Towards climate-smart sustainable management of agricultural soils

GeoPackage Encoding Rules

Of the

Deliverable 6.4

Software framework for a shared agricultural soil information system

Version: 1.0

Date: 31 January 2024

GENERAL DATA

Grant Agreement: 862695
Project acronym: EJP SOIL
Project title: Towards climate-smart sustainable management of agricultural soils
Project website: www.ejpsoil.eu

Start date of the project: February 1st, 2020
Project duration: 60 months
Name of lead contractor: INRAE

Funding source: H2020-SFS-2018-2020 / H2020-SFS-2019-1
Type of action: European Joint Project COFUND

Encoding Rules EDITED BY:	Stefania Morrone (Epsilon Italia)
Authors of the Encoding Rules	Stefania Morrone (Epsilon Italia), Maria Fantappiè (CREA), Giovanni L'Abate (CREA), Andrea Lachi (CREA).
DELIVERABLE NUMBER:	6.4
DELIVERABLE TITLE:	Software framework for a shared agricultural soil information system
DELIVERABLE TYPE:	Software
WORK PACKAGE N:	WP6
WORK PACKAGE TITLE:	Supporting harmonised soil information and reporting
DELIVERABLE LEADER:	Maria Fantappiè, Council for Agricultural Research and Economics (CREA)
DISSEMINATION LEVEL:	PU

Table of Contents

List of Figures

List of acronyms and abbreviations

API	Application Programming Interface
GML	Geography Markap Language
GPKG	Geopackage
INSPIRE	Infrastructure for Spatial Information in the European Community
IR	Implementing Rules
O&M	Observation and Measurements
OGC	Open Geospatial Consortium
SQL	Structured Query Language
STA	Sensor Things API
UML	Unified Modeling Language
URL	Uniform Resource Locator

1 Introduction

This document describes the Geopackage model for the INSPIRE data theme Soil, as developed in the context of the EJP SOIL project, deliverable D6.4 "Software framework for a common agricultural soil information system". Such a software framework, aimed to enable transcoding and streamlining of interoperable and harmonised national agricultural soil data, builds on the solid foundations of the INSPIRE Directive, taking advantage of the latest developments in the process of modernisation and simplification of its technical requirements. In particular, while the INSPIRE Soil conceptual model has been referenced as data model, the INSPIRE GeoPackage alternative encoding has been selected as data encoding format, as it is considered more appropriate for the exchange of soil data compared to the GML 3.2.1, which is the default encoding of INSPIRE data. As a (SQLite3) database, in fact, a single GeoPackage file can store multiple vector layers and handle large numbers of spatial features and tables in a small file size. Its database capabilities speed up (compared to e.g. shapefiles and even more so to complex features GMLs) common operations such as loading, updating and querying vector layers in QGIS or similar geographic information management tools. The GeoPackage format is also broadly compatible across software environments, and it is particularly efficient where connectivity and bandwidth are limited.

The Soil GeoPackage model described hereafter is to be considered as a specific implementation for the Soil data theme of the INSPIRE Good Practice on the GeoPackage encoding of INSPIRE datasets, which describes the mechanism for creating INSPIRE data sets in the GeoPackage encoding format in compliance with the INSPIRE Implementing Rules (IR).

The INSPIRE compliance of the soil GeoPackage datasets encoded according to the model described in this deliverable has been demonstrated by meeting all the requirements of the abovementioned Good Practice (see Section 6: *INSPIRE validation of EJP SOIL GeoPackage* datasets). Therefore, the developed model for a common agricultural soil information system not only serves the needs for an efficient exchange of harmonised national agricultural soil data but is also functional to derive soil data that comply with the INSPIRE requirements.

This work will be published in the INSPIRE GitHub repository that gathers best practices for GeoPackage encodings of INSPIRE datasets, and added to the implementation examples list as specific implementation for the Soil data theme.

2 Normative References

INSPIRE Soil data specifications

INSPIRE Good Practice: GeoPackage encoding of INSPIRE datasets

GeoPackage Encoding Standard version 1.3.0

OGC SensorThings API Part 1: Sensing Version 1.1

3 Model Transformation Rules.

To transform the Soil UML conceptual model to corresponding logical GeoPackage schema, the following transformation rules have been applied:

- generic (encoding agnostic) INSPIRE model transformation rules,
- encoding-specific INSPIRE UML-to-GeoPackage encoding rule,
- transformation rules developed specifically for the soil data theme and in the scope of the EJP SOIL GeoPackage work.

In this transformation process, simplifications have been made to limit the number of tables and facilitate data management and use, while retaining all the information needed to describe the complex objects of INSPIRE Soil. Wherever deemed appropriate, the complex (nested) data structures and the unbound cardinalities present in the INSPIRE model have been reduced e.g., using simple attributes and /or fixed multiplicity.

3.1 Generic and encoding-specific transformation rules

The generic and encoding-specific transformation rules that have been applied are listed below:

1) *Flattening of hierarchical structures and data types:*
    a) For multiplicities 0 or 1 the applied rule is based on both the INSPIRE model transformation rule Flattening of Nested Structures (MT001) and the GeoPackage flattening rule.
    b) For multiplicities greater than 1 (not addressed by abovementioned INSPIRE flattening rules), tables for the data types have been created. Associations of the original tables have been managed according to the rules 4) and 5) of this list, which target management of the associations.
2) *Properties*

    GeoPackage Properties rule has been applied.
3) *Handle associations with 0..1 or 1:1 multiplicity :*

    Attributes of the destination table are incorporated in the source table.
4) *Handle associations with 1:n multiplicity*

    A related attribute is introduced in the destination table which is foreign key to the source table.
5) *Handle associations with n:m multiplicity*

    Many-to-many relationships have been represented using a relationship table. A relationship table is composed of a primary key and two foreign keys (one per associated table).
6) *Voidable*

    The GeoPackage Voidable rule has been applied

3.2 Soil data theme specific transformation rules

Find below sections 3.2.1 and 3.2.2 the specific rules which have been applied respectively to the spatial feature types describing common artifacts of soil sampling (such as profiles, layers, horizons) and to the set of features (O&M Observations) providing information that has been observed or measured on the spatial feature types.

3.2.1 Rules for the spatial feature types

1) *SoilProfile*

The *ObservedSoilProfile* and *DerivedSoilProfile*, specialization with no attributes of the *SoilProfile* feature type, have no corresponding tables in the GPKG model. To distinguish between an observed and a derived profile the "*isderived*" BOOLEAN attribute has been added to the soilprofile table. The attribute is set to TRUE when the profile is derived. In addition, a specific trigger checks that derived soil profiles have null values for the location attribute (association to soilplot).

Figure 1 below illustrates the flattening of the SoilProfile feature type.

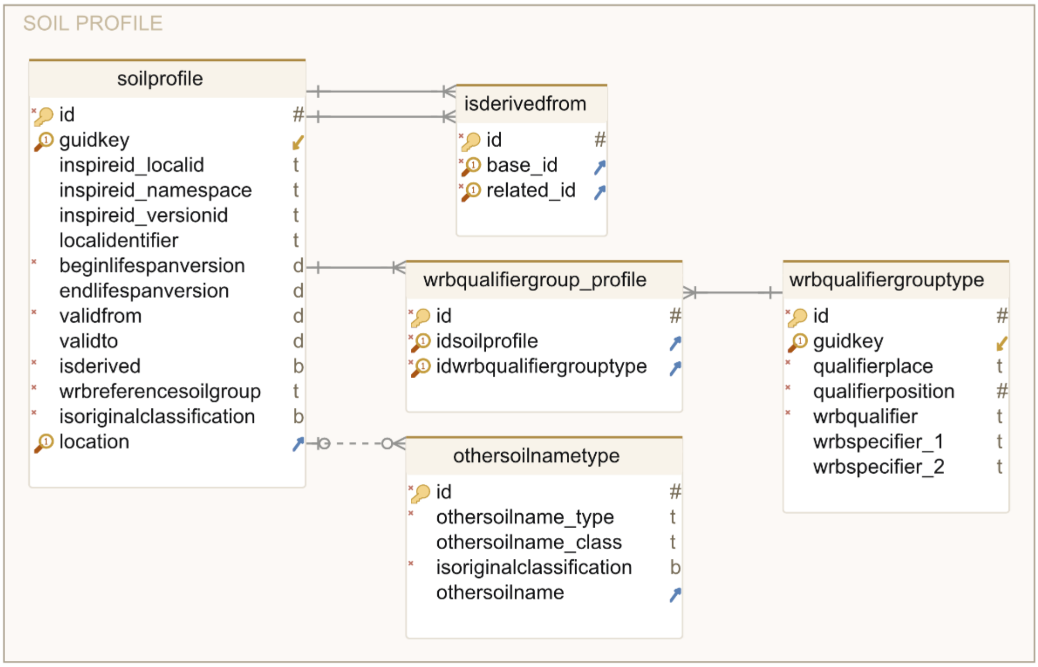


Figure 1 Soil Profile

2) *SoilPlot: soilplotlocation POINT NOT NULL*

In the INSPIRE model, the location of the soil plot (soilPlotLocation) can also be provided in the form of a textual field (this element is defined as "a reference to a location on the earth; it can be a point location identified by coordinates or a description of the location using text or an identifier"). However, in the context of EJP SOIL, it is preferred that the coordinates of the plot are always available. The soilPlotLocation field has been assigned the POINT data type.

Figure 2 below illustrates the flattening of the SoilSite and SoilPlot feature types.

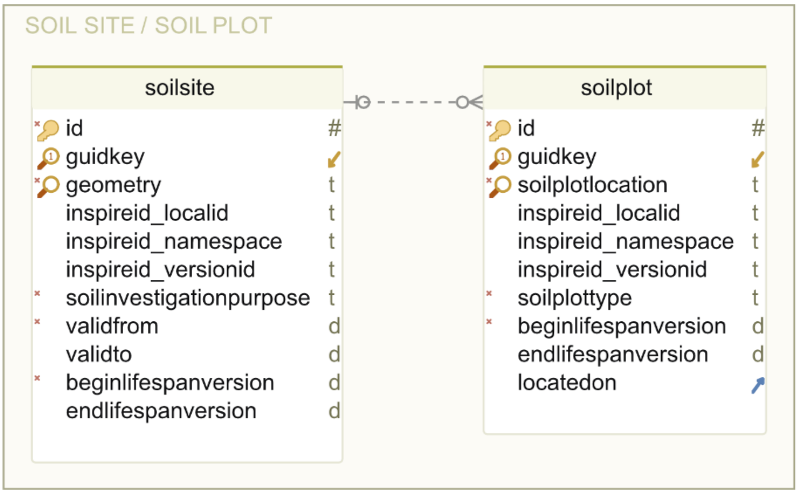


Figure 2 SoilSite and SoilPlot

3) *ProfileElement*

The *SoilHorizon* and *SoilLayer*, specialization of the *ProfileElement* feature type, have no corresponding tables in the model. All the SoilLayer attributes have been incorporated in the ProfileElement table. Conversely, the Horizon attributes (FAOHorizonNotation, complex data type with multiplicity 1, and otherHorizonNotation complex data type with multiplicity 0..n) are represented as tables, namely the FAOHorizonNotationType and OtherHorizonNotationType tables, linked to the ProfileElement by related foreign keys.
Depending on the ProfileElement feature being a layer or horizon certain fields are (or are not) compulsory. This is managed through the creation of relevant triggers.
Figure 3 below illustrates the flattening of the SoilProfileElement feature type.

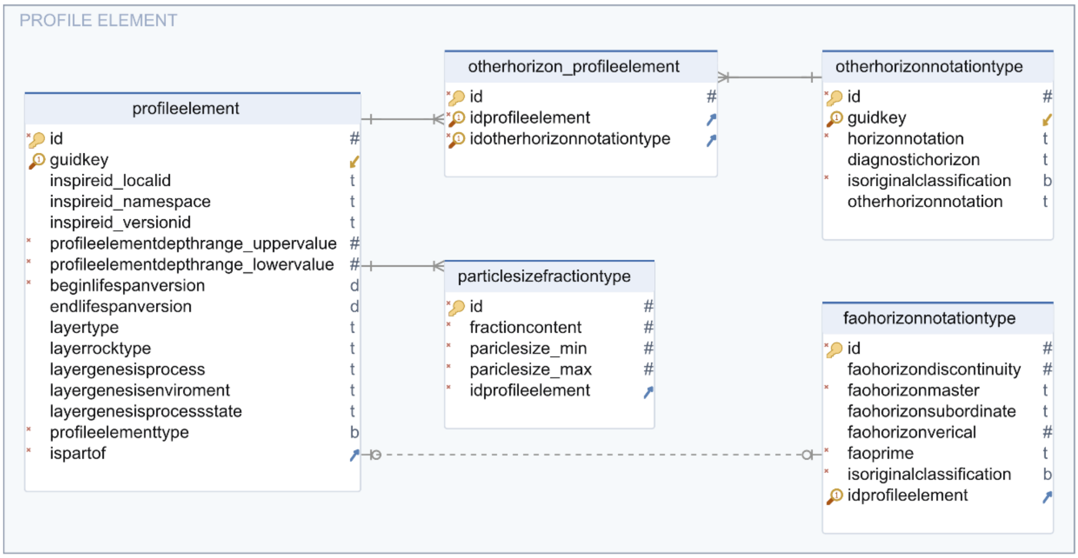


Figure 3 ProfileElement

4) *SoilBody geometry*

The same soil body refers different geometries representations (soil maps at different scales and level of details). For this reason, the geometry of a soil body is handled in a dedicated soilbody_geom table. The spatial objects in the soilbody_geom table are linked to the related soil body via a foreign key.
Figure below illustrates the flattening of the SoilBody feature type.

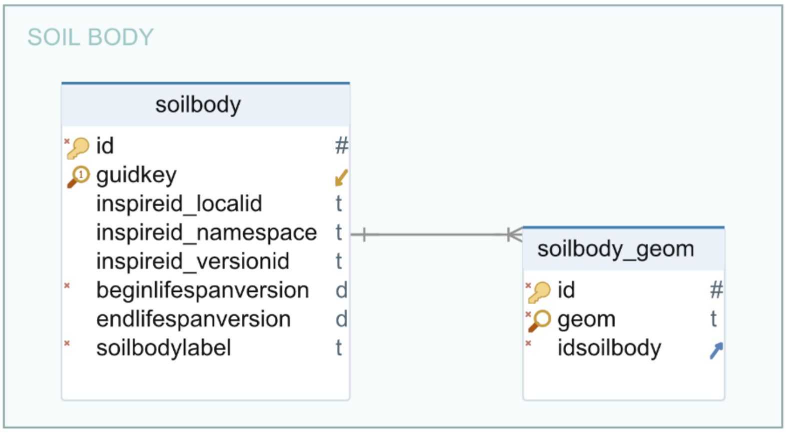


Figure 4: soilbody and soilbody_geom tables

3.2.2 Rules for the Observation and Measurements

For the type of soil data of interest, which is mostly (though not exclusively) based on sensor detections, the observation and measurement part has been modelled based on the OGC SensorThings API (STA) model. This model provides a standard way to manage and retrieve observations and metadata from sensor systems and thus meets the requirements for managing sensor data and their datastreams. However, it is always possible to conform to the INSPIRE soil data model as all the necessary information is available: each observation is linked by a foreign key to its datastream, which in turn is linked to its feature of interest (foreign keys to feature type specific tables, e.g. profileelement, soilderivedobject, etc.) as well as to its sensor and related observed property and process.

The list of specific rules applied accompanied by a brief description is provided below:

1) *Datastream and DatastreamCollection*

In general, a measuring instrument (modelled by the STA “Thing” concept) contains multiple sensors that share the same observed area and the same phenomenon and result time, but each with its own observed property and process and its own data stream. The output of the instrument, however, is a single file containing the data streams of all the sensors.
In order to efficiently model this scenario and to be able to access both the specific information on the measuring instrument output (single file with measurements from all the sensors) and the measurements of individual sensors, the STA model has been extended to include the concept of data stream collection (DatastreamCollection).
Specifically,

- the objects in the datastreamcollection table correspond to physical files produced by the measuring instrument,
- the objects in the datastream table correspond to measurements (datastreams) of the specific sensors.

The association between a datastream and its related datastreamcollection is handled by the corresponding foreign key in the datastream table.
In particular, the datastream table allows to link the observations to the related:

- feature of interest (soilsite, soilprofile, profileelement, soilderivedobject),
- observed property,
- process,
- sensor,
- observed area, phenomenon and result times (via the link to the datastreamcollection table).

To be noted that, in this context, the sensor is intended as “an entity capable of observing a phenomenon and returning an observed value” i.e., a physical sensing instrument.

Figure 5 below illustrates the modelling of observation, datastream and datastreamcollection.

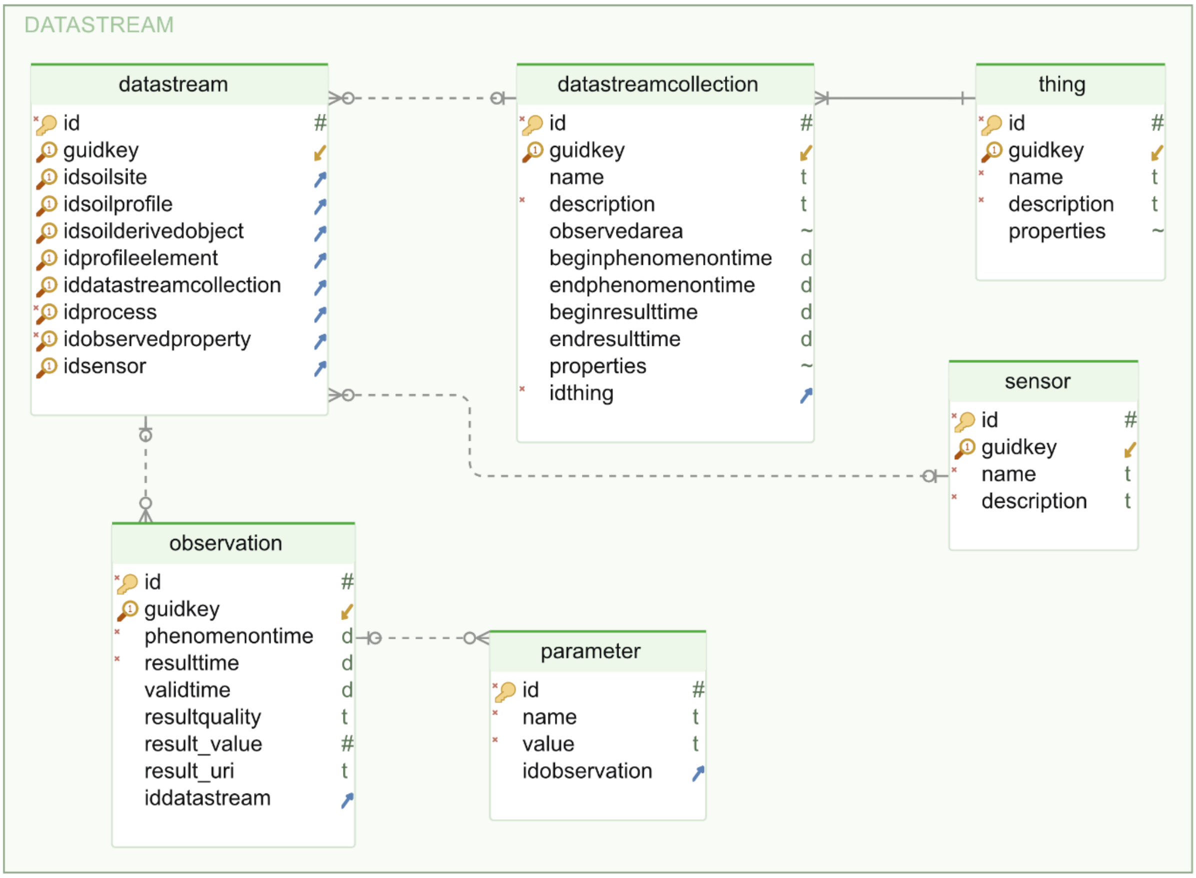


Figure 5 observation, datastream, datastreamcollection tables

Specific constraints are implemented to ensure that:

- only one feature of interest is associated to a specific datastream,
- a feature of interest in a datastream corresponds to a unique combination of property/process/sensor/datastreamcollection
- only meaningful property/process pairs exist.

2) *ObservableProperty and Process*

To efficiently implement the checks ensuring the consistency of observable properties (which feature of interest can be associated with which property, what is the domain of a particular property, etc.), the concept of observable property, which in the INSPIRE data model is characterized by a label, a base phenomenon and a unit of measure, has been extended to include the property definition and description fields plus information on:

- the domain (min and max values, type of value and code),
- the phenomenon type (biological, chemical, physical),
- the feature of interest.

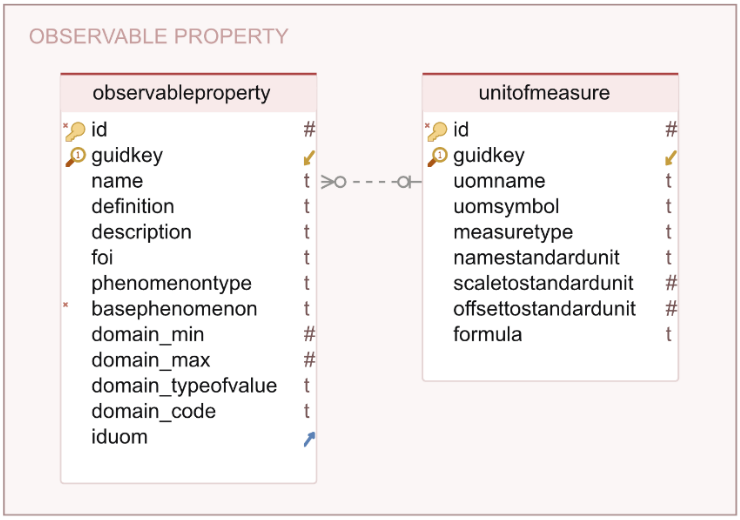


Figure 6 observable property and unit of measure tables

Moreover, the observableproperty_process table illustrated below allows to associate an observedproperty to the related process. Through a specific constraint on the datastream table, only (observedproperty, process) pairs in the observableproperty_process table are allowed to be part of the same record in the datastream table.

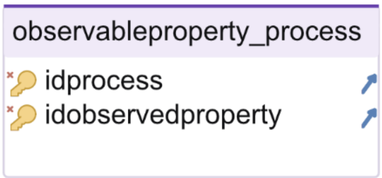


Figure 7 observableproperty_process table

4 Model-embedded Validation Rules

Foreign keys and triggers have been created to enforce INSPIRE Soil Data Specification constraints and maintain the data integrity of the GeoPackage database.

In order not to violate the triggers (and thus be able to save the information in the GeoPackage file), it is necessary to follow the recommended loading order provided below.

In any case, a version of the template without triggers was also created to facilitate reuse of the work done.

4.1 Recommended" loading order for feeding data into the Geopackage.

Find in bold the elements whose order must be strictly respected.

VECTOR

- **soilsite** (Feature)
- **soilplot** (Feature)
- **soilprofile**
- isderivedfrom
- **profileelement**
- **soilbody**
- **soilbody_geom** (Feature)
- derivedprofilepresenceinsoilbody
- **soilderivedobject** (Feature)
- isbasedonsoilderivedobject
- isbasedonsoilbody
- isbasedonobservedsoilprofile
- wrbqualifiergrouptype
- wrbqualifiergrou_profile

- othersoilnametype
- faohorizonnotationtype
- particlesizefractiontype
- otherorizonnotationtype
- otherhorizon_profileelement

O&M/Sensor Thing

- **unitofmeasure**
- **observableproperty**
- relatedparty
- **process**
- **observableproperty_process**
- thing
- datastreamcollection
- sensor
- **datastream**
- **observation**
- processparameter
- documentcitation
- parameter

5 Codelist management

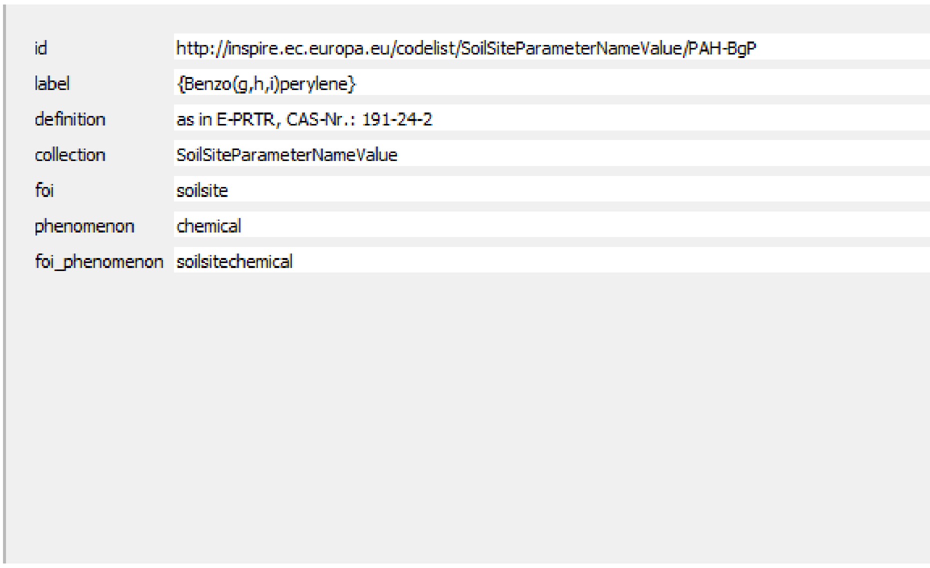
In data modelling activities, the use of code lists is a fundamental means of ensuring data consistency and providing a common vocabulary for the community.

In the development of this model for an interoperable and harmonised national agricultural soil data, a particular care has been paid to the management of code lists. This especially because, in the case of the Soil data theme, the INSPIRE code lists are currently mostly empty or contain too few values (only for more relevant concept) to guarantee the richness and completeness of the soil vocabularies.

A codelist table has been created, in which each codelist value:

- is described through a label and a definition,
- classified through the relevant collection (codelist) name,
- assigned unique identifier, corresponding to the codelist value URL,

- is associated with the type of phenomenon and the type of feature of interest.



id	http://inspire.ec.europa.eu/codelist/SoilSiteParameterNameValue/PAH-BgP
label	{Benzo(g,h,i)perylene}
definition	as in E-PRTR, CAS-Nr.: 191-24-2
collection	SoilSiteParameterNameValue
foi	soilsite
phenomenon	chemical
foi_phenomenon	soilsitechemical

Figure 8 codelist record example.

The structure of this table is aligned to the template for entering codelist / codelist values in INSPIRE registries. This will facilitate extension of INSPIRE codelists through INSPIRE national registries to provide new /more detailed concepts (narrower terms) complementing the ones already available via the INSPIRE codelist registry.

The codelist table also allows for the creation of quality checks for data validation purposes.

6 INSPIRE validation of EJP SOIL GeoPackage datasets

The suitability of the model described in this document (and the associated geopackage template) to deliver INSPIRE soil data is guaranteed by its compliance with all requirements of the INSPIRE Good Practice on the GeoPackage encoding of INSPIRE datasets.

Specifically, the following evidence has been provided:

- description of the UML-to-Geopackage model transformation rules (see Section 3: Model Transformation Rules.)
- empty geopackage template acting as database schema (also provided via SQL scripts)
- executable model for data transformation of GeoPackage datasets into INSPIRE GML datasets. Specifically, this executable model is provided in the form of a hale studio project, mapping soil GeoPackage datasets conforming to the model therein described into corresponding INSPIRE Soil GML datasets.

- a sample dataset obtained through the above-mentioned hale data transformation project, successfully validated against INSPIRE requirements using the [INSPIRE Validator.]()

It is worth highlighting that, even if not strictly required by the Good Practice, this model also makes use of triggers to enforce constraints of the INSPIRE UML data model and maintain data integrity.

A geopackage file, structured according to the model described in this document, has been filled with soil data (from the CREA soil database). Using a [hale studio]() project, this geopackage data has been transformed into an INSPIRE Soil GML dataset which has been successfully validated with the INSPIRE Validator.

Find below one small excerpt of the abovementioned GML illustrating the encoding of a SoilHorizon object with related observations and in figure 10 a screenshot of the successful INSPIRE validation result:


```xml
<gml:featureMember>
  <so:SoilHorizon gml:id="_A3ADD08A-F2AB-4EB6-8D8C-1982BDE43FA2">
    <so:particleSizeFraction xsi:nil="true"/>
    <so:profileElementDepthRange>
      <so:RangeType>
        <so:upperValue>50.0</so:upperValue>
        <so:lowerValue>55.0</so:lowerValue>
        <so:uom uom="cm"/>
      </so:RangeType>
    </so:profileElementDepthRange>
    <so:beginLifespanVersion>2008-01-01T00:00:00Z</so:beginLifespanVersion>
    <so:profileElementObservation xlink:href="#_A953E51B-6A92-4680-BC31-4D76DDB7B34C"/>
    <so:profileElementObservation xlink:href="#_C6F804AE-6EE8-4578-88D2-5B22999DEB70"/>
    <so:profileElementObservation xlink:href="#_9B9C7490-2BEA-4AE3-A39A-27B1101BF4F3"/>
    <so:profileElementObservation xlink:href="#_3A6A8743-754B-41C8-BACA-D662B2753CEC"/>
    <so:profileElementObservation xlink:href="#_A87B1640-7E67-4E02-91DB-070CA60CE70B"/>
    <so:isPartOf xlink:href="#_4F914E54-AB37-41D5-855F-3BBEFC4D1F97"/>
    <so:FAOHorizonNotation>
      <so:FAOHorizonNotationType>
        <so:FAOHorizonMaster xlink:href="https://aims.fao.org/aos/agrovoc/c_016fddeb"/>
        <so:FAOPrime xlink:href="https://inspire.ec.europa.eu/codelist/FAOPrimeValue/0"/>
        <so:isOriginalClassification>false</so:isOriginalClassification>
      </so:FAOHorizonNotationType>
    </so:FAOHorizonNotation>
  </so:SoilHorizon>
 </gml:featureMember>

<gml:featureMember>
  <om:OM_Observation gml:id="_A953E51B-6A92-4680-BC31-4D76DDB7B34C">
    <om:phenomenonTime>
      <gml:TimeInstant>
        <gml:timePosition>2024-01-11T17:51:52.829Z</gml:timePosition>
      </gml:TimeInstant>
    </om:phenomenonTime>
    <om:resultTime>
      <gml:TimeInstant>
```

```xml
      <gml:timePosition>2024-01-11T17:51:52.844Z</gml:timePosition>
    </gml:TimeInstant>
  </om:resultTime>
  <om:procedure xlink:href="#_31C6269A-877E-462A-8E5A-38504F5CEE50" xlink:title="Water Content Volumetric"/>
  <om:observedProperty xlink:href="http://www.eionet.europa.eu/gemet/concept/7874" xlink:title="Water Content Volumetric"/>
  <om:featureOfInterest xlink:href="#_A3ADD08A-F2AB-4EB6-8D8C-1982BDE43FA2"/>
  <om:result>
    <gco:Measure uom="%vol">0.043</gco:Measure>
  </om:result>
</om:OM_Observation>
```

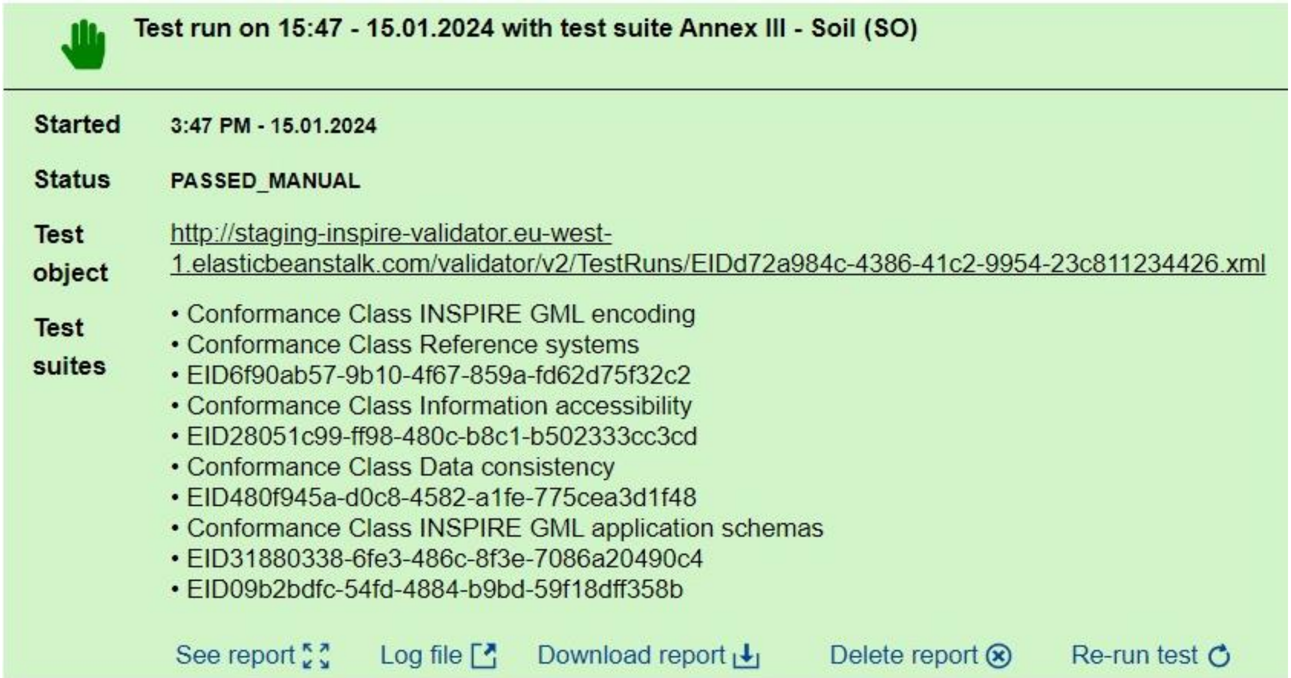



Figure 9 successful INSPIRE Validation

7 Conclusions

A software framework, aimed to enable transcoding and streamlining of interoperable and harmonised national agricultural soil data, has been designed and tested. The underlying model builds on the solid foundations of the INSPIRE Directive, taking advantage of the latest developments in the process of modernisation and simplification of its technical requirements. In particular, it is to be considered as a specific implementation of the INSPIRE Good Practice on the GeoPackage encoding of INSPIRE datasets for the INSPIRE Soil data theme. As detailed in Section 6, all requirements of the INSPIRE Good Practice have been met and therefore the model not only serves the needs for an efficient exchange of harmonised national agricultural soil data but is also functional to derive soil data that comply with the INSPIRE requirements.

This work will be published in the INSPIRE GitHub repository that gathers best practices for GeoPackage encodings of INSPIRE datasets, and added to the implementation examples list so that it can be shared and reused by the INSPIRE community.

Regarding the usability and efficiency of the developed geopackage model, tests in the QGIS environment have shown how straightforward it is to access the information contained in the derived geopackage soil data. Clicking on a spatial feature will in fact display all the information associated with it (e.g. given a site, it will show which plots are associated with it, with details of associated profiles, observable properties and observations). In this regard, it is worth mentioning the work done by CREA in the creation of data entry masks in the QGIS environment that are functional for the use of geopackage soil model in operational environments and for field surveys.

Of particular note is the work to facilitate semantic harmonisation of soil data described in Section 5 *Codelist management*.

The work carried out has also been noted by the European Environment Agency for possible reuse in terms of developing soil-related environmental reporting models.