Development of Multipurpose Land Administration Warehouse

Miodrag ROIĆ, Saša VRANIĆ, Tomaš KLIMENT, Baldo STANČIĆ and Hrvoje TOMIĆ (Croatia)

Key words: Multipurpose Land Administration, Cadastre, Land Management, LADM, Metadata

SUMMARY

A multipurpose land administration system (MLAS) should integrate various types of data that are stored in many official registers as public sector information. All information of the public sector, if available, is the potential which can contribute to social and economic development. The right of access to public sector information in recent years is guaranteed by regulations in many countries. Although the regulations differ, their objectives are common. Informing the population is an important task for the development of democracy, fighting corruption and increasing the accountability of the governing structures.

The availability and redundancy of data can be determined by comparing the key registers' data. Results of comparisons are used in the development and implementation of a prototype of an MLAS warehouse and implementation of a portal for access to (meta) data.

In order to establish a prototype warehouse, open source technologies were analysed in terms of their applicability to the development of an appropriate data model of a warehouse. Based on the strengths and weaknesses, we opted for the approach to modelling and establishment of a repository for various types of data.

Typical data are migrated to storage. Each migrated dataset is described in a standardized manner with metadata. In that way, discoverability and searchability is facilitated, which results in significantly improved ability to use the warehouse of an MLAS. Metadata/data infrastructure is established as a set of publicly available services suitable for research of improvements of usage of public sector information.

Two portals with the ability to view/download data have been implemented. These provide access to the metadata with the ability to search/view/download data. This allows further research on the possibilities of using the data for various purposes such as real estate mass valuation and land consolidation.

Development of Multipurpose Land Administration Warehouse

Miodrag ROIĆ, Saša VRANIĆ, Tomaš KLIMENT, Baldo STANČIĆ and Hrvoje TOMIĆ (Croatia)

1. INTRODUCTION

Efficient management of public sector information is the basis of effective governance. In a business and technological manner, systems of public sectors are largely based on the registers. Integration of different data stored in them should be enabled through multipurpose land administration systems (MLAS) in order to achieve many benefits for the whole society. Availability of public sector information through an MLAS enables a higher level of informed population, a more efficient fight against corruption and a higher responsibility level of the governing structures, all of which results in democracy development.

MLAS core registers are cadastre and land book which are essential state registers and are part of so-called key registers. Key registers are primarily those that contain information about the people and the land as well as about their relationships (Ellenkamp and Maessen, 2009). This definition can be extended to other registers which meet a set of requirements and whose data at some point becomes of key significance to society (Mađer et al., 2015).

The concept of MLAS implies avoiding redundancy among connected data as explained in article of this topic (Mađer et al. 2015), and also involves the establishment of procedures and standards for data sharing (Lemmen et al., 2015). To achieve this, adequate technologies need to be used. Open source technologies give the possibility to develop an inexpensive and efficient MLAS.

Modern thinking in multipurpose cadastral systems is increasingly directed at developing the MLAS; that is, building national land information infrastructures (Bennett et al., 2012). Aside from the cadastre as the fundamental register of land information, many countries have other registers of land and land tenure that need to be combined. One characteristic they have in common is that most information is registered in relation to the basic spatial unit – the cadastral parcel – but also with regard to specific spatial features (e.g. a national park). If such features are not originally linked to parcels, difficulties in management often arise (Roić et al., 2016).

Apart from for land information, which they register and maintain, public authorities are also responsible for many other types of information – they produce and maintain statistical, financial, economic, traffic, legal and numerous other types of official information. All such information represents an enormous potential that can be used to achieve economic progress if it is accessible and reused, which means utilized in new ways, adding new values, combining information from different sources and analyses, and finding new applications (EU, 2013).

This paper describes the development of an MLAS warehouse as a basis for research on land data linking and integration. The paper provides an overview of the different features registered in land administration systems and the possibilities for increasing their efficiency by combining registered land data from other registers.

2. REGISTERS AND INSTITUTIONS

All registers in a given country are governed by regulations, and their management and maintenance are entrusted to a public administration body. That body must maintain the register in keeping with regulations and guarantee the integrity of the information recorded in it. For every register, a purpose(s) for which it is established is defined on establishment. Key components of an MLAS are registers of land information and land tenure.

Cadastres are fundamental land data registers. In general, they can be divided according to registration type, on deed and title registration systems. Among cadastres, in many countries there are other registers which deal with land and interests on it. In most cases they are not linked and do not share the data mutually. Heterogeneous registers are not sufficient to address the requirements of users and it is necessary to unify them into a homogenous infrastructure and effective land information system available to users.

Land Administration Domain Model (LADM) is a conceptual model for establishment of the MLAS. It defines cadastre and land book as core registers. Besides them, in LADM are defined so-called external registers which are connected to core ones. Connection between core and external registers is also provided in LADM through external classes. Data of external registers foreseen for connecting to core data registers are:

- ortho photos, satellite imagery, and Lidar and elevation models
- topography (planimetry), topographic maps (or databases with topographic data)
- geology, geotechnical and soil information
- (dangerous) pipelines and cable registration
- address registration (including postal codes)
- building registration, both (3D) geometry and attributes (permits)
- natural person registration
- non-natural person (company, institution) registration
- polluted area registration
- ship and aeroplane (and car) registration
- cultural history, (religious) monuments registration

Weak or non-existing linkage of official registers, which maintain those data and the data redundancy as an inevitable outcome of such a state, are the causes of various unwanted consequences for the relevant public authorities, as well as for citizens and companies as the end-users of that data (Mađer et al., 2015). By conducted research, registers which contain information about the land and the people who have an interest in it in the Republic of Croatia were defined as a key registers and analysed (Table 1).

Table 1. Analysed registers (Mađer et al., 2015)

Register	Public authority
Register of natural persons	Ministry of public administration
Register of non-natural persons	Judicial authority

Register of personal identification numbers	Tax administration
Land registry	Judicial authority
Cadastre	State Geodetic Administration
Register of spatial units	State Geodetic Administration
Utility cadastre	State Geodetic Administration

The obstacles for registers' linking were identified through analysis. Some overlapping of institutional responsibilities have been noted that indeed create difficulties in registers' linking. Apart from institutional overlap, a significant level of data redundancy in the data has been noticed, which is a direct consequence of non-existing linkage between the official registers. Non-existing linkage is the primary motivator for copying of data from one register to another, thus producing redundancy. More on register linking problems and redundancy can be found in article dealing with this topic in detail (Mađer et al., 2015).

3. DATA TYPES

Data for an MLAS are provided from different sources and as different data types. Data that are provided can be in analogue or electronic format and given as raster, vector or even textual data (Figure 1). All the data should be attained in an electronic format to be used in an MLAS.

INSPIRE defines 34 different data themes. Data that belong to one theme usually are provided in different data types. In our case, data for INSPIRE theme cadastral parcels is stored in several data types. Spatial data (e.g. cadastral plan) is stored in a MLAS warehouse as vector data in a spatial database and as separate vector files for every cadastral municipality. Archived sheets of the same cadastral plans are stored as scanned and/or georeferenced raster files. Descriptive (mainly historical) data about cadastral parcels are stored as scanned raster files of cadastral documentation while part of the descriptive data provided from official cadastral database is stored as text files.

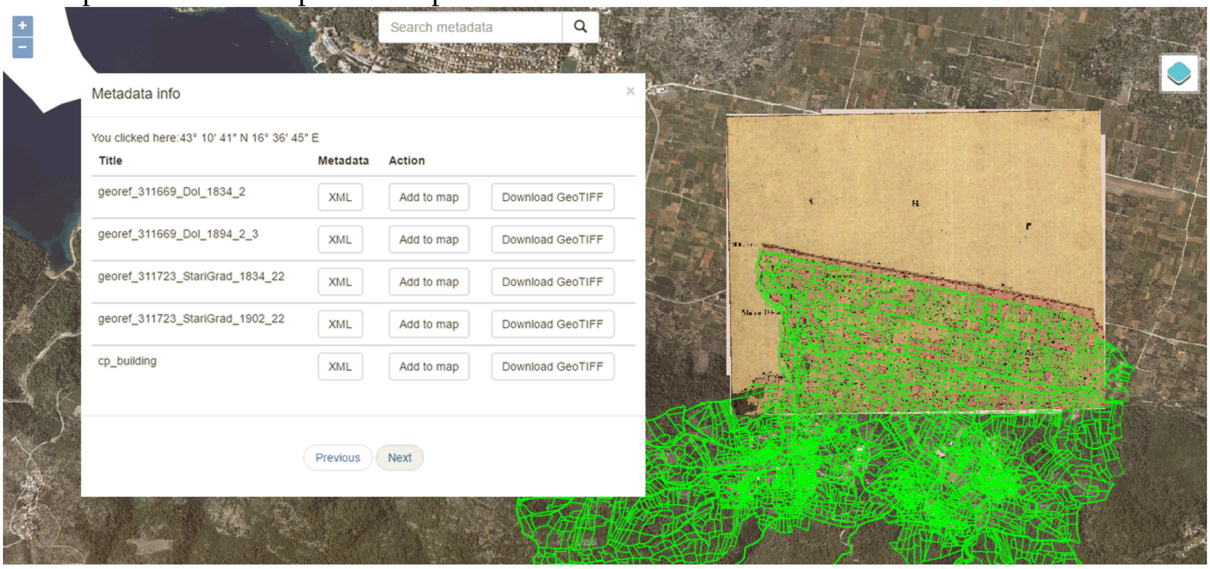

Figure 1. Data types in MLAS

Metadata of all those data should also be provided. Metadata are one of the fundamental and vital components of an MLAS. In general, they represent a set of attributes that describe the content, quality, availability and access to data, conditions and other characteristics of data. Therefore, they are used in an MLAS to enable simple finding of available (land) data.

3.1 Spatial

Existing spatial data is stored or provided through an MLAS. Another functionality that should be enabled in an MLAS is storing and providing raw surveyed data. Adequate procedures of uploading data along with their respective metadata is needed. These procedures are necessary in order to store the data in a way that they can later be easily found and retrieved.

Apart from currently valid data, an MLAS should be able to store and provide insight into historical data. It is necessary to combine them with data from different official registers through one place (MLAS portal). The following spatial datasets are loaded into an MLAS warehouse:

- official and historical digital cadastral plan,
- scanned georeferenced historical cadastral plan,
- ortho photos,
- elevation model.

3.2 Non-spatial

Searching and retrieving non-spatial data stored as raster files, text or in some other format also needs to be enabled in an MLAS. The easiest way to georeference these data is through adequate metadata elements (bounding box) so that they could be provided through an MLAS portal in the same way as spatial data. Scanned property sheets and other alphanumerical content of cadastral data are stored in an MLAS warehouse.

4. WAREHOUSE DATA AND METADATA MANAGEMENT SYSTEM

The data and metadata publishing workflow is based on the methodology for automatic creation, validation and publication of geospatial metadata proposed by group of authors (Kliment et al., 2013) and previously implemented in another project (Bordogna et al., 2016). The overall idea is schematically represented in Figure 2.

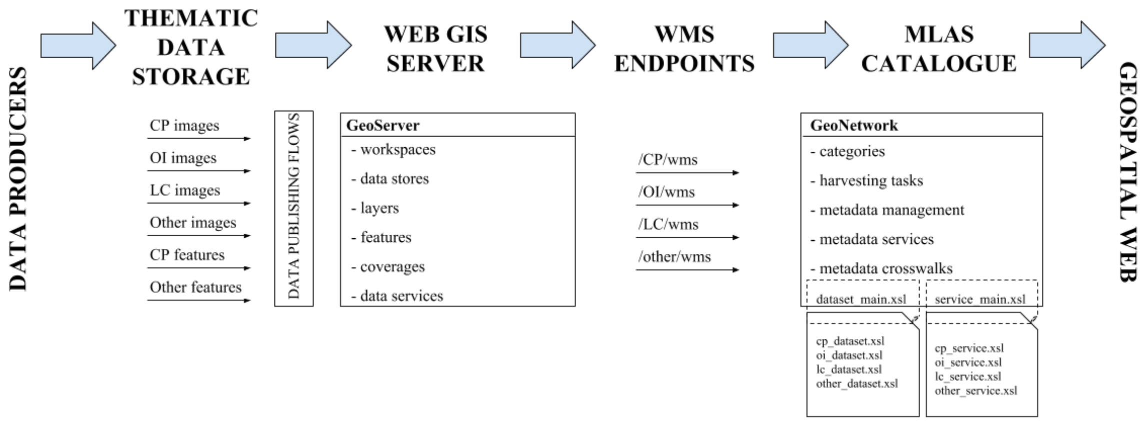

Figure 2. Schematic representation of the workflow designed to automate the process of data and metadata publishing within an MLAS warehouse

Starting from the left-hand side, the workflow inputs are represented by either historic or newly collected datasets provided by the registers (MLAS stakeholders) by moving them into a predefined folder structure entitled 'thematic data storage' represented by the file system in the virtual machine deployed in the university cloud centre.

Application infrastructure, which was used to implement a proposed solution, is open source and most of them implement OGC specifications

4.1 Logical model

Initially the land data were divided into the four thematic groups based on the spatial data themes defined by the INSPIRE directive (EU, 2007) as follows:

1. Cadastral parcels (CP)
2. Land cover (LC)
3. Orthoimagery (OI)
4. Elevation (EL)

These INSPIRE data themes were extended with another thematic data group named observations and measurements (OM). This group encompasses raw survey data stored in an MLAS warehouse. All observation and measurements data are described with metadata during the process of data entry in the warehouse.

In order to publish the datasets available within the thematic data folders into an MLAS warehouse, data publishing flows have to be implemented and incorporated into the workflow. A reported research (Bordogna et al., 2016) used Java-based open source tool GeoBatch; however, the researchers addressed some drawbacks, among which was a statement that the project is obsolete and complex to be customized by project specific needs. Since the needs of the MLAS required additional functionality as support of other data source types and automatic creation of projection files for raster files, it was decided to develop an in-house data publishing flow using PHP language and Apache web server. The current pilot version of the data publishing flow reads the TIF files in a predefined data folder (e.g. /opt/demlas/cp/georef), writes a project file (tiffilename.prj) and creates a coverage store on the MLAS warehouse deployed using GeoServer open source application. Once the data are published by GeoServer they are made available via open APIs as, for example, Web Map Service (WMS) to be portrayed, or Web Feature Service (WFS) and Web Coverage Service (WCS) to be downloaded and used (Cetl et al., 2016). In order to automate the next step of the workflow, which is to generate geospatial metadata, an MLAS catalogue is introduced. Basic (most time dynamic) metadata are collected in the catalogue by use of harvesting tasks applied on WMS service endpoint, which bears basic metadata about the service and dataset visualized as WMS Layers within its capabilities document. Other metadata such as lineage, abstract and keyword set are defined in metadata templates created for each data theme; for example, cadastral parcels, georeferenced maps, using an online excel sheet. These metadata are later integrated into individual records during the harvesting task execution. Related resources such as georeferenced cadastral and scanned map sheets are linked via a single metadata record providing the geospatial metadata. In addition, any other related data sources like additional info to cadastral parcels is described by individual metadata records and linked to the parent geospatial dataset metadata. The metadata are finally made available for querying by users using Catalogue Service for Web (CSW) interface. Two pilot solutions have been tested, one based on GeoNetwork as a complex back-end and front-end solution and another, pyCSW, as a standalone back-end catalogue.

5. ACCESS TO THE WAREHOUSE

The data from the MLAS warehouse can be searched and downloaded based on the metadata collected for the datasets in the back-end catalogues in order to make usage of data easier. Two graphic user interfaces have been developed, one searching for the data from the GeoNetwork (Figure 3) and the second from the pyCSW catalogue (Figure 4). They are basically almost identical from the end user point of view. The main difference is that the first solution communicates with the GeoNetwork catalogue using its internal codebase to search (API) and to display the results whereas the second solution is more generic, since it uses standardized communication to the catalogue; however, it requires additional coding to represent the results and internal logic of bidirectional relations among metadata as parent to child, service to dataset, etc. Both solutions passed the OSGeo incubation process; however, pyCSW leaves much lower impact on a web server and we wanted to explore its possibilities regarding to more robust GeoNetwork.

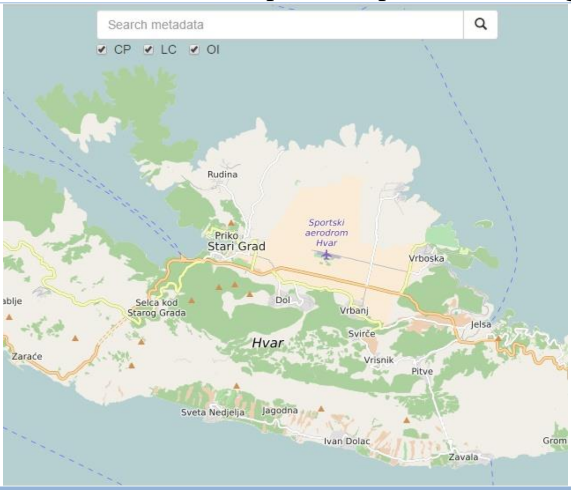

Figure 3. GeoNetwork interface

Figure 4. pyCSW interface

5.1 Searching

The user can use either full-text or spatial search defining the Point of Interest (POI) by clicking on the map (Figure 5). When the user clicks on a certain POI a query is made on the database and the geometry of the clicked point is compared with the geometry (bounding box) stored in the metadata. If the user is searching data using a full-text search then entered text is compared to all textual fields in metadata records (title, abstract, organization, etc.). A search term is made of words the user entered in the search box and added wild cards at the beginning and at the end of the search term.

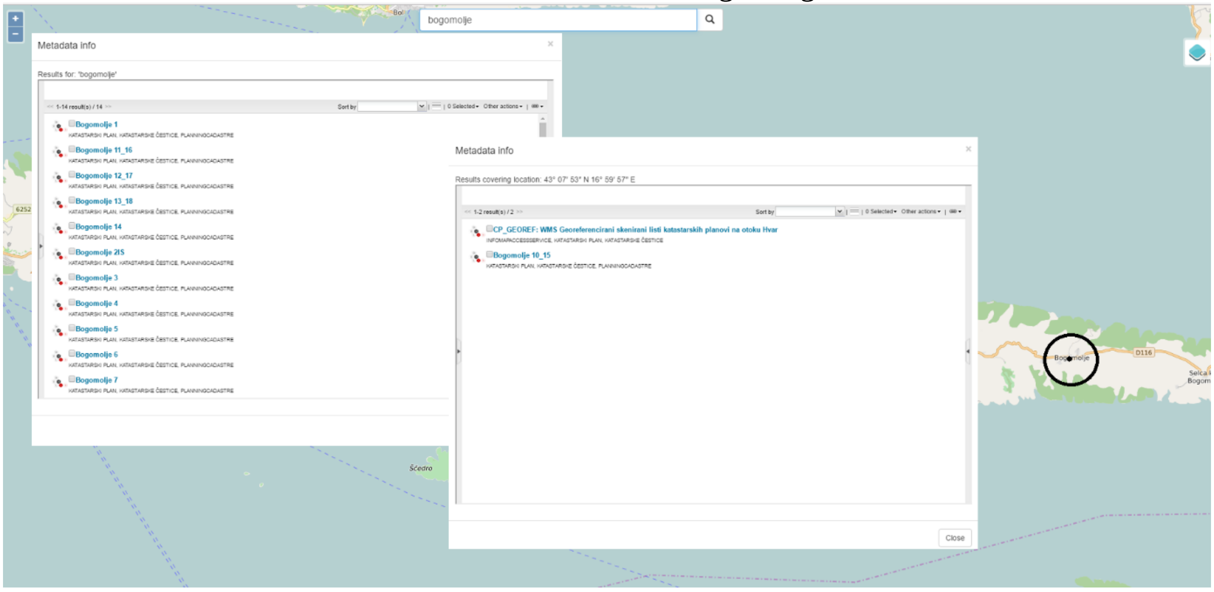

Figure 5. Searching in MLAS by free text or spatial query launched by click on a POI

5.2 Displaying, browsing and evaluating the results

Once the query has been sent to the back-end catalogue server the modal window appears and displays a list of results provided by the server (Figure 6).

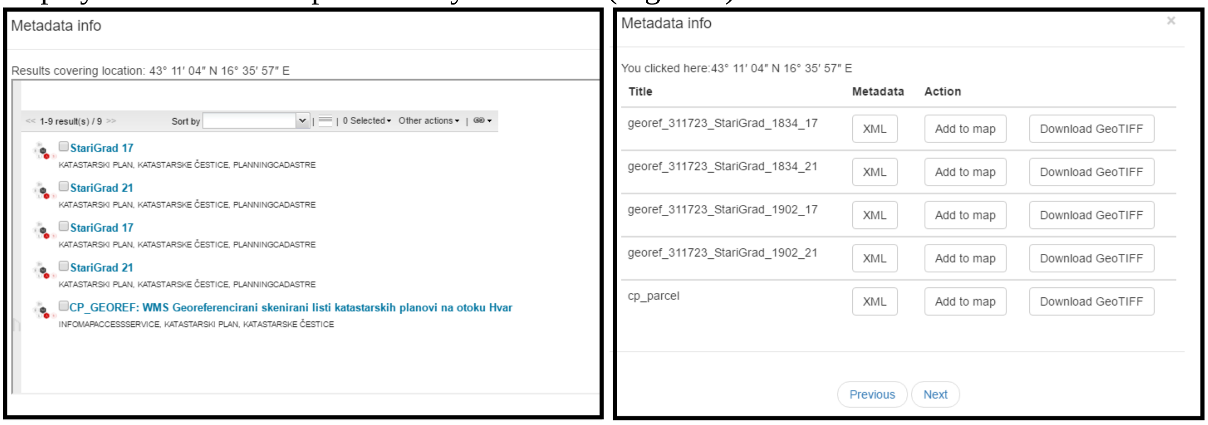

Figure 6. Result lists: left native GeoNetwork representation, right parsed from standard CSW GetRecords response XML

Using the solution based on GeoNetwork the user can click on a metadata record and display detailed information. In addition, if any related metadata are available, the user can browse through them; for example, Geospatial dataset (GeoTIFF) -> non-spatial dataset DPKC (PDF) -> another geospatial dataset from the series (GeoTIFF) -> geospatial service (OGC WMS), and vice versa (Figure 6).
Each search result can be added to the map, downloaded, or additional information can be viewed in the resulting XML file. Since pyCSW does not have interface which would enable user friendlier environment one can use existing solutions (CKAN, GeoNode), or develop special components which would enable easier viewing and interaction.

5.3 Downloading the data

In the last step, the user wants to obtain the actual data. Each metadata record about the georeferenced map contains a link to the original dataset in GeoTIFF format published via WCS. In addition, the respective scanned dataset can be downloaded via simple HTTP web link. Additional textual information about cadastral data can be downloaded as PDF files. Each georeferenced dataset can be used in any client application which supports standard geospatial data formats and OGC services.

By initially loading only a base set of layers we enabled faster loading of web application. Based on search results one can include only layers which are of interest.

6. CONCLUSION AND RECOMMENDATIONS

Official registers contain much public information. According to the regulations, access and use of data is often very restrictive. This prevents research in order to develop new applications by reuse of this data. By creating a prototype of the warehouse of official data this is enabled.

Access to the actual official data, without restrictions that exist for a particular register, opens the possibility of developing new applications of an MLAS in the technical sense. For any future application it will be necessary to analyse what the restrictions are as an obstacle to the implementation and propose appropriate changes in the legislation.

Another goal of this paper was to prove that data can be accessed in the same way with different applications through standards-based access. Data were organized into data themes and workflow was designed to publish those data through standardized API. Thus, open source applications were used. Most of these are OGC reference implementations of certain standard.

By comparing GeoNetwork and pyCSW we can conclude that GeoNetwork is more robust but leaves a much higher footprint on a web server, while pyCSW is lightweight and leaves a low footprint on a web server, which can be of great importance in high-level concurrency systems. On the other hand, pyCSW does not have a rich set of functionalities and components which can be used in creating a rich user interface; however, it can be integrated with other systems to overcome this deficiency. Both metadata/data warehouse approaches meet MLAS requirements and allow easier discoverability and usability of data.

ACKNOWLEDGEMENT

This work has been fully supported by Croatian Science Foundation under the project 7714.

REFERENCES

Bennett, R., Rajabifard, A., Williamson, I. and Wallace, J. (2012): On the need for national land administration infrastructures, Land Use Policy, 29(1):208-219.

Bordogna, G., Kliment, T., Frigerio, L., Brivio, P. A., Crema, A., Stroppiana, D., Sterlacchini, S. (2016). A Spatial Data Infrastructure Integrating Multisource Heterogeneous Geospatial Data and Time Series: A Study Case in Agriculture. ISPRS International Journal of Geo-Information, 5(5):73.

Cetl, V., Kliment, T., and Kliment, M. (2016): BORDERLESS GEOSPATIAL WEB (BOLEGWEB), Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLI-B4, 677-682, doi: 10.5194/isprs-archives-XLI-B4-677-2016, 2016.

Ellenkamp, Y., Maessen, B. (2009): Napoleon's registration principles in present times: The Dutch System of Key Registers. GSDI 11 World Conference, Rotterdam, Netherlands.

EU (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).

EU (2013): Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information.

Kliment, T., Granell, C., Cetl, V., & Kliment, M. (2013). Publishing OGC resources discovered on the mainstream web in an SDI catalogue. 16th AGILE International Conference on Geographic Information Science. Leuven, Belgium.

Lemmen, C.H.J., Van Oosterom, P. and Bennett, R. (2015): The Land Administration Domain Model. Land Use Policy, 49:535-545.

Mađer, M., Matijević, H., Roić, M. (2015): Analysis of possibilities for linking Land Registers and Other Official Registers in the Republic of Croatia based on LADM. Land use policy, 49:606-616.

Roić, M. Mastelić Ivić, S. Matijević, H. Cetl, V. Tomić, H. (2016): Towards a Standardized Concept of Multipurpose Land Administration. 78th FIG Working Week 2016: "Recovery from disaster". Christchurch, New Zealand.

URL 1: http://bolegweb.geof.unizg.hr/gnk_query/
URL 2: http://www.pg.geof.unizg.hr/metadata_viewer/index.html

BIOGRAPHICAL NOTES

Miodrag Roić graduated in Geodesy from the University of Zagreb, Faculty of Geodesy. In 1994, he received a PhD from the Technical University Vienna. Since 1996, he is a professor at the University of Zagreb, Faculty of Geodesy. He was Dean of the Faculty 2011-2015. The topics that he specializes in are Cadastre, Land Administration Systems, Engineering Geodesy and Geoinformatics. He was an editor-in-chief of Geodetski list. He is a corresponding member of the German Geodetic Commission (DGK) and many other national and international scientific and professional institutions.

Saša Vranić graduated at University of Zagreb, Faculty of Geodesy in 2009th with diploma thesis Interface of cadastral database. After graduation he worked in Croatian geoinformatic company Geofoto L.L.C. for several years as GIS Consultant on projects of implementing spatial information systems. Since 2012, he is employed at the University of Zagreb, Faculty of Geodesy as University Assistant on Chair for land surveying. He is also PhD student and his main research interests are land survey data management, land administration systems, spatial databases and web applications in service of cadastre and land survey.

Tomáš Kliment works currently as a Postdoctoral Researcher at Faculty of Geodesy, University of Zagreb. He finished his PhD at the Slovak University of Technology, Dept. of Theoretical Geodesy within the study program Geodesy and Cartography, with a specialization Geoinformatics. Dissertation thesis entitled: Discovery of geospatial resources available on the Internet dealt with the problem of geospatial information resources discovery via the Internet with emphasis on Spatial Data Infrastructures and the mainstream web represented by search engines. Main research interests are SDI, OGC Web services, geoinformatics and INSPIRE.

Baldo Stančić works as a Senior assistant (Postdoctoral) at the Chair of Spatial Information at the Faculty of Geodesy in Zagreb. He defended his Ph.D. thesis in 2013. at the same University, with thesis: "Modelling archival space-time cadastral data in the modern technological environment" Main research areas are Land Administration Systems. Baldo Stančić has participated on several projects and has published several papers.

Hrvoje Tomić works as an Assistant Professor at Department of Applied Geodesy, University of Zagreb, Croatia. In 2010 he received his Ph.D. from University of Zagreb for the thesis: "Geospatial Data Analysis in Purpose of Real Estate Valuation in Urban Areas". His main research interests are GIS and DBMS technology in spatial data handling. Hrvoje Tomić has participated on several projects and has published several papers.

CONTACTS

Miodrag Roić
University of Zagreb, Faculty of Geodesy
Kačićeva 26
HR-10000 Zagreb
CROATIA
Tel. + 385 1 4639 229
Fax + 385 1 4828 081
Email: mroic@geof.hr
Web site: http://www.geof.unizg.hr/~mroic

Saša Vranić
University of Zagreb, Faculty of Geodesy
Kačićeva 26
HR-10000 Zagreb

CROATIA
Tel. + 385 1 4639 210
Fax + 385 1 4828 081
Email: svranic@geof.hr
Web site: http://www.geof.unizg.hr/~svranic

Tomáš Kliment
University of Zagreb, Faculty of Geodesy
Kačićeva 26
HR-10000 Zagreb
CROATIA
Tel. + 385 1 4639 367
Fax + 385 1 4828 081
Email: tkliment@geof.hr
Web site: http://www.geof.unizg.hr/mod/page/view.php?id=33

Baldo Stančić
University of Zagreb, Faculty of Geodesy
Kačićeva 26
Zagreb
CROATIA
Tel. + 385 1 4639 366
Fax + 385 1 4828 081
Email: bstancic@geof.hr
Web site: http://www.geof.unizg.hr/osobna.php?ISVU_oznaka=BS029

Hrvoje Tomić
University of Zagreb, Faculty of Geodesy
Kačićeva 26
Zagreb
CROATIA
Tel. + 385 1 4639 522
Fax + 385 1 4828 081
Email: htomic@geof.hr
Web site: http://www.geof.unizg.hr/~htomic