

Collection of soil data in SOTER format from 14 Danube strategy countries, at scale 1:250 000

EU-JRC Service Contract #936108

Deliverable D.A6 & D.A7 and part of D.B1 & D.B2

Exporting Danube Basin SOTER databases to INSPIRE compliant
XML/GML Developer and User manual

Paul van Genuchten

October 2021



Contents

- Background.....1
- Alignment using Hale Studio1
- User Manual2
 - Setting up the SOTER database2
 - Run the Alignment2
 - Verify the outputs in the INSPIRE validator.....3
- Developer Manual4
 - Alignment for Derived Profiles based on Horizons.....4
- Alignment findings.....6
- Acknowledgements7
- References.....7
- Annex 1: Preparing the SOTER database8
- Annex 2: Alignment based on alternative classes11

List of figures

- Figure 1.UML diagram Soil from D2.8.III.3 Data Specification on Soil – Technical Guidelines (Templer and others, 2013).....11

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 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 and d) ultimately, exporting the data in SOTER format to XML/GML INSPIRE compliant format (this report).

Alignment using Hale Studio

A common tool for transforming spatial data to INSPIRE (2013) compliant GML (INSPIRE 2013) is Hale Studio (Templer and Others 2021). Hale Studio uses the term 'alignment' for the process of transforming a dataset from one model to another.

The open source tool has been originally developed by the Humboldt Institute and is now maintained by a group around the company WeTransform. This section describes how the tool has been used to set up an alignment between SOTER and INSPIRE GML.

The deliverable of the work is a Hale project file. This project file can be called from the Hale command line interface to run an alignment on arbitrary SOTER databases merged with a shapefile for soil body geometry. A batch script is provided to facilitate triggering of the alignment from command line.

User Manual

Setting up the SOTER database

During the project we have identified that the alignment can run efficiently by adding some aspects to the database prior to the alignment. The aspects to add to the database are described in Annex 1.

Run the Alignment

This procedure describes the situation in MS Windows. Scripts operate in a similar way in Linux or MacOS Operating systems.

Unzip the provided package (Soter2Inspire.zip) to a folder close to the database(s) to be converted. The zip file contains the Hale-Cli executable used to run the alignment, a project file which defines the alignment and some batch scripts to trigger the alignments.

In order to run an alignment, you need a [Java Runtime Environment](#) (v8+) and you have to indicate to the scripts where Java is installed. Verify you have Java Runtime Installed, by opening console and typing:

```
Java --version
```

Verify that a parameter exists with the path to the Java folder. In normal situations JAVA_HOME is set. The application will throw an error if it is not.

```
Echo %JAVA_HOME%
```

You can set a temporary parameter with:

```
SET JAVA_HOME="C:\Program Files\Oracle\Java\jre.x.x"
```

Or set an [environment variable using Control Panel](#) to persist the value.

Now navigate the console to the folder where you unzipped the package.cd c:\temo\SOTER2INSPIRE

```
CD c:\temp\SOTER2INSPIRE
```

Then run the alignment by typing:

```
SOTER2INSPIRE {path-to-shp} {path-to-mdb} {namespace}
```

The script takes 3 parameters:

- The path to the shapefile, it can be a relative path. Quote the path if it has spaces.
- Path to the access database OBSERVATIONS (i.e., materialised view prepared as described in Annex 1).
- The namespace parameter is an organisational prefix which is used in INSPIRE ID.

For example:

```
SOTER2INSPIRE "data/CZ/CZ.shp" "data/CZ/CZ.mdb" "http://data.isric.org/edanube"
```

The alignment creates a local file `out.zip`, which contains the exported GML.

If an error occurs during the alignment, verify the paths in the call.

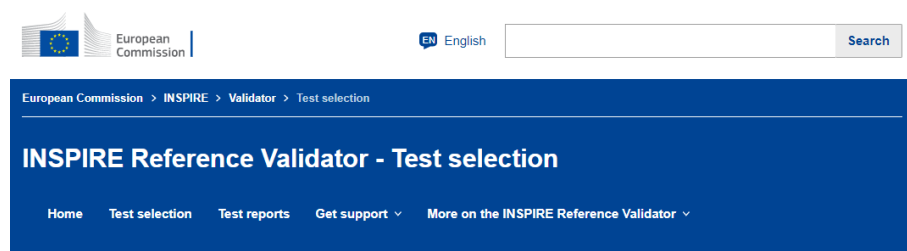
If an error keeps occurring, the most likely problem is that the model of your database files does not match with the model used in the definition of the alignment. Best option is to install Hale Studio (see section Developer Manual) and open the relevant project file and choose 'Transform external data' from the 'transformation' menu. This option starts a wizard which will guide you through all the steps of configuring the alignment.

GML files for the 14 Danube Basin countries (e.g. out-at.gml) as well as the whole basin are delivered with this report (see folder `Soter2Inspire\gml`).

Verify the outputs in the INSPIRE validator

The INSPIRE validator at <https://inspire.ec.europa.eu/validator> facilitates verification of the output of the alignment on INSPIRE compliance.

On the first page of the wizard select 'Data' on theme 'Soil', follow instructions of the `antispam widget`, select the generated GML file.



Configure your test

Select the INSPIRE resource you would like to test

- Metadata
- View Service
- Download Service
- Discovery Service
- Data set

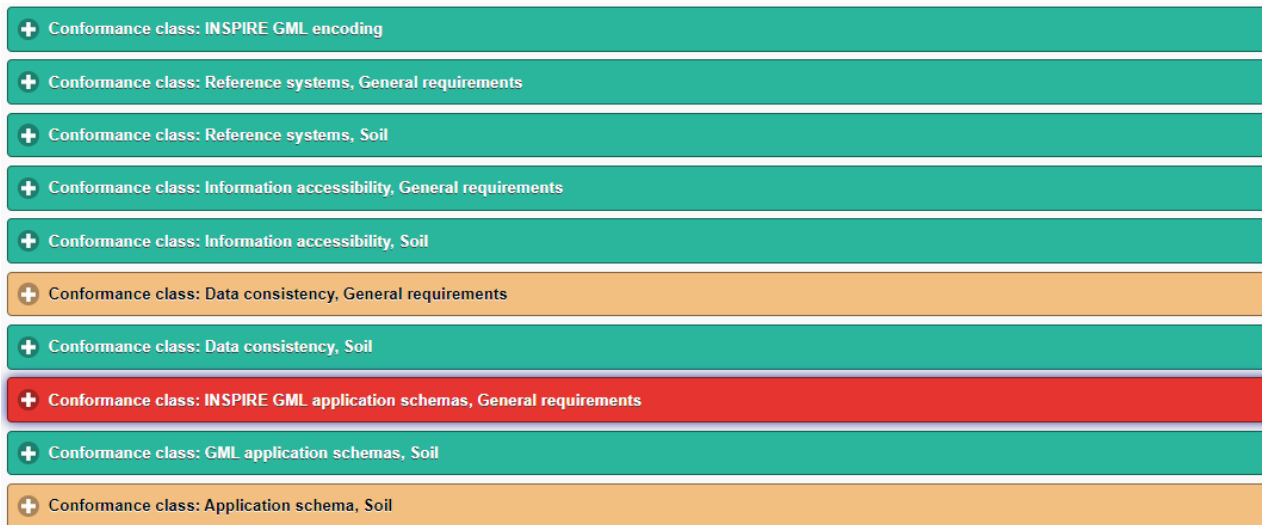
Select the data theme(s) relevant to the data set

- You can select more than one theme at a time.
- If no theme is selected, the data set will be only tested against the Conformance Classes for general requirements.
- For data themes displayed in grey below, specific conformance classes are not yet available. Proceed without any selection to test the dataset against the Conformance Classes for general requirements.

Annex I	Annex II	Annex III
-	-	Soil (SO)

The GML will be uploaded and the 'Start test' button will become available.

When the test is finished a report about the test run is available. In case a test failed (or is not machine testable) a warning will be shown. Click on the warning to find more details about the test result.

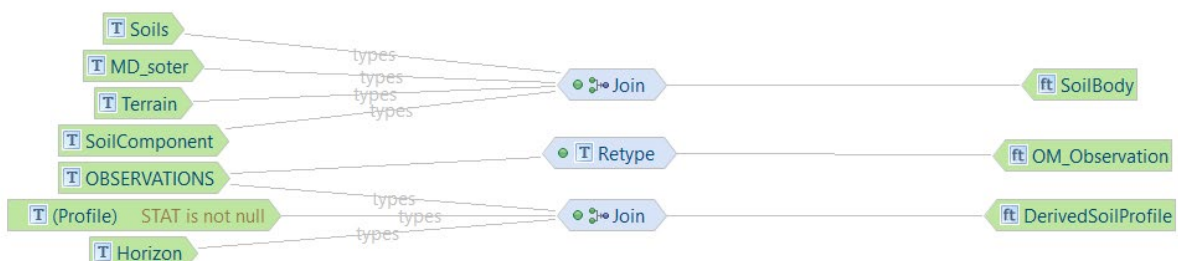


Developer Manual

In cases that the data model does not match the reference data model, or the output has unexpected aspects, the project can be opened in Hale Studio to polish the alignment. Hale Studio can be downloaded from <https://www.wetransform.to/downloads/> for Windows, macOS and Linux.

Alignment for Derived Profiles based on Horizons

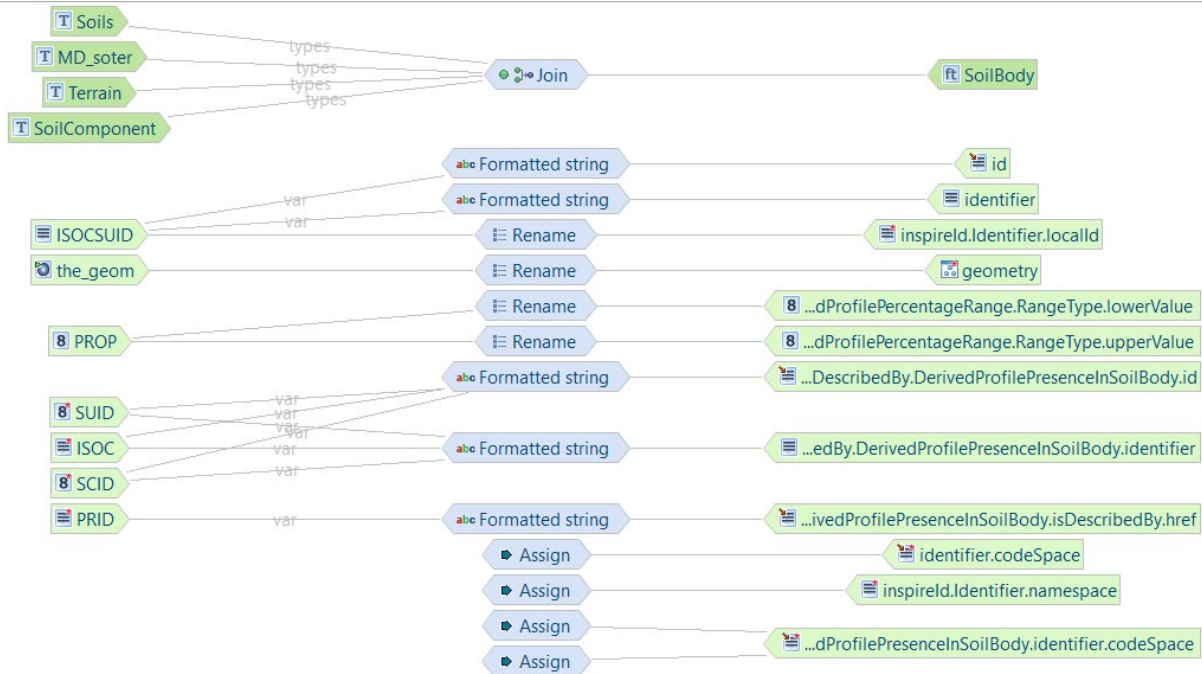
The alignment project contains 3 alignments for 3 main classes, SoilBody, DerivedSoilProfile and OM_Observation.



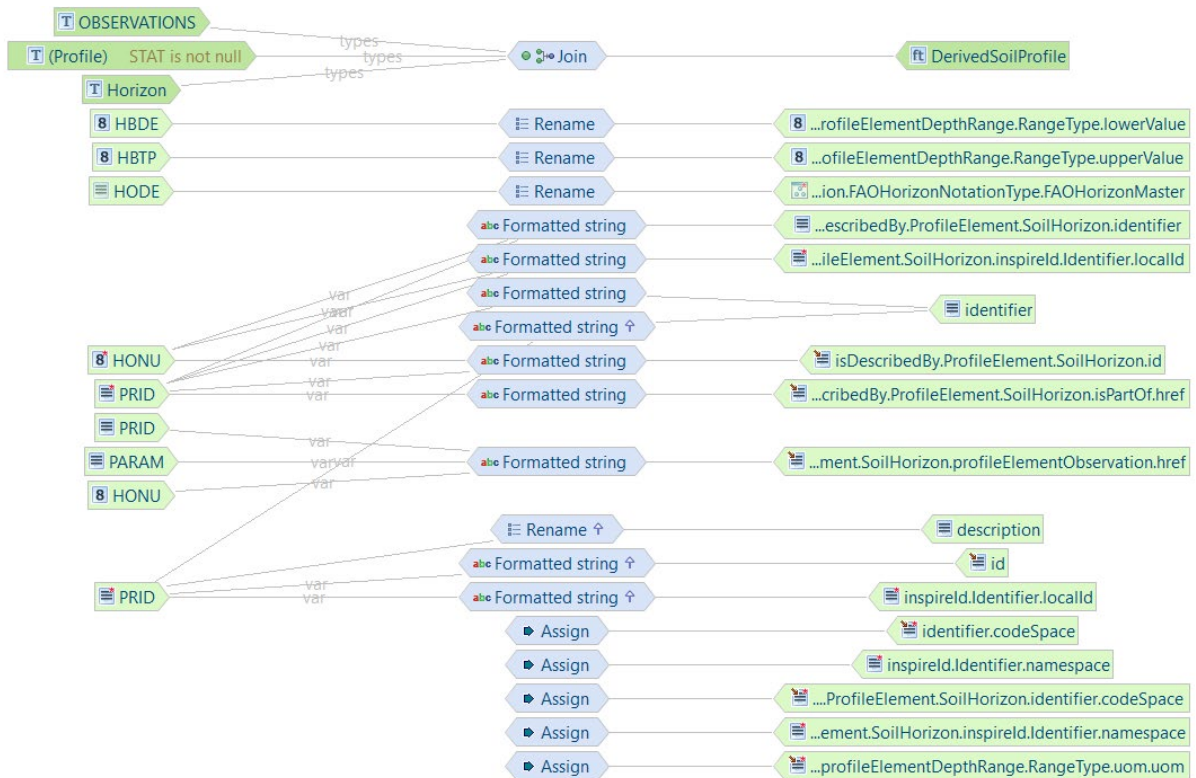
- A collection of Soil bodies, derived from a join of the shapefile (geometry), terrain, SoilComponents and Soils.

- A collection of DerivedSoilProfiles (including SoilHorizons) which reference the Soilbody they represent for a certain percentage.
- A collection of Observations which link to the soil profile they describe.

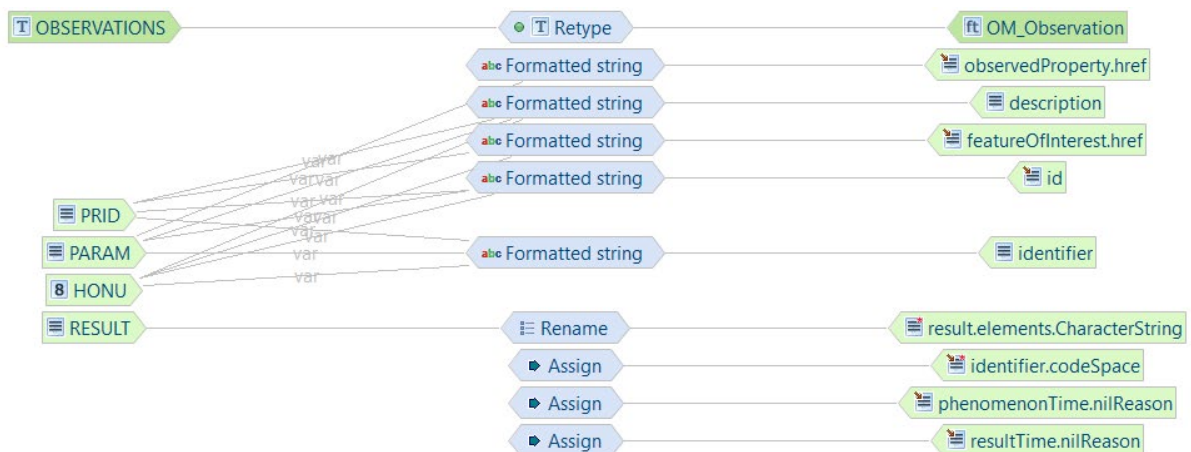
Only profiles are included which have STAT (i.e., soil profile description status) not null.



Above schema lists the elements which are used to populate the SoilBodies.



Above schema lists the properties which are included for each Soil Profile (and its horizons).



Above schema lists the properties of each measurement.

For defining identifiers we have followed the practice as described in <https://www.wetransform.to/news/2018/02/12/best-practices-for-inspire-ids>.

Alignment findings

This paragraph lists some aspects of the source data which we found while running some of the alignments.

A challenging aspect of the alignment design is the conversion of multiple observations in a single horizon row to individual OM_Observations as suggested by the INSPIRE model. We've tested two approaches for this aspect.

- Breaking up the horizon row in multiple observation entities using Hale filter mechanism.
- Adding a new table to the SOTER database which contains a row per observation (i.e., materialised view OBSERVATIONS).
- Using the Hale Scripting option to provide a Groovy script which creates the observation for each Horizon property.

We have selected the second approach, with the materialised view, as being much more robust than the first one. However, it requires database administrators to set up the table before exporting the data (see Annex 1). The third option may have given the same result, but requires advanced skills in a scenario that for example the database model has been modified.

Some of the datasets have a space or back slash in the profile identification. This is not according to the SOTER specification and leads to problems while transforming and validating the GML in the INSPIRE validator. The alignment now removes these characters from the identification.

Some shapefiles had illegal geometries (self intersecting). These issues have been resolved using the 'fix geometries' option in QGIS.

The INSPIRE validator 'complains about counterclockwise' presentation of the coordinates of the SoilBody polygons. Those polygons, however, validate fine in QGIS. This issue may be to discussion <https://github.com/inspire-eu-validation/ets-repository/issues/60>.

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.

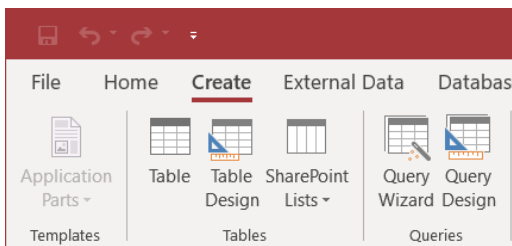
ISRIC has been supported by Wetransform gmbh, Darmstadt, Germany in the development of the Alignment project.

References

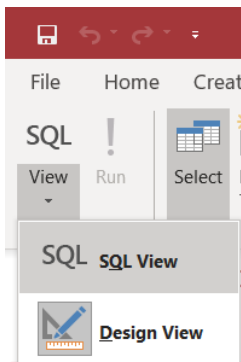
- Batjes NH 2021. *Integrating 14 Danube basin country databases into Danube_SOTER (Contributes to Deliverable D.B2, EU JRC Service Contract #936108, Collection of soil data in SOTER format from 14 Danube strategy countries, at scale 1:250 000). [Project report 4],, 7 p*
- Batjes NH and Ribeiro E 2019a. *Guidelines for compiling a 1:250,000 SOTER database (Deliverable 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, 59 p*
- Batjes NH and Ribeiro E 2019b. *Updated MS-Access SOTER template for the EU Danube basin project with worked examples (Deliverables 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, 30 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>
- INSPIRE 2013. *D2.8.III.3 Data specifications on soil - Technical Guidelines. INSPIRE Infrastructure for Spatial Information in Europe*, European Commission Joint Research Centre, Ispra, Italy.
http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_SO_v3_0.pdf
- Ruiperez Gonzalez M and Batjes NH 2019. *Country terrain unit shapefiles of the EU Danube basin (Deliverable 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, 25 p.*
https://www.isric.org/sites/default/files/Report02_SOTER_Danube_GIS_201906.pdf

Annex 1: Preparing the SOTER database

A new table needs to be created by running a query on the SOTER database. Open the given database, go to the Ribbon 'Create' tab, click and select 'Query Design'.



And switch to SQL view.



Copy the SQL script below in the empty screen (note: save query as e.g. 'Query-OBSERVATIONS'), then run the query. This will create a new table 'OBSERVATIONS', see screenshot below for an example. Update this table (e.g. run the 'Query-OBSERVATIONS'-query), before any alignment, so the latest changes in the database are included.

HONU	PRID	PARAM	RESULT
1 MD 1	1 MD 1	BULK	1,278702
1 MD 1	1 MD 1	CECS	29,73318
1 MD 1	1 MD 1	CLPC	41,59031
1 MD 1	1 MD 1	ORGC	35,42011
1 MD 1	1 MD 1	PHAQ	7,218373
1 MD 1	1 MD 1	SDTO	15,1702
1 MD 1	1 MD 1	STPC	43,24241
1 MD 1	1 MD 1	TOTN	3,523519

```

SELECT * INTO OBSERVATIONS FROM (
Select HONU,PRID,'SCMO' as PARAM,SCMO as RESULT from HORIZON where SCMO is not null union
Select HONU,PRID,'SCDR' as PARAM,SCDR as RESULT from HORIZON where SCDR is not null union
Select HONU,PRID,'MOCL' as PARAM,MOCL as RESULT from HORIZON where MOCL is not null union
Select HONU,PRID,'MOAB' as PARAM,MOAB as RESULT from HORIZON where MOAB is not null
union
Select HONU,PRID,'MOSZ' as PARAM,MOSZ as RESULT from HORIZON where MOSZ is not null union
Select HONU,PRID,'STGR' as PARAM,STGR as RESULT from HORIZON where STGR is not null union
Select HONU,PRID,'STSI' as PARAM,STSI as RESULT from HORIZON where STSI is not null union
Select HONU,PRID,'STTY' as PARAM,STTY as RESULT from HORIZON where STTY is not null union
Select HONU,PRID,'MINN' as PARAM,MINN as RESULT from HORIZON where MINN is not null union
Select HONU,PRID,'MINA' as PARAM,MINA as RESULT from HORIZON where MINA is not null union
Select HONU,PRID,'MINS' as PARAM,MINS as RESULT from HORIZON where MINS is not null union
Select HONU,PRID,'CFRA' as PARAM,CFRA as RESULT from HORIZON where CFRA is not null union
Select HONU,PRID,'CFRS' as PARAM,CFRS as RESULT from HORIZON where CFRS is not null union
Select HONU,PRID,'SDVC' as PARAM,SDVC as RESULT from HORIZON where SDVC is not null union
Select HONU,PRID,'SDCO' as PARAM,SDCO as RESULT from HORIZON where SDCO is not null union
Select HONU,PRID,'SDME' as PARAM,SDME as RESULT from HORIZON where SDME is not null union
Select HONU,PRID,'SDFI' as PARAM,SDFI as RESULT from HORIZON where SDFI is not null union
Select HONU,PRID,'SDVF' as PARAM,SDVF as RESULT from HORIZON where SDVF is not null union
Select HONU,PRID,'SDTO' as PARAM,SDTO as RESULT from HORIZON where SDTO is not null union
Select HONU,PRID,'STPC' as PARAM,STPC as RESULT from HORIZON where STPC is not null union
Select HONU,PRID,'CLPC' as PARAM,CLPC as RESULT from HORIZON where CLPC is not null union
Select HONU,PRID,'PSCL' as PARAM,PSCL as RESULT from HORIZON where PSCL is not null union
Select HONU,PRID,'BULK' as PARAM,BULK as RESULT from HORIZON where BULK is not null union
Select HONU,PRID,'ELEC' as PARAM,ELEC as RESULT from HORIZON where ELEC is not null union
Select HONU,PRID,'ELCO' as PARAM,ELCO as RESULT from HORIZON where ELCO is not null union
Select HONU,PRID,'SSO4' as PARAM,SSO4 as RESULT from HORIZON where SSO4 is not null union
Select HONU,PRID,'HCO3' as PARAM,HCO3 as RESULT from HORIZON where HCO3 is not null union
Select HONU,PRID,'SCO3' as PARAM,SCO3 as RESULT from HORIZON where SCO3 is not null union
Select HONU,PRID,'EXCA' as PARAM,EXCA as RESULT from HORIZON where EXCA is not null union
Select HONU,PRID,'EXMG' as PARAM,EXMG as RESULT from HORIZON where EXMG is not null union
Select HONU,PRID,'EXNA' as PARAM,EXNA as RESULT from HORIZON where EXNA is not null union
Select HONU,PRID,'EXCK' as PARAM,EXCK as RESULT from HORIZON where EXCK is not null union
Select HONU,PRID,'EXAL' as PARAM,EXAL as RESULT from HORIZON where EXAL is not null union
Select HONU,PRID,'EXAC' as PARAM,EXAC as RESULT from HORIZON where EXAC is not null union
Select HONU,PRID,'CECS' as PARAM,CECS as RESULT from HORIZON where CECS is not null union
Select HONU,PRID,'TCEQ' as PARAM,TCEQ as RESULT from HORIZON where TCEQ is not null union
Select HONU,PRID,'GYPS' as PARAM,GYPS as RESULT from HORIZON where GYPS is not null union
Select HONU,PRID,'P2O5' as PARAM,P2O5 as RESULT from HORIZON where P2O5 is not null union
Select HONU,PRID,'TOTP' as PARAM,TOTP as RESULT from HORIZON where TOTP is not null union
Select HONU,PRID,'PRET' as PARAM,PRET as RESULT from HORIZON where PRET is not null union
Select HONU,PRID,'FEDE' as PARAM,FEDE as RESULT from HORIZON where FEDE is not null union
Select HONU,PRID,'ALOE' as PARAM,ALOE as RESULT from HORIZON where ALOE is not null union
Select HONU,PRID,'FEOE' as PARAM,FEOE as RESULT from HORIZON where FEOE is not null union
Select HONU,PRID,'CLAY' as PARAM,CLAY as RESULT from HORIZON where CLAY is not null union
Select HONU,PRID,'TOTC' as PARAM,TOTC as RESULT from HORIZON where TOTC is not null union
Select HONU,PRID,'TOTN' as PARAM,TOTN as RESULT from HORIZON where TOTN is not null union
Select HONU,PRID,'ORGC' as PARAM,ORGC as RESULT from HORIZON where ORGC is not null union
Select HONU,PRID,'PHAQ' as PARAM,PHAQ as RESULT from HORIZON where PHAQ is not null
)

```


Annex 2: Alignment based on alternative classes

The INSPIRE Soil Schema facilitates 3 typical Soil Use cases (Figure 1):

- Storage of derived soil profiles based on horizons to represent a percentage of a soil body
- Storage of derived soil profiles based on layers to represent a percentage of a soil body
- Storage of soil profiles based on horizons from actual soil observations at soil plots

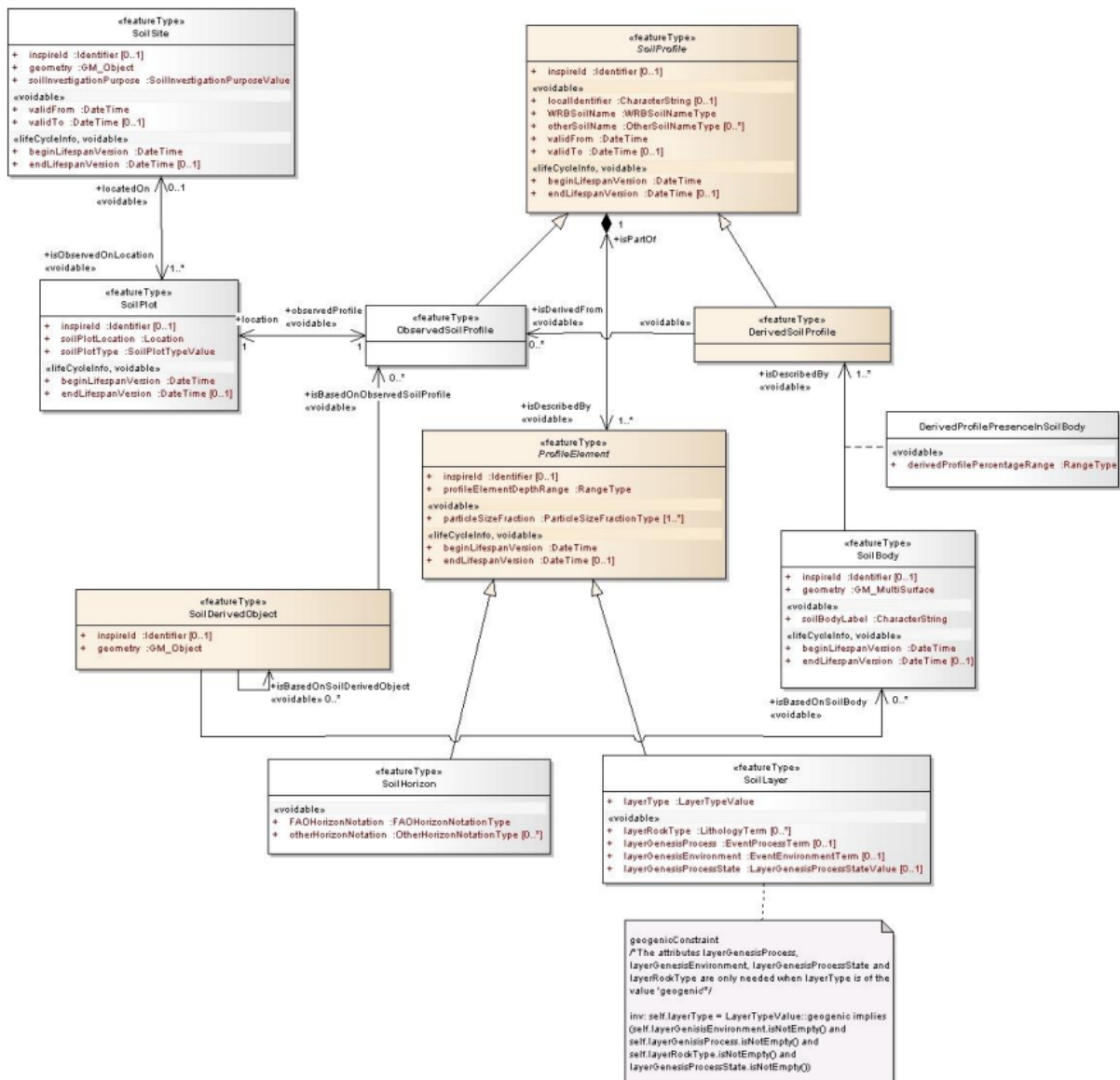


Figure 1. UML diagram Soil from D2.8.III.3 Data Specification on Soil – Technical Guidelines (Templer and others, 2013).

In this , we have only worked with the first approach, ‘Storage of derived soil profiles based on horizons to represent a percentage of a soil body.’ However, we noticed different approaches were used by the data compilers when filling the attribute databases, for a summary see Batjes (2021). We have started

alignment projects for the second and third approach, but these have not been completed and these are not a formal part of the deliverable.

Alignment for Derived Profiles based on Layers

SoilGrids is a typical example of a dataset oriented on layers. Layers have fixed heights and lack a FAO classification. The model is similar as above, but instead of SoilHorizons, SoilLayers are used. Typically, such cases should not occur in SOTER databases that follow the actual specifications (Batjes and Ribeiro 2019a).

Alignment for Observed Profiles

The Soilbody type cannot be used if the dataset contains profiles based on actual observations (which would be the case for fully compliant SOTER databases, see procedures manual (Batjes and Ribeiro 2019a)). Instead the SoilDerivedObject is used, which has a relation to ObservedSoilProfile via the isBasedOnObservedSoilProfile property. ObservedSoilProfile is linked to a SoilPlot which contains the location of the ObservedSoilProfile. SoilSite is not populated, since SOTER does not capture these aspects.