OTN Hub, is responsible for storing meta-information about available data and also supports the discovery of existing geospatial data and services. A key objective of MICKA is to facilitate the editing of metadata profiles according to various standards including INSPIRE-compliant ISO 19115 (a standard for describing data), INSPIRE-compliant ISO 19919 (a standard for describing services), DC/ISO (a standard known as Dublin Core). As previously mentioned, OTN aims to bridge the gap between the spatial world of INSPIRE with Open Data initiatives and standards such as the Data Catalogue Vocabulary (DCAT), a metadata schema designed to facilitate interoperability between data catalogues published on the Web. DCAT is a metadata format proposed for European portals based on principles of the semantic web^{14 15}, and has been adopted by the OTN data model solution.

OTN has implemented the DCAT RDF-XML output in MICKA using an XSL template according to the rules defined in the INSPIRE profile of DCAT-AP - Extended version¹⁶. Keywords for describing data have been mapped to origin thesauri (e.g. GEMET) and the INSPIRE Registry¹⁷.

In addition to MICKA, the popular CKAN catalogue used predominantly for Open Data management will be used in OTN. CKAN currently supports harvesting from OGC catalogues (CSW 2.0.2) including ISO AP 1.0. However, the harvested metadata does not cover all the INSPIRE required metadata elements. The RDF output is available, but not fully compatible with DCAT-AP. Therefore, an extension module was developed in addition to the existing harvester to import additional INSPIRE metadata to DCAT. It is based on XSL templates so it is simply configurable without need to change the code. An extension for the metadata visualization was developed using the CKAN Jinja2 templating system plus Python background. This module takes metadata saved in the database and transforms them into a visually improved form.

Data harmonisation: integration of data into the INSPIRE based data model for transport networks

Data harmonization is necessary for combining data from heterogeneous sources (e.g. Regional datasets) integrated, consistent and unambiguous information products (e.g. European datasets). Such datasets can then be easily used in combination with other

OTN data hubs: stimulating innovation by improving access to open geographic information

harmonized data for viewing as well as querying and analysing. Data harmonization is a complex task without a universal solution that can cover all possible scenarios. The ideal technical solution (system architecture, software) is always determined by many specific facts including the way the original data sets are stored, the size of the data and the type of harmonization required.

The result of data harmonization process is data structured according to selected standards based on user needs. The needs for the OTN initiative were first identified and specified by the 14 project partners, before being further elaborated at OTN co-design workshops with 80 end-users from SMEs and public administrations. The gathered user requirements were analysed and ranked in order of priority before being translated into OTN Hub architecture design and technical specifications.

The INSPIRE Data Specification on Transport Networks was selected as the target data model for transport data within the OTN Hubs. A major source of transport data is found within a layer with the collaboratively built Open Street Map (OSM). The data model based on the INSPIRE specification was designed and data from the OSM was harmonized and imported into the OTN Hub. This new model now serves as a basis for majority of OTN's applications.

Exposure: analysis of diverse data and delivering visualisations for analysis

OTN's results will be validated by deploying data hubs in four European cities including Birmingham (UK), Liberec Region (Czech Republic), Antwerp (Belgium) and Issy les Moulinaux (France). The requirements and societal problems for OTN to solve differ between the pilot cities:

- **Birmingham – road safety;**
- Liberec Region – crisis management;
- Antwerp – infrastructure maintenance;
- Issy les Moulinaux – traffic congestion.

However, all the pilots will share a common platform with standardized interfaces and harmonized transport data. The high-level system architecture is depicted in Figure 1. The design is based on a trusted architecture used in Plan4business project and includes three basic layers: 1) storage, 2) geospatial middleware and 3) the front-end.

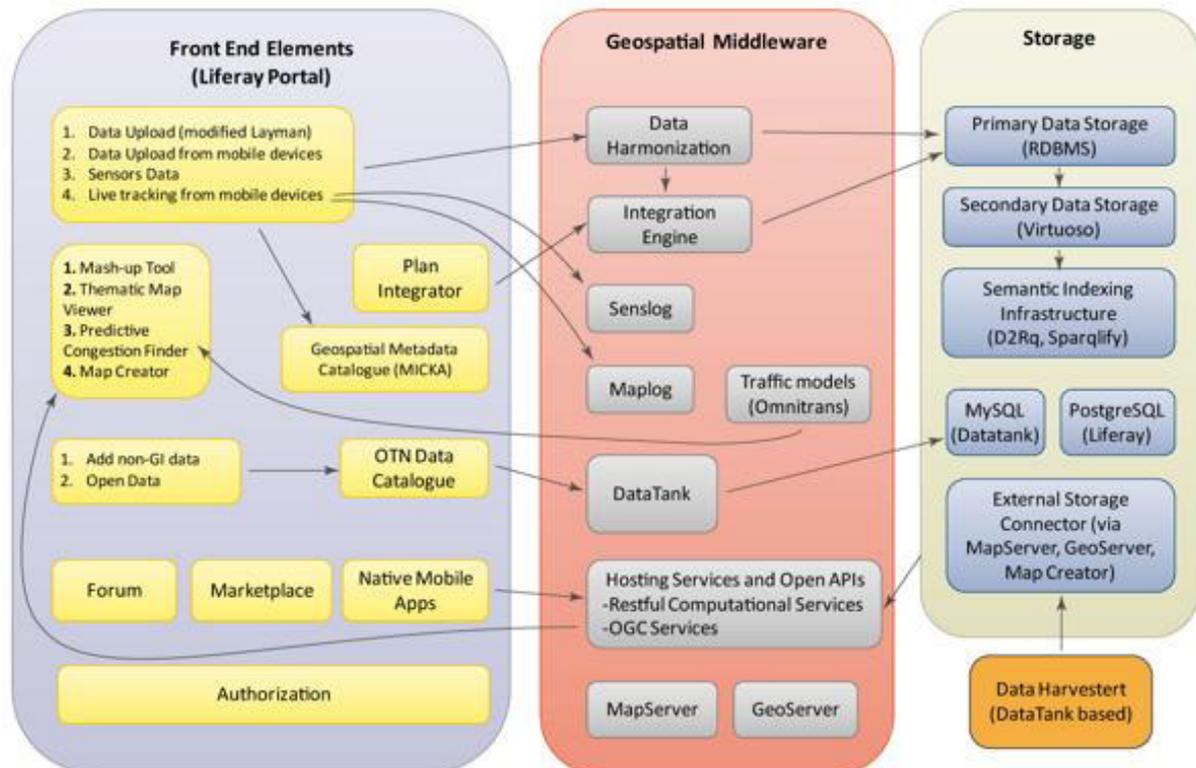


Figure 1. OTN high-level system architecture

The front-end layer takes advantage of previously developed applications in Plan4business and other European projects. These applications include, for example, the Thematic Map Viewer and Map Composer. The main objective of the Thematic Map Viewer is to visualize data stored in the OTN Hub in a user friendly way. Due to the fact that the database will contain many data-layers, a grouping of these layers will take place and predefined transport related thematic maps can be visualized. The Map Composer will enable creation of custom made thematic maps by the user.

When Geospatial data is made public, several steps are needed: a) Upload the data to the server, b) import the data into the database, c) publish the data through some kind of map server, and, if needed, d) configure the access rights so only the users with the proper privileges can display the data. LayMan - a Layer Manager – is a useful tool for sorting out the data into layers of information. Data can be published either from the uploaded files, or from the tables or views already present in the database. The various parameters of a layer can be set, with access control being of a special interest.

While manipulation with the published layer (write access) is limited to the members of the group that layer is published to, the read access (showing the layer in a map) can be granted to any other group. Users and groups are managed within the Liferay portal, which encapsulates the whole system. Once published, the layers can be styled with OpenGeo Styler.

Of course, the layers can be shown on the map interface. Currently, the Map Composer enables visualization of data from the following services: WMS, WFS, WCS, KML, GeorSS, GML, GeoJSON and SOS. Users can then combine selected layers into a thematic map (Figure 2) that can be saved as .hsl file (JSON styled text document that contains all necessary definitions and metadata of each layer and composition itself). It is further intended that the user will publish the thematic map and share it with other users.

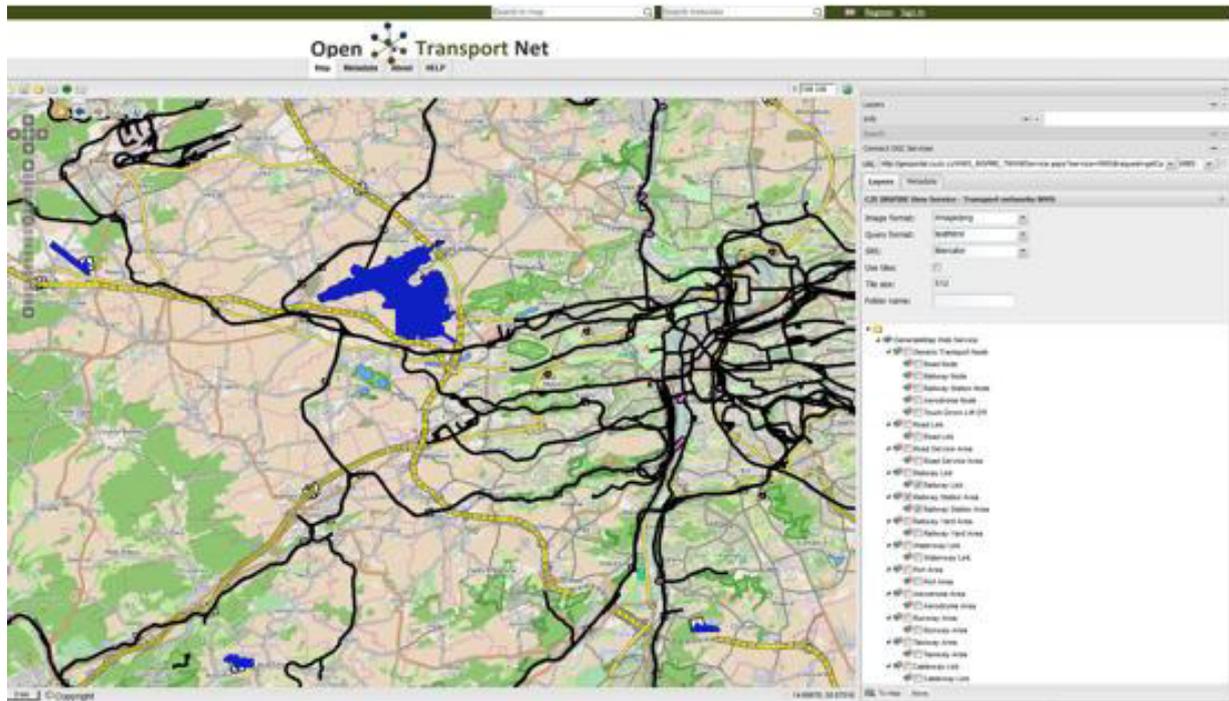


Figure 2. OTN thematic map

HSlayers NG is an online mapping library that operates in a web browser. It extends OpenLayers 3 functionality and takes some ideas from the original HSlayers library, but doesn't use Ext3 as the frontend javascript framework and is more lightweight in general. That's why the NG or "Next Generation" is added to its name. It is still under development and published under GNU/GPL licence version 3.

To use the library a developer has to link a couple of JavaScript files, setup the configuration parameters and include one html tag in his document. Further customization is possible and encouraged, but in this case some level of knowledge in JavaScript, CSS and possibly AngularJS is required from the developer.

Eurostat explorer is a demo application (module), which queries Semantic Web data sources via SPARQL endpoints. It demonstrates the feasibility of automatic query building for Eurostat report data and displaying it on a map of NUTS2 regions (specified in GeoJSON file) according to the calculated transparency ratios. On the server side it uses a Virtuoso Universal Server which is a middleware and database engine hybrid that combines the functionality of a traditional RDBMS, ORDBMS, virtual database, RDF, XML, free-text, web application server and file server functionality in a single system.

Used mainly for touristic purposes this component creates a special layer that contains the thumbnails of user-generated scenic landscape photos within an open database called "Panoramio". The photos are displayed online in the place where they were taken. It uses Panoramio's API, to get the most popular images in the current map extent. The number of images returned is dependent on the screen size of the users device.

It is worth noting that the OTN mobile application is a hybrid application written with web technologies HTML5, CSS and Javascript. It will run inside a native container and leverage the device's browser engine (but not the browser) to render the HTML and process the

JavaScript locally. Application will use a web view control (WebView on Android) to present the HTML and JavaScript files in a full-screen format, using the native browser rendering engine, therefore functionality will be accessible via WebView control (Figure 3).

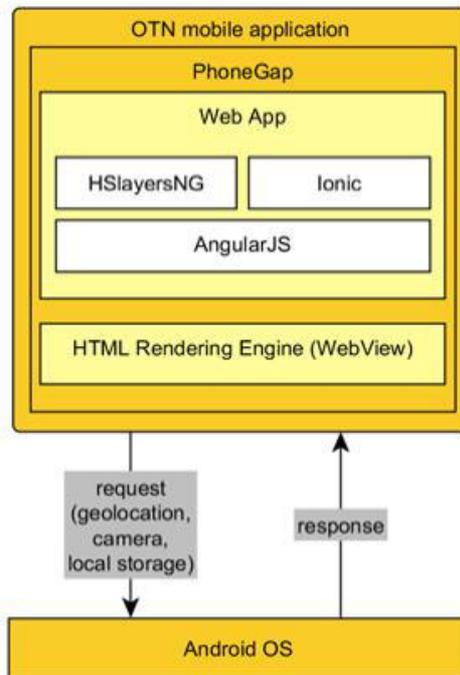


Figure 3. OTN mobile application

Conclusion

During the first year of its project lifecycle, OpenTransportNet set out to address the shortcomings found in current INSPIRE data model standards. Its aim was to create a solution that allowed transport related GI and Open Data to be viewed together in simple visual forms. In so doing, OTN's data model would remove the reliance that small businesses currently have upon the APIs of large information systems providers, such as Google and Navtech, and create an easy-to-use innovation environment where business intelligence could be extracted from the harmonization and layering of different transport data sets.

To date over 100 data sets have been selected for harmonization within the Hubs, and these have formed the basis for OTN's research and development work. This paper contains the high level technical results from the first year of the project, first identifying the challenges that need to be overcome to make the OTN vision possible, before reviewing the key elements of OTN's enriched data model blueprint. The resulting Hub specifications shifts INSPIRE architecture from a two-tier (User-Provider) to a three-tier (User-Broker-Producer) solution. Like an "application server," Brokers can do much more than just facilitate discovery and access of available resources: they can implement new capabilities for users and data/information providers such as advanced semantic discovery, clustering of results and quality control.

The validation of the enhanced data models will take place in 2015, when the newly developed data hubs will be rolled out to end-users in the four pilot cities. By 2016 all cities in Europe will have the potential to deploy their own instances of OTN's transport data hubs.

OTN data hubs: stimulating innovation by improving access to open geographic information

Acknowledgements

Acknowledgement of the whole OTN consortium and the European Commission for their valuable input and support.

References

1. Infrastructure for Spatial Information in the European Community (INSPIRE) website (2015). www.inspire.ec.europa.eu
2. Copernicus website (2015). <http://copernicus.eu/>
3. OpenStreetMap website (2015). <http://www.openstreetmap.org/>
4. Open Geospatial Consortium website (2015). <http://www.opengeospatial.org/>
5. European Commission (2000). *Innovation and SMEs – Keys to Prosperity*. https://ec.europa.eu/research/sme/leaflets/pdf/smes_en.pdf
6. Plan4Business website (2015). www.plan4business.eu
7. EnviroGRIDS website (2015). <http://www.envirogrids.cz/home>
8. Habitats website (2015). <http://www.habitats.cz/home>
9. HUMBOLDT website (2015). www.esdi-humboldt.eu
10. Bridging Services, Information and Data for Europe (BRISEIDE) website (2015). <http://www.briseide.eu/>
11. Jakobsson , A. (2012). *European Location Framework White Paper v1.0*. <http://www.eurogeographics.org/sites/default/files/E.L.F%20white%20paper%20v.1.0.pdf>
12. Plan4all website (2015). www.plan4all.eu
13. Horizon 2020 (2015). *Smart, Green and Integrated Transport*. <http://ec.europa.eu/programmes/horizon2020/en/h2020-section/smart-green-and-integrated-transport>
14. Data Catalogue Vocabulary (DCAT) (2014). *W3C Recommendation 16 January 2014*. <http://www.w3.org/TR/2014/REC-vocab-dcat-20140116/>
15. European Commission (2015). *DCAT application profile for data portals in Europe*. https://joinup.ec.europa.eu/asset/dcat_application_profile/home
16. European Commission (2014). *Inspire profile of DCAT-AP – Extended version*. https://ies-svn.jrc.ec.europa.eu/projects/metadata/wiki/INSPIRE_profile_of_DCAT-AP_-_Extended_version
17. European Commission (2015). *INSPIRE registry*. <http://inspire.ec.europa.eu/registry/>