Annex

1. Database System Overview

1.1. Data Curation

In AsphaltMine, users add data to the database either via an online form or in bulk via a predefined excel sheet. The difference between these two is as follows:

- **AsphaltMine Web Form:** an interactive web entry form that presents fields matching the database schema, where the user fills in values for a single mixture with corresponding one or multiple performance tests and submits it to the database. In this form, users select the test types and complete the required fields for each record in the browser. The form shows a detailed description of each input field, including the unit of measurement, the standard number, and a tooltip with an example. The web form is suitable for entering few mixtures at a time.
- **AsphaltMine Predefined Excel File:** It is a predefined excel file (Download here: Asphalt-Mine Excel Sheet) that matches the fields of the aforementioned online form. The excel file includes the measurement unit under each field and the description of the field when hovering over its name. There are two data entry sheets in the file which are identical in terms of fields, but they differ as follows:
  - **Columns sheet:** the parameters are organized in columns, and the records are to be filled in rows as shown in Figure 1.
  - **Rows sheet:** the parameters are organized in rows, and the records are to be filled in columns.

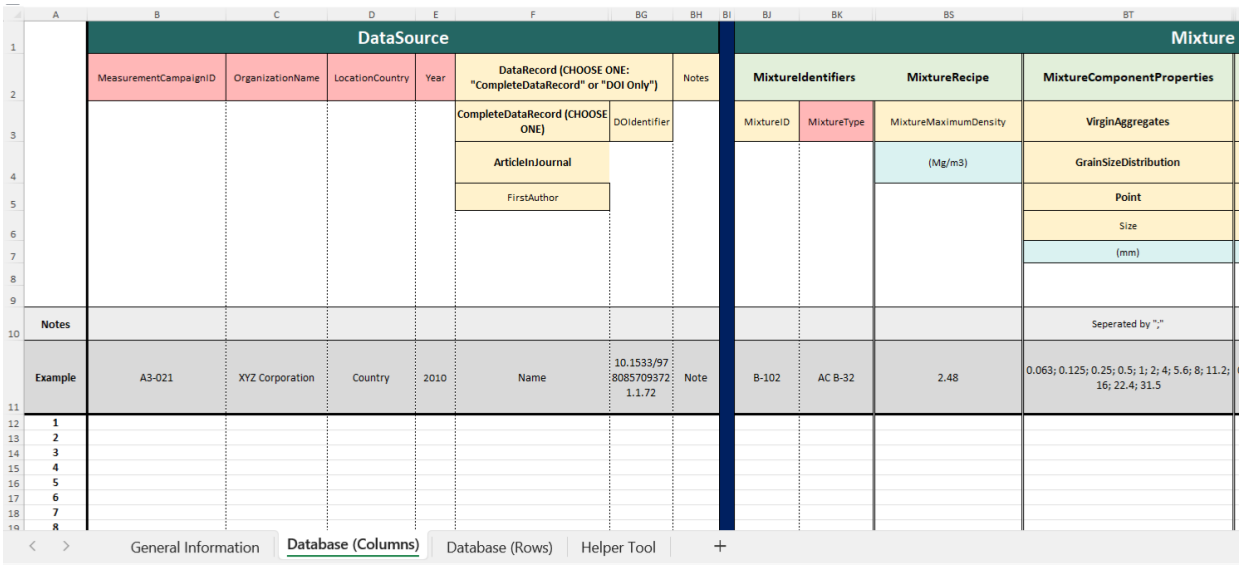


Figure 1: The column sheet of the AsphaltMine predefined excel file

The choice of the sheet is based on the user's preference and allows rapid copying of the data from local databases. The excel file is useful for inputting larger datasets, as it can read hundreds of records at once for different tests and then upload them in the web form in a batch. The file is available to download locally through the web application, allowing to compile it over longer time before uploading the data into AsphaltMine. Plugins can also be developed to transform large unified excel files into AsphaltMine predefined excel format, facilitating curation of data.

1.2. Internal Data Saving Process

This internal data saving process is summarized in Figure 2 and explained in more detail below.

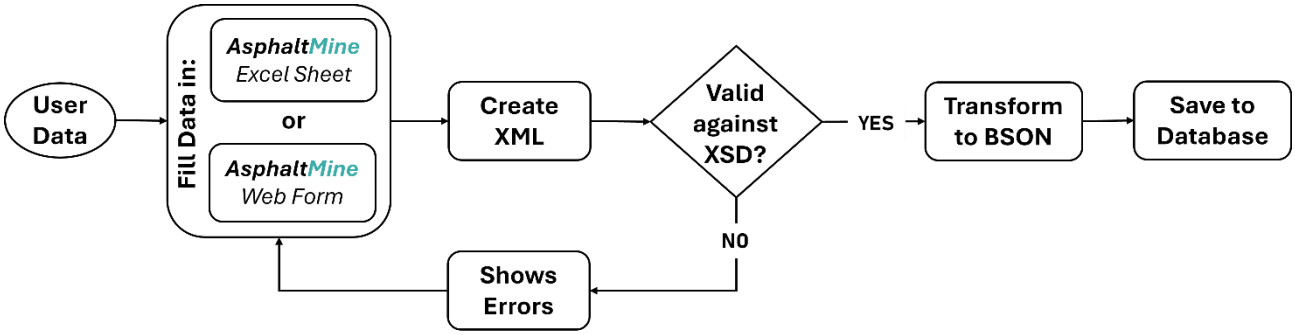


Figure 2: AsphaltMine database records saving process

Once AsphaltMine web form or excel sheet is used to fill in records and then submit to the system, the system will generate each record and initially save it as a structured XML schema-based format. XML is a flexible, text-based format widely used for storing and transporting data, which allows the data to be structured hierarchically. It ensures that the data is both human-readable and machine-readable, and providing a standardized format that can be easily integrated with other applications and platforms. A sample XML for mixture is shown in Figure 3.

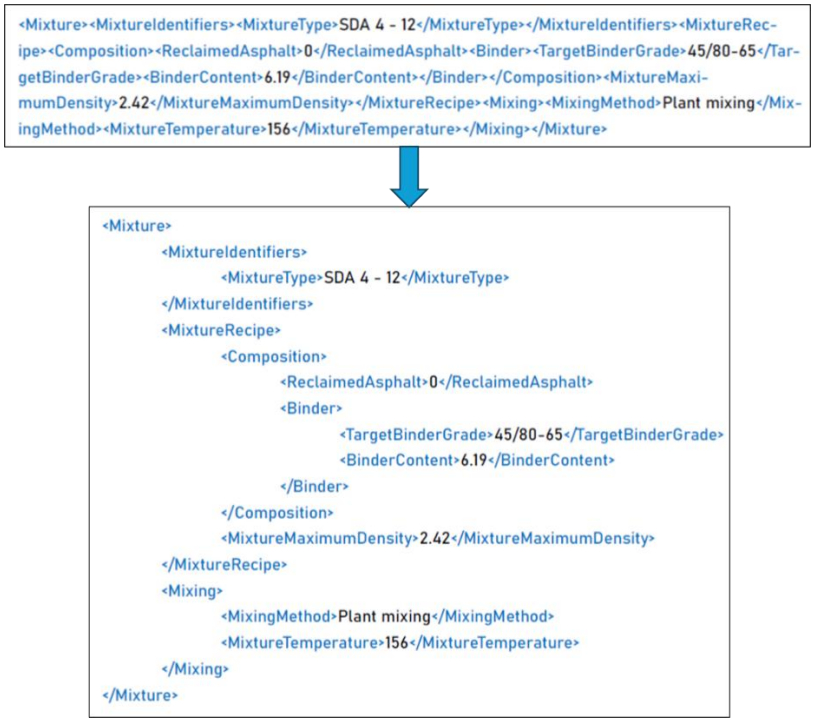


Figure 3: XML sample representation of a mixture node and its hierarchical tree structure after re-arrangement

The XML schema itself is validated and structured using an XML Schema Definition (XSD), which is a file that defines the structure, constraints, and data types of the XML documents used throughout the system. It ensures that the XML data is consistent, well-formed, and follows the agreed-upon rules for each record and test. The XSD defines various elements and attributes within the XML, including their data types (e.g. string or integer), units of measurement and enumeration values (a list of

predefined requirements for validating the data). This allows the system to effectively capture, validate, and store data in a standardized way. By linking the XSD to the XML documents, it becomes possible to validate data in real-time during the creation or editing of records within the AsphaltMine. Any violations of the defined schema, such as missing required fields, incorrect data types, or invalid units, are detected, providing users with feedback to correct errors before the data is saved. This eventually ensures that data errors or discrepancies are minimized, and the system is more reliable.

Once a record is saved, it is transferred to a MongoDB database where it is stored. MongoDB [1], a NoSQL database, stores the data in Binary JSON format called BSON, which is more efficient than XML for performance and scalability. BSON is a binary representation of JSON-like documents, designed for faster data processing and storage. Unlike XML, BSON (which is non-human readable format) is a compact format that facilitates faster access and retrieval while maintaining flexibility in schema design. The use of BSON ensures that MongoDB can scale easily and handle large volumes of data, while the initial XML format offers a structured way to define and exchange records.

By default, the data is stored on a server in Switzerland. However, AsphaltMine is a federated database system in which each user can store their own data in their own location, while still contributing the data to the online platform.

2. Records Overview

The XML records in AsphaltMine are based on the XSD files that define the permitted schema or structure of the record. In what follows, the structure of the record in AsphaltMine will be discussed.

2.1. Record Structure

The XSD of each test defines the record detailed tree structure for the test. In AsphaltMine, each single record has four main branches as shown in Figure 4.

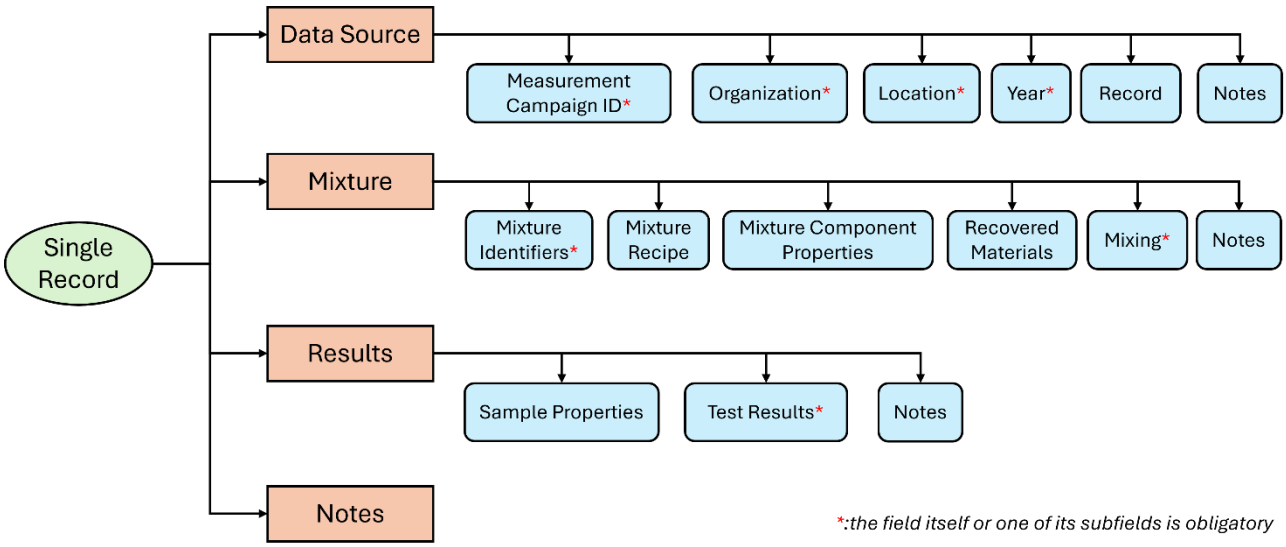


Figure 4: The tree structure of AsphaltMine database records, showing the first two levels.

The branches can be described as follows:

- **Data Source:** This branch contains source information about the record, including the organization that provided the record, the location (country), the year the test was performed, and the measurement campaign ID, which is a unique identifier for the record within the organization. Additionally, an optional Record sub-branch can be filled to specify where the record was published (e.g., journal, conference, report, or other).

- **Mixture:** This branch contains all the mixture information. It is the largest branch made of several subbranches. The mixture branch direct child and grandchild branches are shown in Figure 5. The mixture child branches are as follows:
  - **Mixture Identifiers:** This branch contains the optional Mixture ID, which is the local identifier of the mixture, and the Mixture Type (e.g., AC 22).
  - **Mixture Recipe:** This branch contains two optional subbranches: (1) the composition subbranch, which includes several child elements describing percentage by content of each material in the mixture, as well as the total aggregates distribution, and (2) the mixture maximum density.
  - **Mixture Component Properties:** This branch includes information about each material within the mixture, such as virgin aggregates and virgin binder, as shown in Figure 5. Each material contains several predefined test results that can be filled, along with an Additional Property parameter that allows the user to add test methods and results not included in the predefined structure. This parameter can be used an unlimited number of times for each material. Similarly, the Additive branch can also be used an unlimited number of times (e.g., to include multiple additives). If needed, other materials that do not fit within the predefined categories can be added by the user.
  - **Recovered Materials:** This branch includes information about the materials recovered from the asphalt mixture through binder extraction and recovery, specifically the aggregates and the binder. Both have multiple predefined parameters to be filled in and an Additional Property parameter.
  - **Mixing:** This branch contains information about the mixing process, such as the method used (e.g., plant mixing), the type of mixer, temperature, and duration of mixing. It also allows to define other mixing parameters an unlimited number of times.
  - **Notes:** Contains optional notes about the mixture.

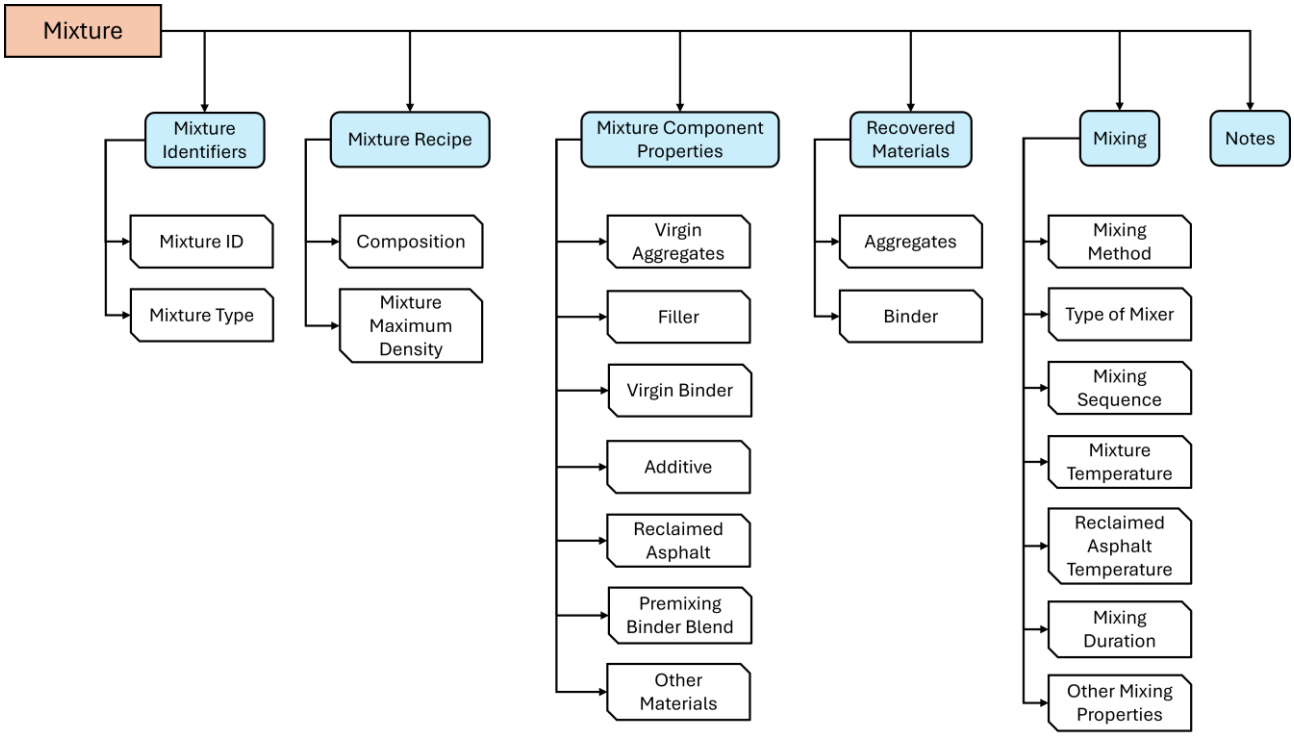


Figure 5: The tree structure of the Mixture branch of the records, showing the first two sub levels

- **Results:** This branch contains information about the test performed on the mixture, organized in two main subbranches. The Sample Properties branch describes sample preparation

properties (e.g., information about the compaction process), and the Test Results branch that contains parameters of the corresponding test (e.g., Test Temperature). The latter contains several mandatory fields to fill for each test (e.g., Test Temperature in Rutting Test), and the bulk density and air voids are always required parameters within this branch for all tests.

- **Notes:** This branch allows adding optional general notes about the saved record, and it has no subbranches.

As an example, we provide a closer look at one of the dense subbranches, the Recovered Materials subbranch, within the Mixture branch. The structure is shown in Figure 6. One can see several pre-defined parameters that can be filled with their units defined in the XSD and shown in both the Web form and the AsphaltMine excel file. In addition, the standard number used for each parameter is clearly stated in both methods. One can see the aforementioned 'Additional Properties' parameter in the BTSV branch and in the binder branch itself. This parameter, as discussed above, can be used multiple times for each case. In addition, one can see that Multiple Stress Creep and Recovery Test (MSCRT) branch has a nonstandard stress branch that allows the user to add stress not defined by its standard [2], and to fill multiple other parameters for this user defined stress.

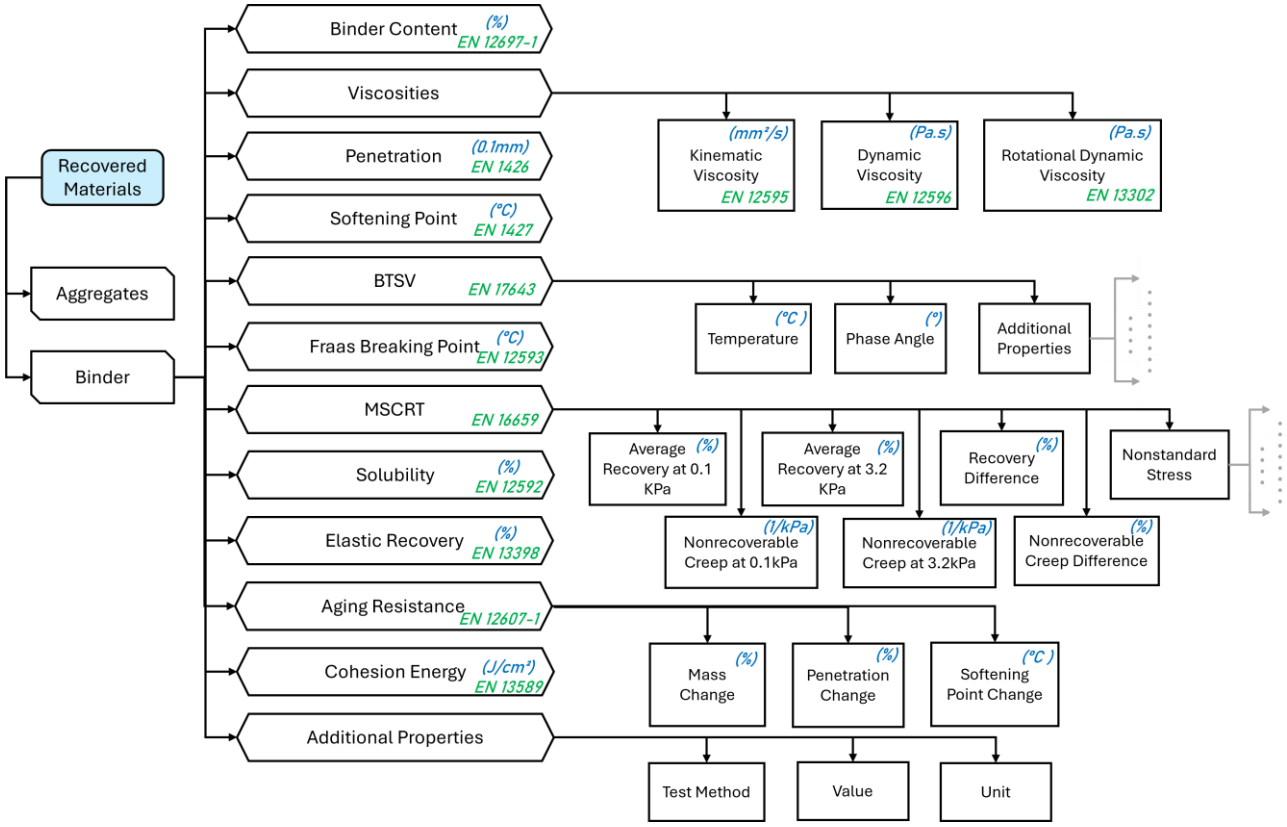


Figure 6: The tree structure of the recovered binder subbranch of the records, showing their units and the corresponding European standards

3. General User functionalities

AsphaltMine platform provides users with several functionalities, which are as follows:

- **Uploading Records:** Users can upload records.
- **Editing Records:** Users can edit records within the web form. Only the owner of the record (user that uploaded it) and any users that have been assigned "write" access can edit the record. Bulk Editing will also be possible through the predefined excel sheet in the future.

- **Downloading Records in Excel:** Users can download records in MS Excel format. This applies to records that the users own, that are public, or that the user has been given right to access. The download buttons are placed inside each workspace page and in the user's own records page ("My Data" page). Downloading is an important feature, as it allows merging records from different sources in a unified format for use in different purposes, including training data-driven performance prediction models.
- **Deleting Records:** Only the owner of the record can delete it.
- **Reassigning ownership:** Users can give ownership of their records to any other user.
- **Attaching Files:** Users can attach files to an already existing record at any time. This includes adding an image, a graph, a special sheet, pdf or another information based on the user's preference.
- **Visualization:** Users can visualize database insights, including parameter distribution and custom data plots.

References

[1] MongoDB 2025. https://www.mongodb.com/.

[2] EN 16659:2015 - Bitumen and Bituminous Binders - Multiple Stress Creep and Recovery Test (MSCRT). iTeh Standards n.d. https://standards.iteh.ai/catalog/standards/cen/b25fd0ad-65be-4590-94dd-a674cbfd3535/en-16659-2015.