May, 2018

Current achievements in Romania for integration of soil data into the infrastructure for spatial information of the European Community

Ruxandra Vintila
CURRENT ACHIEVEMENTS IN ROMANIA FOR INTEGRATION
OF SOIL DATA INTO THE INFRASTRUCTURE FOR SPATIAL
INFORMATION OF THE EUROPEAN COMMUNITY

Ruxandra Vintila

Dr. Ruxandra Vintila
National Research and Development Institute for Soil Science, Agrochemistry and Environment
ICPA Bucharest
61 Marasti Str, 011 464 Bucharest, Romania
ruxandra.vintila@icpa.ro

Abstract
This article presents the steps taken to ensure the consistency and usability of the geographic information system of soil resources of Romania SIGSTAR-200 within INSPIRE, the Infrastructure for Spatial Information of the European Community. At present, SIGSTAR-200 contains tens of thousands of polygons characterized by three attributes collected from legacy paper maps, namely (i) soil mapping unit (SMU), including soil association, (ii) topsoil texture class and (iii) skeleton class, all defined as stated by the national methodologies. In addition, each soil is characterized concerning the risk of land degradation (by water erosion, wind erosion, salinization, alkalinization, gleyzation and waterlogging) through attributes inferred by expert system rules built on pedogenesis.

To achieve the compatibility and usability within INSPIRE, SIGSTAR-200 has been transformed in accordance with the common Implementing Rules in force. To this end, first, the SMUs have been correlated at the dominant soil type level with the international soil classification system World Reference Base of Soil Resources (WRB) to ensure the semantic interoperability. Second, the SMUs, modeled as INSPIRE feature type SoilBody, have been populated with new attributes. This step was performed for SIGSTAR-200 dataset in three coordinate reference systems (CRSs): EPSG:3844 (CRS of Romania), EPSG:3035 and EPSG: 4258 (the last two CRSs being required or recommended by INSPIRE). Finally, the data transformed in INSPIRE-compliant GML have been checked by the INSPIRE Validator (Executable Test Framework), passing all the tests currently available for soil datasets, i.e., regarding (i) data consistency, (ii) INSPIRE GML application schemas, (iii) information accessibility, and (iv) reference systems. The next steps of the work, synchronized with the availability of the Executable Test Suites for the themes defined in Annex III of the Directive, aim to fully validate SIGSTAR-200 GML and to conclude its first integration into INSPIRE.

Keywords: digital soil mapping, INSPIRE Directive, interoperability, spatial data infrastructure

INTRODUCTION

Over the last years, it was widely recognized that soils provide key ecosystem services in response to global existential challenges in food security, water security, energy security, climate change, human health, and biodiversity. Besides the focus on ecosystem services, current perspectives also explore moral and ethical values that people may have in relation to soils (e.g., Grunwald et al., 2017). In this context, promising soil policies have emerged, built upon the recently coined concept of soil security, to better address existential challenges as well as societal values and achieve the Sustainable Development Goals set by the United Nations (e.g., Mc Bratney et al., 2017; Montanarella 2017).

This global context may pave the way for an appropriate legal framework dedicated to the protection of soils, which would greatly benefit the ecosystems of Romania (Vintila et al., 2016). Its implementation requires, inter alia, data easily available through the National Spatial Information Infrastructure, which is part of INSPIRE, the Infrastructure for Spatial Information in the European Community (http://geoportal.ancpi.ro/geoportal/catalog/main/home.page; https://inspire.ec.europa.eu/; Pashova and Bandrova, 2017).

This article presents the steps taken to ensure the consistency and usability of the geographic information system of soil resources of Romania SIGSTAR-200 within INSPIRE. SIGSTAR-200 was chosen as a priority dataset because at present it represents the richest source of soil information at the country level.
METHODS

Model Input: Presentation of SIGSTAR-200 soil dataset

The SIGSTAR-200 dataset contains tens of thousands of polygons, originating from the legacy map “1: 200,000 Soil Map of Romania”, which consists of 50 sheets (Florea et al., 1963-1994). We emphasize the difficulty of the data collection, half of the soil delineations being smaller than the minimum legible area on paper defined at Cornell University in the US (e.g., Rossiter 2000), thus having some insuperable effects on the overall digital data quality.

The first version of SIGSTAR-200 (Vintila et al., 2004) contains polygons characterized by three attributes collected from the paper sheets, namely (i) the soil mapping unit (SMU), described according to the Romanian System of Soil Classification SRCS 80 (Conea et al., 1980), along with (ii) the topsoil texture class and (iii) the skeleton class, both defined in accordance with the National Soil Methodology (Canarache et al., 1987). Besides, each soil is characterized concerning the risk of degradation (by water erosion, wind erosion, salinization, alkalization, gleyzation and waterlogging) through attributes inferred by Dr. Ion Manteanu using expert rules built on pedogenesis (Vintila et al., 2004). We also mention that an SMU can be composed of one soil or an association of up to five soils that could not be separated at the mapping scale of the source data (Note: what is called ‘soil association’ by Romanian and other soil surveyors is sometimes called ‘soil complex’, e.g., within USDA). There are 477 different soils and soil associations.

A recently released version of SIGSTAR-200 (2018) adds the correlation - at the levels of dominant soil type and dominant soil class - with the Romanian System of Soil Taxonomy SRTS 2012+ in force (Vlad et al., 2014; Vlad et al., 2015), as well as the correlation with the World Reference Base of Soil Resources (WRB) (Vlad et al., 2012a), endorsed by the International Union of Soil Sciences and recommended by INSPIRE to ensure semantic interoperability (EC 2013). Furthermore, the SMUs are defined in a formalized way, in order to minimize the narrative ambiguity of the soil association descriptions and to facilitate automatic processing by parsing (Vlad et al., 2012b).

We provide in Table 1 the approximate proportions of soils within soil associations, based on the surface area covered (Vlad et al., 2014). They should be henceforth considered in processing (e.g., to assess the spatial extent of the risk of land degradation) and should replace the previous simplified processing based on the dominant soil of the SMU.

Table 1. Approximate proportions of soils within soil mapping units (Vlad et al., 2014)

<table>
<thead>
<tr>
<th>Soil place in the enumeration describing an SMU</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>One soil</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Two soils</td>
<td>60%</td>
<td>40%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Three soils</td>
<td>50%</td>
<td>30%</td>
<td>20%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Four soils</td>
<td>40%</td>
<td>30%</td>
<td>20%</td>
<td>10%</td>
<td>0%</td>
</tr>
<tr>
<td>Five soils</td>
<td>30%</td>
<td>25%</td>
<td>20%</td>
<td>15%</td>
<td>10%</td>
</tr>
</tbody>
</table>

Model Output: Generation of INSPIRE-compliant SIGSTAR-200 GML

In this paragraph, we have intensely used the INSPIRE ‘Implementing Rules’ and ‘Data Specification on Soils. Technical Guidelines’, v 3.0 (EC 2013) that we cite only here for the entire section. According to these documents, the application schema for the theme Soil comprises the following legal feature types: DerivedProfilePresenceInSoilBody, DerivedSoilProfile, FAOHorizonNotationType, ObservedSoilProfile, OtherHorizonNotationType, OtherSoilNameType, ParticleSizeFractionType, ProfileElement, RangeType, SoilBody, SoilDerivedObject, SoilHorizon, SoilLayer, SoilPlot, SoilProfile, SoilSite, SoilThemeCoverage, SoilThemeDescriptiveCoverage, SoilThemeDescriptiveParameterType, SoilThemeParameterType, WRBQualifierGroupType, and WRBSoilNameType.

To model the soil mapping units (SMUs) of SIGSTAR-200, we have directed our attention toward the feature type SoilBody. Figure 1 illustrates the part of the Soil UML Class Diagram focused on SoilBody. The neighboring feature types are voidable, which means that they only are used when the corresponding data are available. For example, DerivedSoilProfile is used when there are data on soil profiles associated to SMUs (Note: the INSPIRE feature type DerivedSoilProfile is comparable to the better known concept of Soil Typological Unit). Thus, given that the SMUs of SIGSTAR-200 are not generally associated with soil profiles, the application schema has allowed us to rely only on the feature type SoilBody in order to transform our dataset into INSPIRE-compliant GML. This feature type is described in UML by two mandatory attributes, inspireId and geometry. However, one of the voidable attributes, soilBodyLabel, is very important in our case, being directly linked to the legend items of SIGSTAR-200.
Figure 1. Part of the Soil UML Class Diagram (EC 2013) focused on the feature type SoilBody

We have used the XML application schema Soil 4.0. Moreover, to achieve the transformation into INSPIRE-compliant GML, we have evaluated several commercial and open source software products. In the end, we have provisionally used INSPIRE Solution Pack for FME and the INSPIRE GML template for Soil (the latter made freely available by S. Dupke of con terra company) that we have adapted to our dataset. Subsequently, we have performed the GML dataset transformation into three coordinate reference systems (CRSs): EPSG: 3844 (CRS of Romania; Note: EPSG: 31700 is deprecated), EPSG: 3035 (INSPIRE for spatial analysis and reporting) and EPSG: 4258 (INSPIRE for View services).

To conclude, the INSPIRE-compliant SIGSTAR-200 GML EPSG: 3035 and EPSG: 4258 have been checked using the tests currently available for the themes included in Annex III of the Directive (http://inspire-sandbox.jrc.ec.europa.eu/validator/). These tests refer to the following aspects (Figure 2): (i) data consistency, (ii) INSPIRE GML application schemas, (iii) information accessibility, and (iv) reference systems.

Figure 2. Example of INSPIRE Validator Test Run for Data Consistency, General requirements
The overall aim of the IT modeling has been to obtain an usable output consistent with all feature types of the Soil application schema (listed at the beginning of this paragraph) as well as consistent with the feature types of the other INSPIRE application schemas (Figure 3), with due attention to the ones related to Soil (such as application schemas for the themes Elevation, Hydrography, Geology, Land Cover, Land Use, Protected sites, Natural risk zones, Biogeographical regions, and Environmental monitoring facilities).

Figure 3. SIGSTAR-200 as part of INSPIRE
RESULTS

INSPIRE-compliant SIGSTAR-200 GML

SIGSTAR-200 GML that we have generated passed all the tests available in December 2017 for the themes listed in Annex III. As an example, Figure 4 shows the results of the test suite shown in Figure 2.

![Figure 4. Example of INSPIRE Validator results for Data consistency, General requirements](image)

Giving the complexity of the INSPIRE GML encoding, this validation process is not trivial. In fact, SoilBody feature type has maximum five attributes in the UML Class Diagram (Figure 1), while its INSPIRE GML encoding relies on numerous properties. Meanwhile, we clarify that an SMU is uniquely identified by the property inspireId.Identifier.localId. Figure 5 briefly illustrates some properties for a few SMUs.

To get a better insight into the INSPIRE-compliant GML description, Figure 6 presents the properties of an arbitrary SMU (highlighted in green on the map), where we notice the complexity of description of the mandatory attributes. We also notice that the only soil information is the SMU definition, stored in the GML property soilBodyLabel.

![Figure 5. INSPIRE-compliant GML SIGSTAR-200 (1/2)](image)

INSPIRE+ SIGSTAR-200

For applications at the national level, we have joined the INSPIRE-compliant SIGSTAR-200 GML with the table containing the other attributes (mentioned in the Model Input paragraph) and generated an INSPIRE+ SIGSTAR-200 (SIGSTAR-200 v2.0). This is illustrated through an example in Figure 7, where we remark five attributes marked within borders. They ensure, in the absence of associated soil profiles, the semantic interoperability, however only at the levels of dominant soil type and dominant soil class, through correlation of the Romanian classifications with WRB.

SIGSTAR-200 v2.0 has a provisional View Service ([http://194.116.136.107/arcgis/services/icpa/sigstar/MapServer?WMSServer](http://194.116.136.107/arcgis/services/icpa/sigstar/MapServer?WMSServer)) characterized by a visualization style where the colors illustrate soil classes, while the SMUs are areas delineated inside the classes to which they belong (Figure 7). An improved style is in progress to deal with the 477 types of SMU (see a dedicated project in the Researchgate Profile 'Virgil Vlad'). We consider this effort a promising solution for the View Service of INSPIRE-compliant SIGSTAR-200 GML.
Figure 6. INSPIRE-compliant GML SIGSTAR-200 (2/2)
Figure 7. INSPIRE+ SIGSTAR-200 (SMU is abbreviated UCS in Romanian, i.e., Unitate Cartografica Sol)
CONCLUSIONS

A significant number of eminent scientists, among whom Arrouays *et al.* (2015), Grunwald *et al.* (2017), McBratney *et al.* (2017) and Montanarella (2017), have highlighted the irreplaceable role of soils in the securitization of socio-environmental processes. Urgent soil protection measures are needed, which in turn request easily available interoperable data. Admittedly, spatial information infrastructures constitute the top tech solution to address this demand in the most appropriate way (Masser *et al.*, 2008; Vintila *et al.*, 2015, Pashova and Bandrova, 2017).

Aligned to the INSPIRE implementation roadmap, SIGSTAR-200 GML passed the available tests for the theme Soil. The next steps, synchronized with the availability of the tests for the themes defined in Annex III, aim to fully validate SIGSTAR-200 GML and to conclude its first integration into INSPIRE.

Meanwhile, we remark that while *datasets with quasi-stable properties in time*, such as SIGSTAR-200, serve as a foundation for the theme Soil, other *time-dependent spatial datasets* should be added to meet the demands with respect to soil security and, more generally, to contribute to disaster risk management. These types of data are mostly acquired by remote sensing (forming time series) and are defined in INSPIRE as dataset series. Their prioritization firstly depends on application - landslide forecasting (e.g., Brocca *et al.*, 2012), soil moisture dynamics estimation (e.g., Prévot *et al.*, 2003; Kim and van Zyl, 2009), yield formation and forecasting (e.g., Baret *et al.*, 2003; Launay and Guérif, 2005) etc. - and, secondly, depends on the feature types foreseen so far in the application schemas (all feature types are provided at [http://inspire.ec.europa.eu/data-model/approved/r4618/fc/](http://inspire.ec.europa.eu/data-model/approved/r4618/fc/)).

In addition, other envisioned steps regard the integration of *point soil datasets*. In fact, only starting from point data, completely interoperable soil digital maps with associated uncertainties can be produced to support multi-level environmental governance (Vintila *et al.*, 2016; Vintila and Carabulea, 2018). Therefore, there is an urgent need to fully leverage the potential of legacy and new soil point datasets by using recognized spatial statistics procedures in order to properly address national demands, as well as to contribute to the Global Soil Map ([http://www.globalsoilmap.net/](http://www.globalsoilmap.net/)) and SOTER ([https://esdac.jrc.ec.europa.eu/](https://esdac.jrc.ec.europa.eu/)), a Soil component of the Danube Reference Data and Services Infrastructure.

ACKNOWLEDGEMENTS

This study was partially funded by the grant PN 16 07 04 03 of the Programs “NUCLEU” managed by the Romanian Ministry for Research and Innovation. We also acknowledge the kind support of Safe Software Inc for the data transformation tools and of the Romanian Space Agency for the View services provided on its ICT infrastructure.

REFERENCES


**BIOGRAPHY**

I am with the National Research and Development Institute for Soil Science, Agrochemistry and Environment (ICPA Bucharest) as a senior research scientist. I graduated from the "Politehnica" University of Bucharest, Faculty of Automatic Control and Computers, and "Paul Sabatier" University of Toulouse, Remote Sensing and Image Processing, MSc.

My current research interests include the use of geomatics, radiative transfer modeling and agro-physiological modeling in agriculture and environment. I am also interested in studying socio-ecological resilience.