Integrating 14 Danube basin country SOTER databases into Danube_SOTER

(Part of Deliverable D.B2)

Niels H. Batjes

ISRIC – World Soil Information (https://www.isric.org)

Contents

Background

To support the Danube strategy the European Commission Joint Research Centre (JRC) has proposed soil data collection within the Danube basin based on SOTER methodologies. The application thereof ensures standard procedures for data harmonisation are applied for the fourteen contributing countries. The production of a harmonised SOTER database at 1:250,000 scale will support soil and land use policy development in the Danube basin region.

The 'SOTER Danube' project (2019-2021) will develop a SOTER database (Danube_SOTER) for the 14 countries represented in the Danube basin. Activities include: a) providing 'empty' SOTER databases

and GIS files of terrain units mapped for the participating countries (based on shapefiles provided by EU-JRC (2017)), preparing guidelines for compiling the SOTER attribute databases (Batjes and Ribeiro 2019a, b; Ruiperez Gonzalez and Batjes 2019), b) national partners filling pre-populated 'national' SOTER databases with their soil data in an activity coordinated by Cranfield University, c) ISRIC integrating the 14 'country' sets into a single database (Danube_SOTER) with accompanying GIS file of Terrain Units (this report) and d) ultimately, exporting the data in SOTER format to XML/GML INSPIRE compliant format (by ISRIC).

Danube basin country SOTER databases

In October 2021, the last of 14 'clean' country SOTER files were submitted by Cranfield University for further processing at ISRIC. According to information provided by the Principal Investigator at Cranfield University (JH, 6/06/2021 pers. comm.), the respective 'country' databases were generated as described below, using AT_SOTER_Clean (Austria) as an example. This approach was necessary as few on the cont(r)acted countries were able to supply new soil data for their respective countries for various reasons.

The steps for filling AT_SOTER_Clean were:

1) "We used the EU soil map (dominant soil) to derive the soil components and proportions for the database.
2) We used the SPADE-14 DB to assign soil profile and horizon data to the soil components
3) We constructed various look up tables to convert to SOTER format in 1 and 2 above and developed scripts to generate the appropriate content."

Further, it was observed that "Please note for both DBs (i.e. AT and DE, the first datasets submitted) not all soil components have a corresponding soil profile and horizon data as this was not available for all soil components. Also, there is no geo-reference for the profile as these were selected as representative profiles for the soil type. Many analytical data have not been measured or recorded in the soil databases used, so there are gaps in the horizon tables. We cannot populate the tables when there is simply no data available to do so!"

As a result, integrity issues arose when the 14[1] different MS Access data sets were merged into a basin-covering Danube_SOTER database. Some critical fields were not filled (e.g. Laboratory methods), which nullified relationships between related attribute tables thus preventing straightforward importing of data tables into Danube_SOTER. Pragmatically, such 'empty' tables were populated with fictional values to re-establish referential integrity, to permit consistent merging of the various country datasets.

The table below shows and example for the LaboratoryMethod table:

[1] Abbreviations: Austria (AU), Bosnia and Herzegovina (BA), Bulgaria (BG), Czech Republic (CZ), Germany (DE), Hungary (HU), Croatia (HR), Moldova (MD), Montenegro (ME), Romania (RO), Serbia (RS), Slovakia (SK), Slovenia (SI), and Ukraine (UA).

LaboratoryMethod

Lab-ID	Attribute	Description
EU001	PSCL	Not defined
EU001	PHAQ	Not Defined

Further, it was noted that some datasets did not meet the requirements for a full SOTER database, as described in the supporting documentation (Batjes and Ribeiro 2019a). For example, the DE-set did not respect the criterion for defining soil horizon depths (upper and lower); inherently, it is up to the national partners themselves to resolve such issues. Further, for the DE dataset, two columns were erroneously added to table SoilComponent by the data provider (i.e., RDEP1 and Flächenanteil); these were removed to ensure consistency. Finally, the DE dataset (submitted as 'DE_SOTER_updated_V2_Clean.mdb') was renamed to DE_SOTER_Clean.mdb to allow for consistent import procedures.

Generating a single SOTER database for the Danube Basin

Attribute database

SQL queries were generated for the various stages of data import to the ultimate Danube_SOTER set. The procedure is described below using AT_SOTER_Clean as an example:

1) Zap all tables in target Danube_SOTER database (except for Codes table: first time only!).

2) Run 10 queries (Qi) in the specified sequence:

   Q1 – Append all data (*) to table LaboratoryMethod (needs to be populated, if duplicates error when running query/ macro just ignore this):

   "INSERT INTO Danube_SOTER IN 'C:\Danube\Danube_SOTER.mdb'

   SELECT AT_soter.* FROM AT_soter;" (Note: Adapt country ISO code for each dataset)

   Q2 – Append * to table Laboratory (needs to be populated, if duplicates error when running query/macro just ignore this):

   "INSERT INTO Laboratory IN 'C:\Danube\Danube_SOTER.mdb'

   SELECT Laboratory.* FROM Laboratory;"

   Q3 – Append * to table ProfileDatabase:

   "INSERT INTO ProfileDatabase IN 'C:\Danube\Danube_SOTER.mdb'

   SELECT ProfileDatabase.* FROM ProfileDatabase;"

Q4 – Append * to table SourceMap (needs to be populated, if duplicates error when running query/macro just ignore this):

"INSERT INTO SourceMap IN 'C:\Danube\Danube_SOTER.mdb'

SELECT SourceMap.* FROM SourceMap;"

Q5 – Append * to table Terrain:

"INSERT INTO Terrain IN 'C:\Danube\Danube_SOTER.mdb'

SELECT Terrain.* FROM Terrain;"

Q6 – Append * to table TerrainComponent:

"INSERT INTO TerrainComponent IN 'C:\Danube\Danube_SOTER.mdb'

SELECT TerrainComponent.* FROM TerrainComponent;"

Q7 – Append * to table SoilComponent:

"INSERT INTO SoilComponent IN 'C:\Danube\Danube_SOTER.mdb'

SELECT SoilComponent.* FROM SoilComponent;"

Q8 – Append * to table Profile:

"INSERT INTO Profile IN 'C:\Danube\Danube_SOTER.mdb'

SELECT Profile.* FROM Profile;"

Q9 – Append * to table Soils:

"INSERT INTO Soils IN 'C:\Danube\Danube_SOTER.mdb'

SELECT Soils.* FROM Soils;"

Q10 – Append * to table Horizon:

"INSERT INTO Horizon IN 'C:\Danube\Danube_SOTER.mdb'

SELECT Horizon.* FROM Horizon;"

All submitted SOTER datasets, being termed 'clean' by the data compiler, were processed 'as is' using the above workflow. All imports to Danube_SOTER processed smoothly. In total, data for 9356 Terrain Units and corresponding tables were imported.

GIS layers

GIS files for the 14 Danube basin countries, as provided with the pre-populated GIS files (Ruiperez Gonzalez and Batjes 2019), were merged into a corresponding single Terrain Units GIS shapefile (part of Deliverable D.B2).

Next steps

A next step will be to develop an application to generate XML/GML files according to INSPIRE specifications (Deliverable D.A6 and D.A7). AT-SOTER_Clean has been selected/used to develop the XML/GML tool, as discussed with the project Principal Investigator at Cranfield University.

Methodological details will be provided in a developer and user manual (Van Genuchten 2021 (*in prep.*)) with accompanying XML/GML files for the 14 countries (part of D.B1) as well as the integrated data set for the whole Danube basin (part of D.B2).

Acknowledgements

The SOTER Danube project is carried out under EU tender RC/IPR/2017/D.3/0002/OC, sub-contracted through Cranfield University (EU-JRC Service Contract #936108). The project is coordinated by Cranfield University.

References

Batjes NH and Ribeiro E 2019a. *Guidelines for compiling a 1:250,000 SOTER database (Deliverabe A5, EU JRC Service Contract #936108, Collection of soil data in SOTER format from 14 Danube strategy countries, at scale 1:250 000). [Project report 3]*, ISRIC - World Soil Information, Wageningen. https://www.isric.org/sites/default/files/Report03_SOTER_Danube_TechGuidelines_20210209.pdf

Batjes NH and Ribeiro E 2019b. *Updated MS-Access SOTER template for the EU Danube basin project with worked examples (Deliverabes A1, A2 and A4 EU JRC Service Contract Nr. 936108, Collection of soil data in SOTER format from 14 Danube strategy countries, at scale 1:250 000.) [Project report 1]*, ISRIC - World Soil Information, Wageningen, 28 p. https://www.isric.org/sites/default/files/Report01_SOTER_Danube_Database_2019-09-12.pdf

EU-JRC 2017. Collection of soil data in SOTER format from 14 Danube strategy countries, at scale 1:250 000. https://themes.jrc.ec.europa.eu/news/view/161358/collection-of-soil-data-in-soter-format-from-14-danube-strategy-countries-at-scale-1250-000

Ruiperez Gonzalez M and Batjes NH 2019. *Country terrain unit shapefiles of the EU Danube basin (Deliverabe A3 EU JRC Service Contract No. 936108, Collection of soil data in SOTER format from 14 Danube strategy countries, at scale 1:250 000) [Project report 2]*, ISRIC - World Soil Information, Wageningen, 23 p. https://www.isric.org/sites/default/files/Report02_SOTER_Danube_GIS_201906.pdf

Van Genuchten P 2021 (*in prep.*). *Exporting Danube Basin SOTER databases to INSPIRE compliant XML/GML format: Developer and user manual. (Deliverabe A6 & A7 EU JRC Service Contract No. 936108, Collection of soil data in SOTER format from 14 Danube strategy countries, at scale 1:250 000) [Project report 5]*, ISRIC - World Soil Information, Wageningen, xx p.