Working package 1 specifications

First version of consortium's platform

"Framework contract for the provision of forest data and services in support to the European Forest Data Centre"
Reference: 2007/ S 194-235358 of 09/10/2007

Contents

1 The Framework context

The objective of the framework contract is to broaden and develop the knowledge base of the European Forest Data Centre (EFDAC) hosted by the Joint Research Centre (JRC) of the European Commission which has been established to supply European Union decision-makers with processed, quality checked and timely policy relevant forest data and information within the EU and territories where EU policies are operating.

The EFDAC will be developed and implemented as the single and central point for forest information at European level in support to relevant EU policies and as the basis of the European Forest Monitoring System proposed in the EU Forest Action Plan. The implementation of EFDAC will contribute to enhancing data harmonization and to streamlining data collection and reporting to international commitments such as the Ministerial Conference of the Protection of Forest in Europe (MCPFE), the FAO Global Forest Resources Assessment (GFRA) and the UN Convention on Biological Diversity (CBD). EFDAC will be built on the basis of the information systems currently existing or under development and in compliance with the guidelines of the Infrastructure for Spatial Information in Europe (INSPIRE). In particular, these systems are the European Forest Fire Information System (EFFIS), the Forest Focus Data Platform, and the European Forest Information and Communication Platform (EFICP). New methods and tools developed for forest and natural hazards monitoring (forest fires, storms, etc) will decisively contribute to the further development and implementation of the Global Monitoring for Environment and Security (GMES) initiative.

Thus, the framework contract will include the provision of the following services or data at EU level:

- **technical assessments** such as reports, models, reviews, etc.
- **data** in the form of either statistics from forest inventory data (e.g. spatially aggregated plot data to provide harmonized estimates of selected indicators/attributes for given areas) or selected plot attributes derived from plot observation/measurements (e.g., forest type, growing stock), to be used as validation or training fields for map production or modeling.

This service provision needs to be organized around a dedicated platform.

2 The Consortium's platform

2.1 Consortium's platform overview

In order to provide the required services, a platform called "consortium's platform" needs to be built. This action represents the first step of this framework contract. This platform will be useful for all consortium's partners to upload required raw data in response to JRC request, harmonize them at an European level and transform them into aggregated data, spatial representation, maps, etc.
In conclusion, this platform will become the central working area for the consortium's members and will act as an intermediary actor between the data provider and the EFICP.

Under the framework contract, the first two working packages concern the building of consortium's platform.
The first working package corresponds to the building of the consortium's platform to perform raw data upload, storage and queries based on a central metadata base.
The second working package corresponds to an enrichment to run specific processes or perform treatments such as harmonization, aggregation, spatial representation, dissemination to EFICP.
The picture below shows the main modules by distinguishing WP1(plain) and WP2 (Hatched).

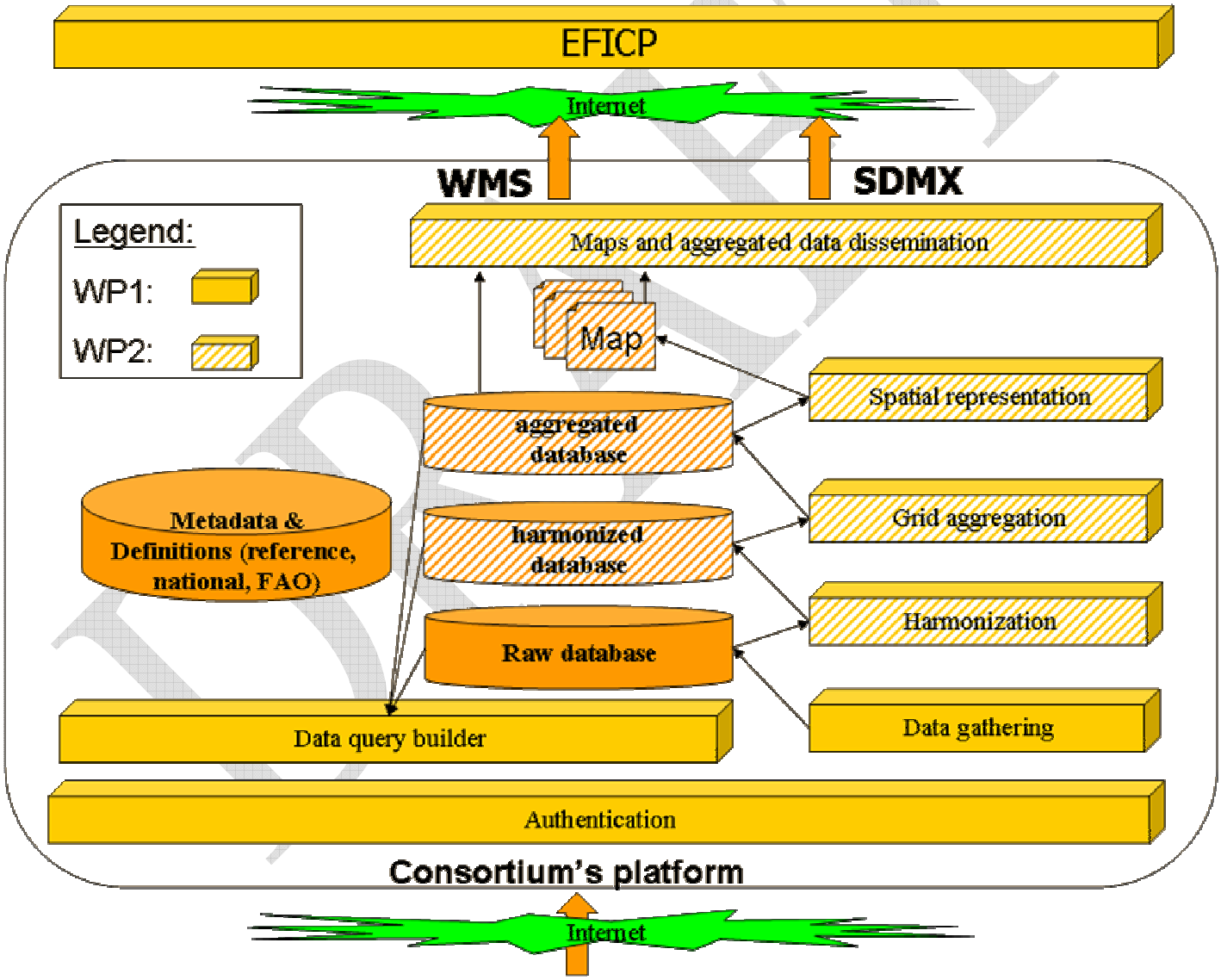


Figure 1: Consortium's platform modules

These specifications only concern the working package 1. Some information concerning the working package 2 are given in this document in order to give an overview of the whole system. In that case, the limitation to WP2 is given.

2.2 Way of working

Taking into account that this platform will be used by all the consortium members, it was necessary to have a feedback from all the framework partners. General rules had been proposed to the consortium members in December 2008. These rules deal with format used to upload file, plot file, managing plot uploads, managing reference year, etc. After a few remarks regarding stratification frames, grid cell with country boundaries, reference year, etc. the main rules of the consortium can be considered as validated. This document is based on these main rules.

2.3 Description of consortium's platform infrastructure

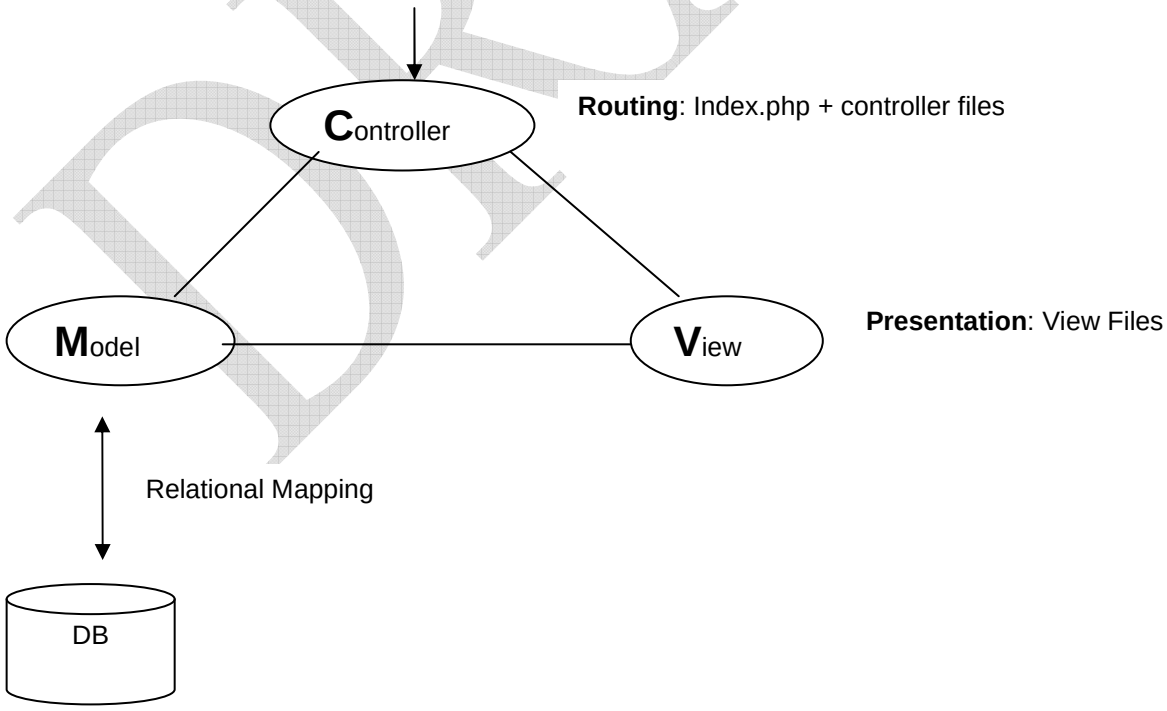
The consortium's platform is mainly developed by the French National Forest Inventory (FNFI) and hosted by the FNFI.
In order to reduce the cost of this platform, numerous modules come from other applications developed by the FNFI.
For each module developed by the FNFI under specific contract, the condition of use will be provided.
All the core or main components are based on Open source software.

2.3.1 Software components

The consortium platform will be based on core components used by the FNFI for more than 5 years.
The server Operating System is **Free BSD** (http://www.freebsd.org).
The web server application is based on **Apache** (www.**apache**.org/) and **Tomcat** (**http://tomcat.apache.org/**) .
The application is developed using **PHP and Java languages** in order to be independent from the operating system.
PHP is a very mature and powerful language. PHP Web servers are robust and can be developed in a very short time.
JAVA is useful for the data integration module and any special treatment such as harmonization process. Java is also used to establish an interface between the application and **R** (see below).

The web site will be built using a Model-View-Controller ("MVC") architecture in order to have a clean code and facilitate its maintenance.

A PHP framework will be used to ensure the good respect of the MVC design pattern.



Mapserver (http://mapserver.gis.umn.edu/) is used to provide OGS services such as Web Map Service, Web Feature Service
BIRT(http://www.eclipse.org/birt/phoenix/intro/) is used as a JAVA service to construct reports based on templates.

The **R Project** (http://www.r-project.org/) is useful for Statistical Computing such as grid aggregation, spatial representation, etc.

The RDBMS is **PostgreSQL** with its spatial cartridge **(PostGIS)**. Thus, it is possible to manage geographic types (point, line, polygons, …).

2.3.2 Web browser compatibility

The web site will be tested with the following browsers:

- Internet explorer Version >=7
- FireFox Version >=2

Javascript must be enabled on the client browser to allow user authentication.

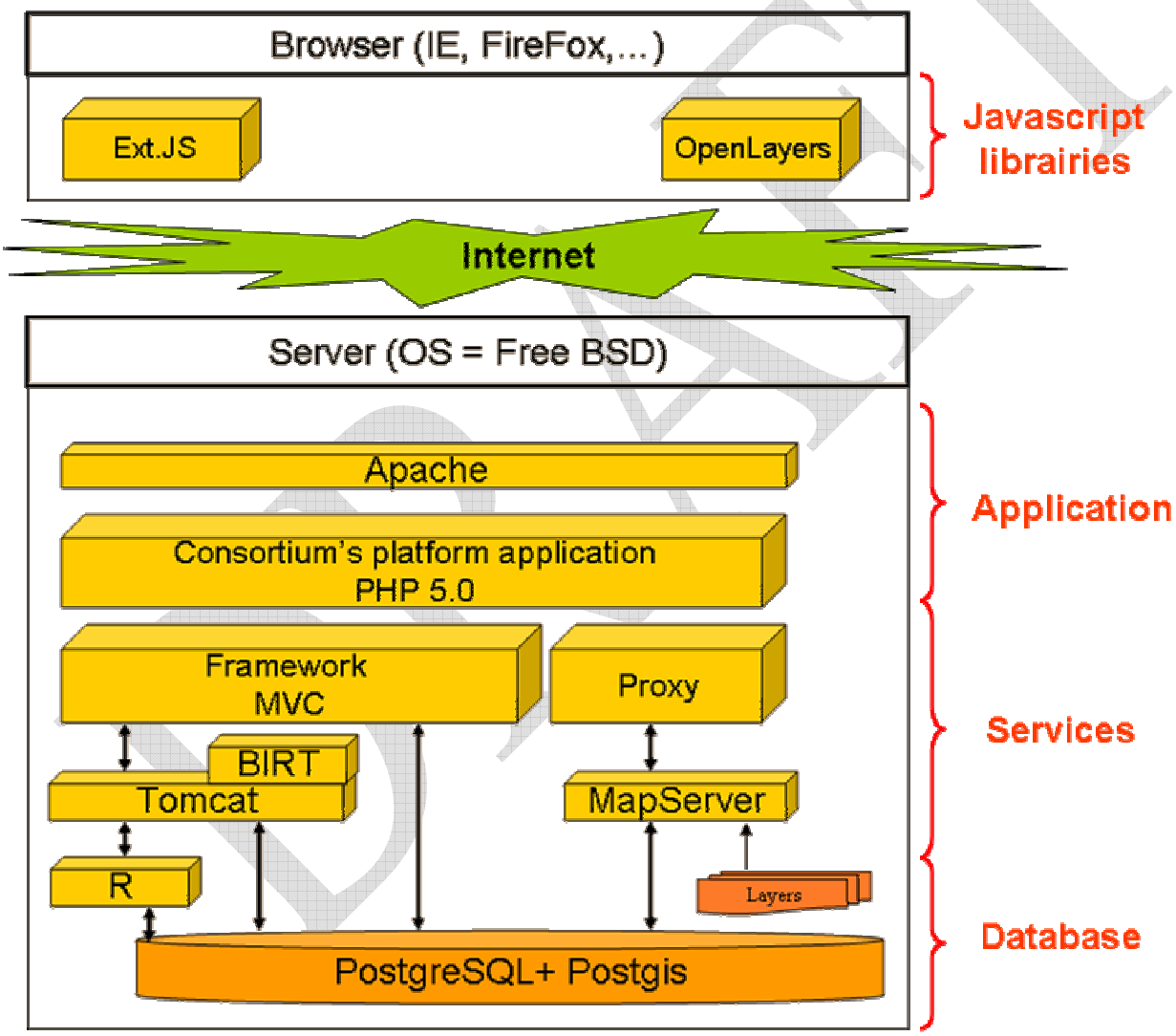


Figure 2: Software architecture

BIRT and R will be installed only within the WP2.

2.3.3 Physical Infrastructure

2.3.3.1 Hardware and Operating system

The consortium's platform is hosted at the premises of the FNFI.

The main server characteristics are listed below (minimum requirements):

- Processor: 2 x Quad Core Intel® Xeon®, 2x6Mo Cache, 2,66GHz or equivalent
- Memory: 8 x 2GB
- RAID 1 with 2 Hard Drive of 73 Go for the OS
- RAID 5 with 6 Hard Drive of 146 Go for the Data
- Redundant Power Supply

The Operating System is Free BSD.
The FreeBSD provide a jail mechanism as implementation of operating system-level virtualization. Thus it is possible to partition a FreeBSD-based computer system into several independent mini-systems called jails.
One jail is used for developing the system.
Another jail hosts the consortium's platform.

In case of hardware failure, the FNFI will transfer the consortium's platform to another server that belongs to FNFI waiting for the hardware service provider intervention.

The server will be provided by the JRC.

2.3.3.2 Internet connection

The French National Forest Inventory has a **symmetrical** access to Internet with a band width up to 4 **MByte/s**. This access is shared by the FNFI users (~50% of the band width) and not dedicated to the consortium's platform. If the connections became too important, it would then be necessary to plan another solution.

2.3.3.3 Security and network implementation

The NFI uses a DMZ to isolate the external network (Internet) from directly referencing the internal Network. Two redundant firewalls based on OpenBSD operating system manage the security rules between the DMZ and the Internet accesses.
The consortium's platform will be protected by this infrastructure.

3 Module description : Working package 1

From a **data provider** point of view, the purpose of the consortium's platform is to provide a way to upload data, to validate the data previously transferred and to carry out queries on its own data.
From a **permanent consortium's member** point of view, the purpose of the consortium's platform is to use and manage harmonized data in response to a service required by the JRC.
From the **JRC** point of view, the purpose of the consortium's platform is to offer to the consortium the possibility to manage with efficiency the data coming from data providers and to access the aggregated data.
The consortium's platform limited to **aggregated data** can be considered as a **part of the EFDAC**.

The Working package 1 corresponds to the basement of the consortium's platform.
The main modules of the consortium's platform are listed below.

The **authentication module** controls all the accesses to the application and data.
The **data gathering module** offers to data providers a friendly way to upload data files containing raw data. In response to these uploads, the data providers receive a PDF file containing errors to correct or warnings.
The **data query builder module** proposes to the authorized user an easy way to carry out queries. In the WP1, the queries are limited to the database containing raw data.

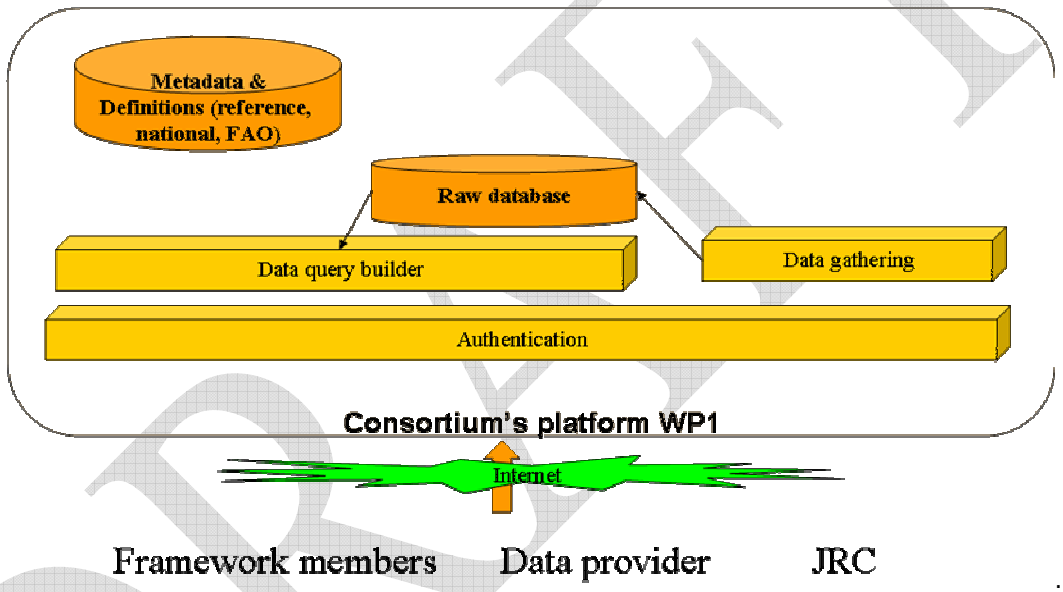


Figure 3: WP1: Consortium's platform

3.1 Authentication module

The application is not accessible without authentication.
Depending on the user's profile, accesses to data are possible.
Data providers can access their own data without any restriction.
The permanent members of the consortium can access all the data stored in the consortium's platform.
The JRC can access only aggregated data whenever the data are validated by the data provider.

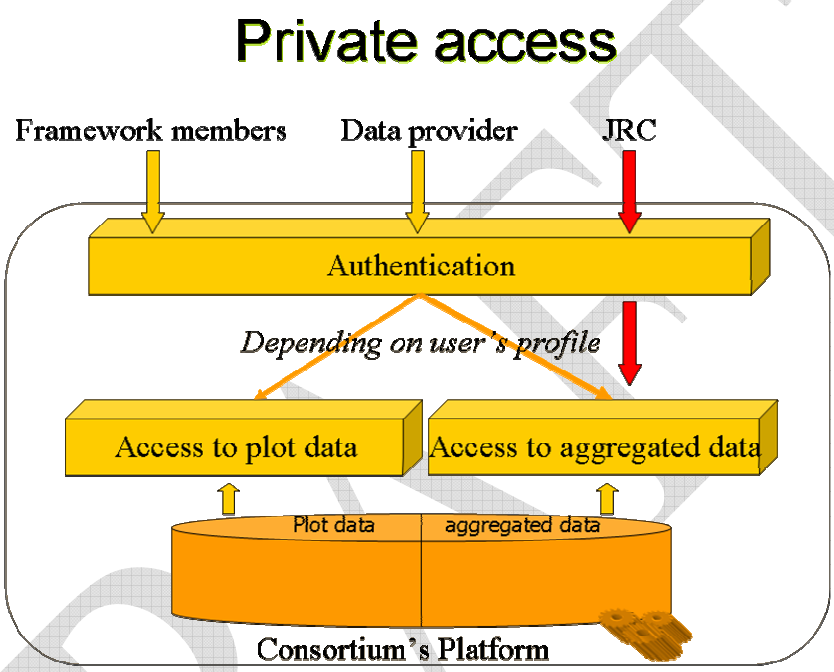


Figure 4: Authentication module

3.1.1 User profile description

- **administrator** : Access to all the modules plus user management.
- **data provider :** Access to the data gathering module to submit the country's data. Access to the data querying module limited to the country data.
- **framework members** : Access to the data querying module with no limitation.
- **JRC** : Access to the data querying module with no limitation about country but limitation about data (raw data won't be reachable by the JRC)

3.1.2 Authentication mechanism

We will use a challenge-response authentication mechanism.

The user passwords will be stored in their encoded form in the database.
When showing the login screen, the server will generate a random value (the challenge).
The login screen will ask the user for a login and a password, and will compute locally a response using the challenge and the password). The server will do the same calculation and will compare the results. The password is never sent over the network.

3.2 Data gathering module

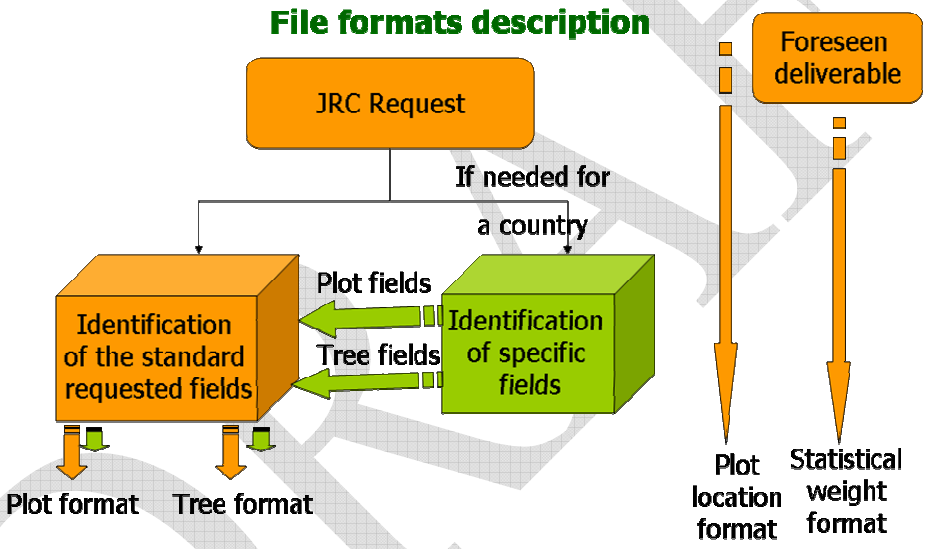
This task represents the first part of the data processing. Its purpose is to ensure that the information stored in the database is “error free” and can be used in further analyses such as harmonization, aggregation,…

The files to upload will be in the CSV (Comma Separated Value) format and will be documented in detail in some separated “file format specification” documents. Whenever it is possible, these specifications are provided to be used by all the data providers. In case of inconsistencies between national data and requested data, the specifications will be adapted.

The module is composed of a web interface and an integration service.
The integration service consists of a java application with servlets responding to the file upload HTTP requests.

The web interface allows the user to :

- Submit some plot definitions (locations).
- Submit plot statistical weights (for statistical analysis).
- Submit data (plot and tree).



3.2.1 Plot location

For each country, the location of the plots used in the different surveys must be defined.

The definition of a plot consists of an identifier for the plot in the country (given to the plot by the data provider) and the geographic position of the plot.

File Format :
The file will contain the following mandatory fields: PLOT_CODE, LATITUDE, LONGITUDE, COMMENT in English

The plots, defined like this, can be considered as “permanents”. Their position cannot evolve in time and a country can re-use a plot for different surveys. A plot can never be deleted from the database once some data has been attached.

A country can add some new plot locations by submitting a new PLOT_LOCATION.CSV file. If the new file contains some plot identifiers already in use it will be rejected.

A web page allows the user to upload a file containing plot locations. Another web page will show the user the files already submitted (with the possibility to cancel a submission) and the total number of plot locations.

Each submission is identified with a unique submission identifier.

The plot code refers to the usual code plot of the country. This code is given in all the other files to identify each plot. An internal code plot (SQL sequence) is used to identity the plot in the consortium's platform databases.

3.2.2 Plot statistical weight

For each country, the plots used in the different surveys must have a statistical weight, according to the stratification frame used (see WP2 for more details).
Different stratification frames are necessary. We need a stratification based on **NUTS** code in order to provide **statistical results** linked with administrative references. We need a stratification based on **grid** to transform grid cells into a **spatial graphical representation**. Thus, the system allows the use of different stratification frames. These frames have to be defined in the database in order to be referred to by the upload file (i.e. STRATIFICATION_CODE).
The statistical weight is useful during the aggregation phase to calculate the contribution of each plot.

File Format :
The file contains the fields: PLOT_CODE, REF_YEAR, STRATIFICATION_CODE, STATISTICAL_WEIGHT, COMMENT in English.

A web page allows the user to upload a file containing statistical weights. Another web page shows the user the files already submitted (with the possibility to cancel a submission), the total number of active plot locations and the number of plot having a weight for each stratification frame.

Each submission is identified with a unique submission identifier.

3.2.3 Plot and Tree Data

Once the plot locations are defined, the user can upload some data related to the plots. These data correspond to data linked to plots or data linked to trees observed on plots.

Before uploading some files, the user has to select the JRC request he wishes to answer.
This JRC request defines file formats: a list of mandatory data for plots and trees.
These formats are defined thanks to the metadata FORMAT table.
The COUNTRY_CODE provided by the login gives also potential complementary data to upload. These complementary data must exist in the metadata database. That means that an analysis is carried out country by country in order to identify whether complementary data are necessary or not.

Plot file format :
The file contains the fields PLOT_CODE, INV_DATE, … a list of data values …, COMMENT.

Tree file format:
The file contains the fields PLOT_CODE, INV_DATE, TREE_CODE, …a list of data values …, COMMENT.

A submission is composed of two files (plots and trees). Each submission is identified with a unique submission identifier calculated after the submission.
Before submitting the files, the country has to fill information about the reference year or reference period of the data submitted. This reference year or period concerns all the plots submitted.

A submission must pass the checks and must be validated by the user before being usable by the other modules. A submission that is not yet validated or used can be deleted by the user (in this case, the plot data is removed from the table).

A country can carry out more than one submission for the same JRC_REQUEST. If a submission contains a plot that has been already submitted, the submission is rejected.

The web page shows to the user some statistics about the last submissions.

3.2.4 Checks

The main part of the validation checks is based on the **Metadata database**.

The validation checks are composed of two types of checks:

- The **compliance checks** verify the accordance of a flat file with the formal aspects specified in the format file specifications and verify if the data are in accordance with the Metadata database (data type, data values, ranges, etc.). If the data are not fulfilling the compliance checks, then they are marked as errors and the whole submission is rejected.
- The **conformity checks** verify if the data are consistent with other data according to a set of rules. These checks are only applied if the compliance checks are passed. If the data are not fulfilling the conformity checks, then they are marked as warnings. The conformity rules are stored in the metadata database in the "Check" table.

Unsuccessful checks	Consequence
Compliance checks	The whole file is rejected
Conformity checks	Values are marked as warning

An example of a conformity check can be the following:

- Check if a value seems out of range. A pH > 14 is a compliance error and is rejected, but a pH > 11 for a forest soil can be defined as suspicious with the help of a conformity rule.

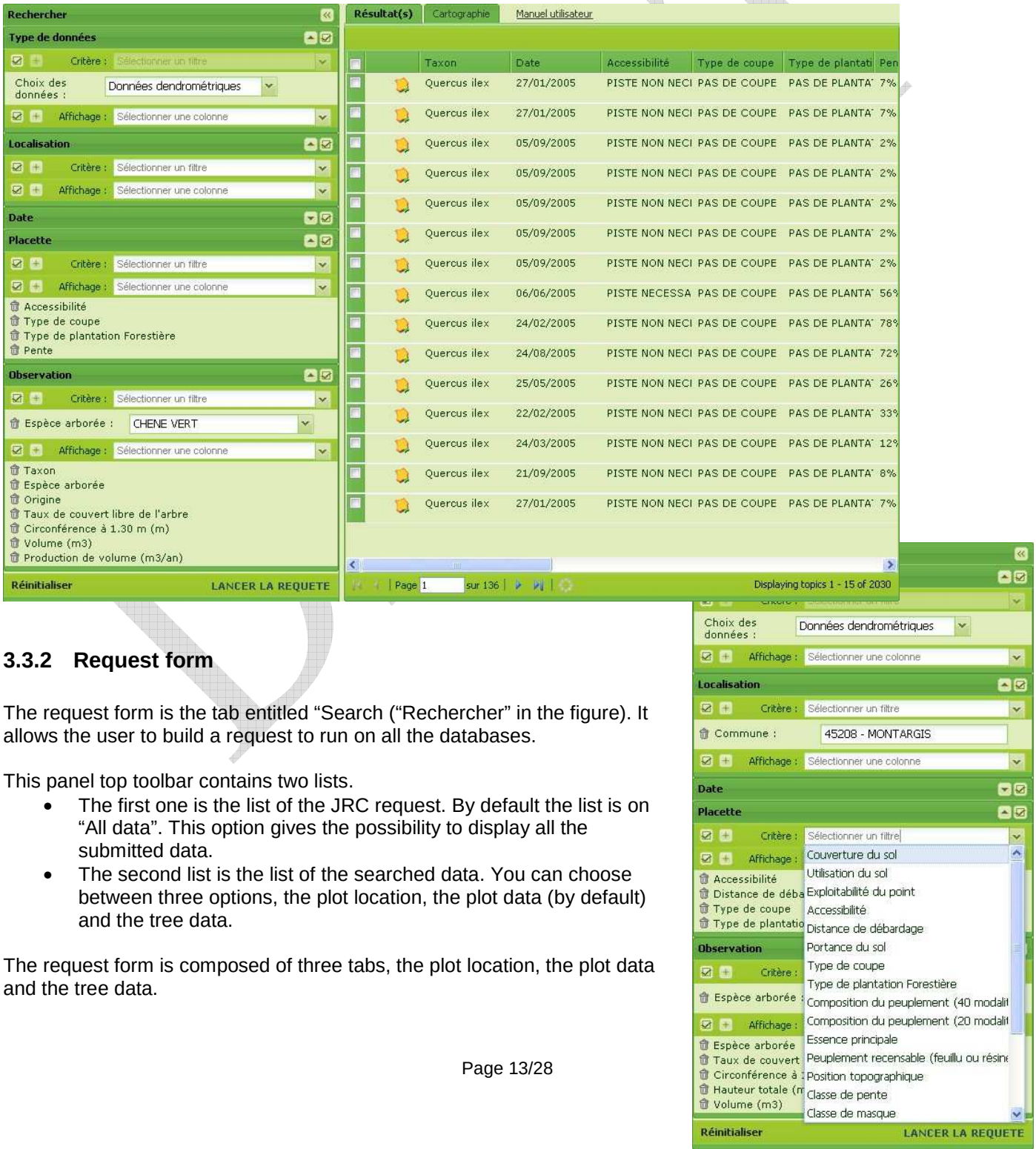
A PDF report is generated at the end of the submission process.

3.3 Data querying module

This module is based on a software developed by FNFI.
The purpose of this module is to offer the possibility to the final user to build a SQL query without knowing SQL language.
The user can carry out his own request by specifying in a user-friendly way the columns to display and the SQL "where" clause.
The data querying module allows the user to query all the databases (raw database in WP1 and aggregated data in WP2).

3.3.1 Web Interface

The query forms are generated dynamically from the metadata database (a "string" field will allow for a search using a input field while a "code" field will show a drop-down list box as search criteria).



3.3.2 Request form

The request form is the tab entitled "Search ("Rechercher" in the figure). It allows the user to build a request to run on all the databases.

This panel top toolbar contains two lists.

- The first one is the list of the JRC request. By default the list is on "All data". This option gives the possibility to display all the submitted data.
- The second list is the list of the searched data. You can choose between three options, the plot location, the plot data (by default) and the tree data.

The request form is composed of three tabs, the plot location, the plot data and the tree data.

For each tab it's possible to add some filters and columns to display via two select lists called "filter" and "columns".

When the JRC request list is changed, all the fields depending on a JRC request in the form are automatically erased and all list "Criteria" are updated to match the new JRC request selected.

Each tab can be disabled using the check box on the right of its title. It can also be folded for more clarity, by clicking on the title, or by clicking on the small arrow in the same area. It is possible to disable only the criterion or column of each tab, respectively deselecting their checkbox. With the buttons plus it's possible to add with one-click, the entire content of the list being at their right. When a field is added, it can be removed at any time by clicking on the little trashbin displayed at its left.

Some columns are proposed by default. It's possible to remove them. To start the request and see the results, the user must press the button at the bottom right of the request form, named "launch the request". At any time the user can reset the query by clicking on the "Reset" button located at the bottom left of the module.

3.3.3 Result array

The result table contains the columns requested in the request form. It contains a tool column on the left. For each line, if the action is possible, a magnifying glass and a map are displayed. The magnifier provides access to details of the observation and the map can display the location of the observation on the map panel.

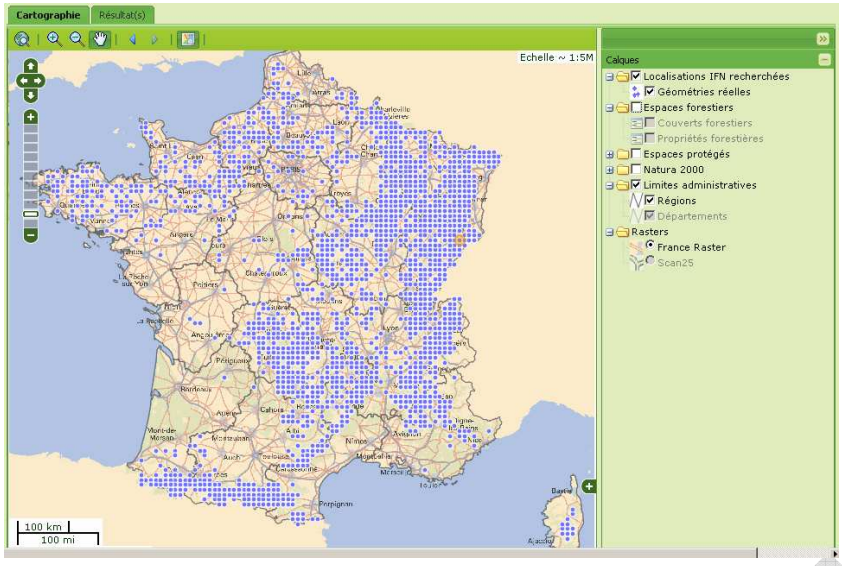
The bottom bar of the table allows the user to browse through the pages of the table. The simple arrow allows to scroll through the results page by page. The arrows with a bar allow to reach the first page or the last. It is also possible to directly enter the page you want into the box next to "Page", then to press on "Enter" to go to the page. The positioning and the total number of data are displayed in the bottom right of the table.

It is also possible to perform different actions by clicking on the arrow that appears when you right over the title of each column,. Users can sort the whole table by ascending or descending. Users can also unselect columns by displaying the menu columns.

Columns can be moved. To do this, the user simply clicks and holds on the title of the column and move the cursor between two other columns before releasing the button.

3.3.4 Graphical representation

The table can be transformed into an interactive map using a web mapping module.
In this representation, each line of the table is represented by a plot in the map.
All the layers used in this module will be provided by the JRC (administrative boundaries, …).


Cartographie
Résultat(s)
Echelle ~ 1:5M
Calques
Localisations IFN recherchées
Géométries réelles
Espaces forestiers
Couverts forestiers
Propriétés forestières
Espaces protégés
Natura 2000
Limites administratives
Régions
Départements
Rasters
France Raster
Scan25
100 km
100 mi

4 Database description

The database called **Raw Database** has to store raw data compliant with metadata according to national definitions or reference definitions.

A **Metadata database** is required to provide structured information about the data managed in the Raw Database and to offer the possibility to harmonize data.
This Metadata database is filled once with code lists, range values, flat file format, methodological aspects, table description, and updatable when needed. This Metadata database allows modules to automatically process the raw data whenever it is possible.

A “check” table, stored in the metadata database will contain some rules used to validate the data during the submission, the result of the checks will be stored in a “check_error” table in the raw database.

The **Web Server Database** is used to configure the web server application (users, profiles, system parameters).

The **Harmonized database** will store the harmonized data, its structure which will be very similar to the raw database will be covered in the WP2 specifications.

4.1 Metadata database

4.1.1 Metadata advantages

Metadata mean descriptive information about an object or resource (**physical** or **numeric**).
The main advantage of a Meta database is to provide **structured data about data**.
All objects used in consortium’s platform have to be described: parameters, definitions (national, reference), flat file format, database tables description, etc.

For each **parameter**, descriptive information means: name, type, units, definition, rules, the associated code list, the possible range values, etc. All these characteristics are stored in this database.

For each **file** and **table**, field information are stored (parameter name, format, position, length, etc.)

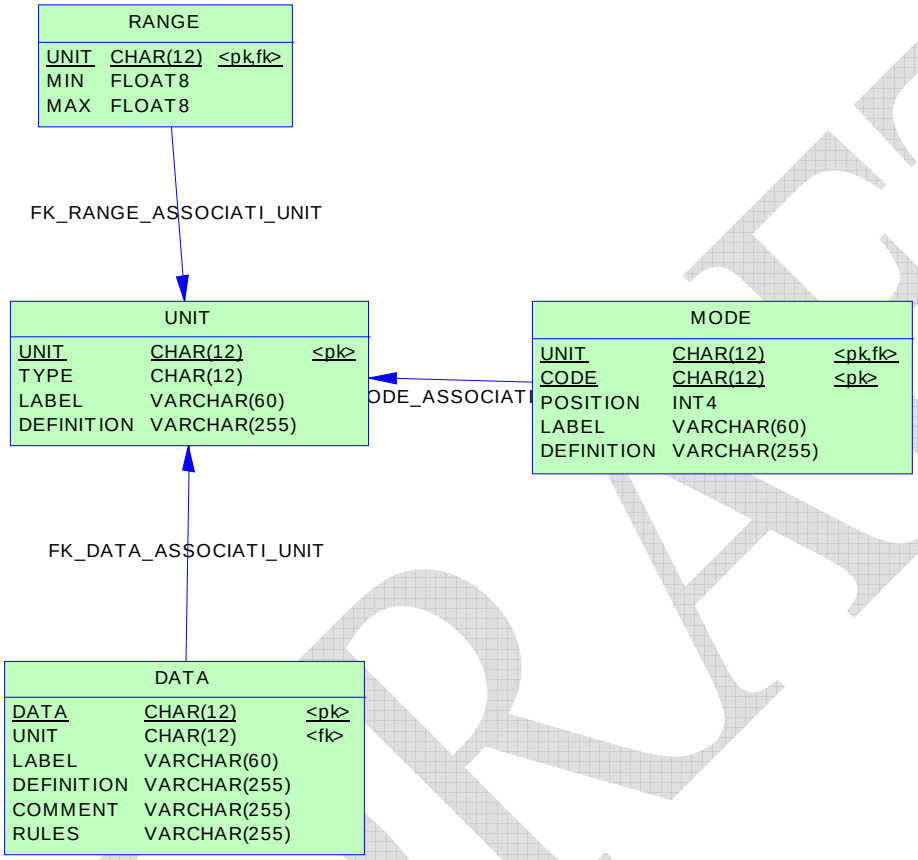
All these information can be used to perform the **automated validity check.**

As a flat file or a database table is described in the Metadata database, it is easy to parse a file and verify if all the records are compliant with their formats.

4.1.2 Metadata specification

The figure below shows a part of the physical Metadata model that will have to be adapted during the specification phase.
The physical model of the Metadata database is given in order to explain the Metadata concepts.
In the following section, automated validation checks are explained using these Metadata tables.

4.1.3 Data description (tables DATA, UNIT, MODE and RANGE)



The data description is used to model information about all parameters used in the consortium's platform and their possible values or ranges.

DATA Table:
This table describes all the data used in consortium's platform.
The mains columns are described below:

Table column	Description	Example
DATA	Data name	TREE_SPECIES
UNIT	Data reference unit	SPECIES
LABEL	Short definition that can be used in the human interface	
DEFINITION	Long definition	
COMMENT	Comment about this data (if needed)	
RULES	Information about how the data have to be noticed, measured, (if needed)	

UNIT Table:
The characteristics can be discrete or continuous. The first ones (nominal or ordinal) are expressed with a number of possible values stored in the MODE table. The second ones are described using universal measurement units. In case of continuous type, the possible range can be specified in the RANGE table.

Table column	Description	Example
UNIT	Unit name	SPECIES
TYPE	Unit type (CODE, RANGE, STRING, NUMERIC)	CODE
LABEL	Short definition that can be used in the human interface	List of tree species
DEFINITION	Long definition	

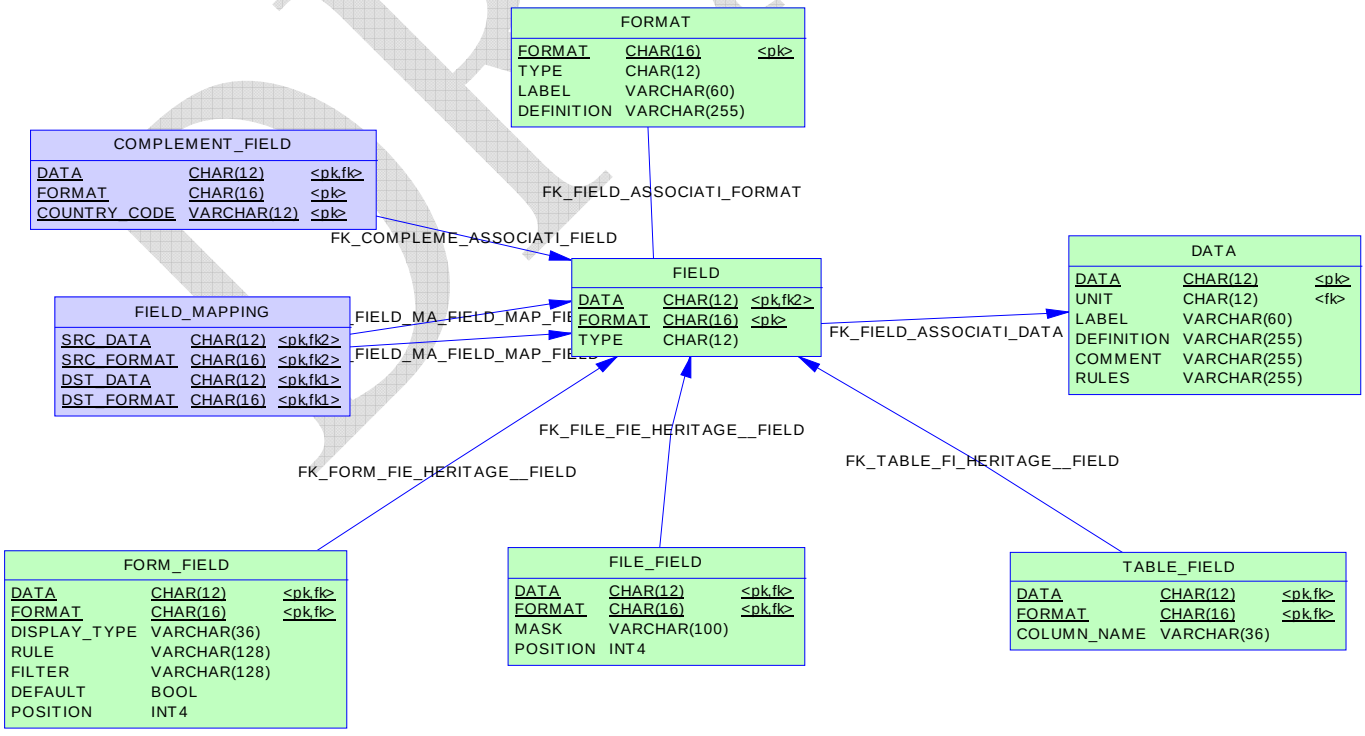
MODE Table:
This table contains a line per mode of discrete or ordinal characteristic.

Table column	Description	Example
UNIT	Unit name	TREE_SPECIES
CODE	Value	1
LABEL	Short code definition that can be used in the human interface	EVERGREEN OAK
DEFINITION	Long code definition	

RANGE Table:
This table contains a line by mode of continuous characteristic.

Table column	Description	Example
UNIT	Unit name	pH (CaCl2)
MIN	Minimum acceptable range	2
MAX	Maximum acceptable range	9

4.1.4 Format description (tables FORMAT and FIELD)
The format specification describes where and how the parameters are stored.



FORMAT Table:
This table contains one record per table or file used.

Table column	Description	Example
FORMAT	Parameter name	PLOT_DATA
TYPE	Format type (TABLE,FILE or FORM)	TABLE
LABEL	Short definition that can be used in the human interface	PLOT WHERE FIELDS OPERATIONS HAVE BEEN CARRIED OUT
DEFINITION	Long definition	If needed

FIELD Table:
This table describes a field in a file or a table.
The inherited tables FORM_FIELD, FILE_FILE and TABLE_FIELD contains specific information.

Table column	Description	Example
DATA	Parameter name	INVENTORY_DATE
FORMAT	Format name	PLOT_DATA
TYPE	Field type (STRING, INTEGER, …)	DATE

The available field types are the following:
STRING
INTEGER
FLOAT
DATE

FORM_FIELD Table:
This table adds the information used to display a field in a web page form.

Table column	Description	Example
DATA	Parameter name	INVENTORY_DATE
FORMAT	Format name (name of the form)	PLOT_FORM
DISPLAY_TYPE	Field display type (SELECTBOX, TEXT…	SELECTBOX
RULE	A rule to apply on the field to validate it	DATE_EXIST
FILTER	A filter to apply on the field	DATE
DEFAULT	Indicate if the field is displayed by default	TRUE
POSITION	The position of the field in the form	2

FILE_FIELD Table:
This table adds the information used to describe a field in a CSV file.

Table column	Description	Example
DATA	Parameter name	INVENTORY_DATE
FORMAT	Format name (name of the file type)	REQ1_PLOT_FILE
MASK	Regular expression used to validate the data	yyyy-mm-dd
POSITION	The position of the field (column) in the file	2

TABLE_FIELD Table:
This table adds the information used to describe a field in a database table.

Table column	Description	Example
DATA	Parameter name	INVENTORY_DATE
FORMAT	Format name (name of the table)	PLOT_TABLE
COLUMN_NAME	The column name	date

FIELD_MAPPING Table:
This table is used to describe the mapping between two fields. This is used to know where to put into the database a field coming from a CSV file.

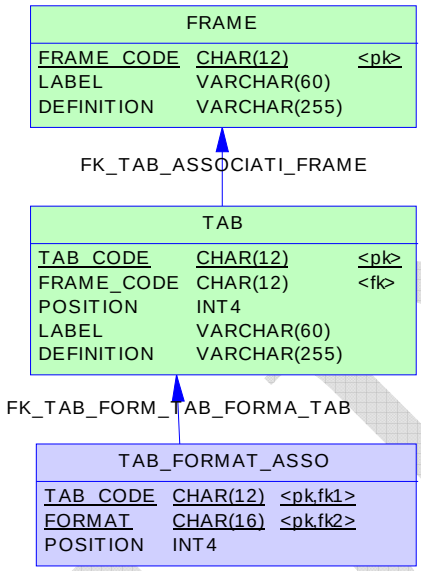
Table column	Description	Example
SRC_DATA	Source parameter name	INVENTORY_DATE
SRC_FORMAT	Source format name	REQ1_PLOT_FILE
DST_DATA	Destination parameter name	INVENTORY_DATE
DST_FORMAT	Destination format name	PLOT_TABLE

COMPLEMENT_FIELD Table:
This table is used to list the complementary field to add to a format for a given country.

Table column	Description	Example
DATA	Parameter name	IR_10
FORMAT	Format name	REQ1_PLOT_FILE
COUNTRY_CODE	The code of the country	1

4.1.5 Web site FORM description (tables FRAME, TAB and TAB_FORMAT_ASSOC)

These tables describes how the data is shown on the web site.



FRAME Table:
This table contains a line per navigator frame.

Table column	Description	Example
FRAME_CODE	The frame code	CONSULTATION
LABEL	Short code definition that can be used in the human interface	The consultation page
DEFINITION	Long code definition	

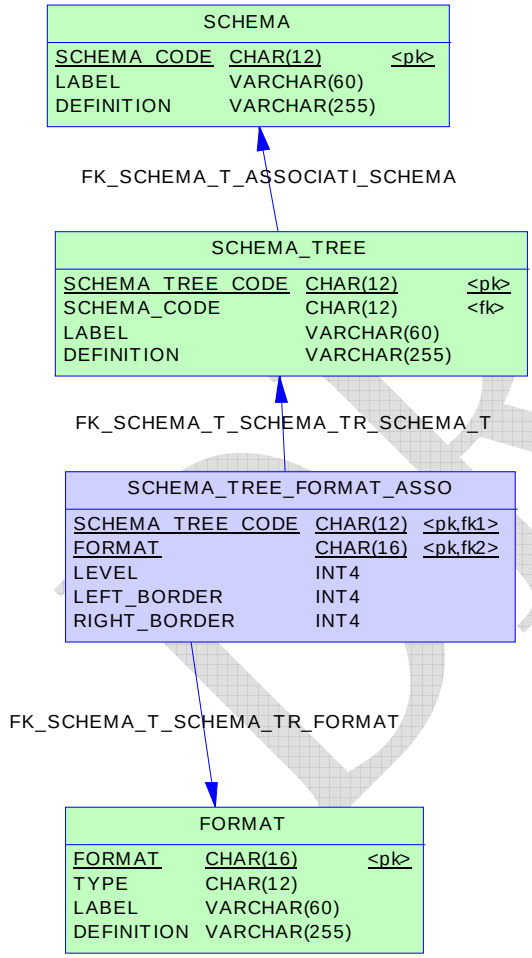
TAB Table:
This table contains a line per frame tab.

Table column	Description	Example
TAB_CODE	The tab code	CONSULTATION_PLOT
FRAME_CODE	The frame code	CONSULTATION
POSITION	The tab position in the frame	4
LABEL	Short code definition that can be used in the human interface	The plot tab
DEFINITION	Long code definition	

TAB FORMAT ASSO Table:
This table contains a line per association between a format and a tab.

Table column	Description	Example
TAB_CODE	The tab code	CONSULTATION_PLOT
FORMAT	The format (form)	PLOT_FORM
POSITION	The format (form) position in the tab	4

4.1.6 Database structure description (FORMAT, SCHEMA and SCHEMA_TREE)



The SCHEMA table lists the different database schemas (RAW, HARMONIZED, AGREGATED, …).

For each SCHEMA, a list of tables (FORMAT) and their relationship (SCHEMA_TREE) is described.

SCHEMA Table:
This table contains a line per database schema.

Table column	Description	Example
SCHEMA_CODE	The schema code	RAW_DATA
LABEL	Short code definition that can be used in the human interface	The raw data schema
DEFINITION	Long code definition	

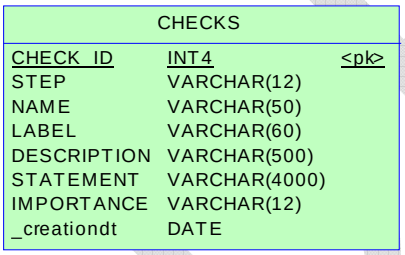
SCHEMA_TREE Table:
This table contains a line per database schema tree.

Table column	Description	Example
SCHEMA_TREE_CODE	The tree code	RAW_DATA_CONSULTATION
SCHEMA_CODE	The schema code	RAW_DATA
LABEL	Short code definition that can be used in the human interface	The raw data schema
DEFINITION	Long code definition	

SCHEMA_TREE_FORMAT_ASSO Table:
This table contains a line per database schema tree.

Table column	Description	Example
SCHEMA_TREE_CODE	The tree code	RAW_DATA_CONSULTATION
FORMAT	The format code (table)	PLOT_TABLE
LEVEL	The level of the format (table) in the tree	3
LEFT_BORDER	The left border of the node in the tree	56
RIGHT_BORDER	The right border of the node in the tree	57

4.1.7 Check rules



The **CHECK** table represents the knowledge database based on a list of check rules to apply to the data. It contains all information that is requiring an in-depth knowledge of the Forest environment at the European scale.

Table column	Description	Example
CHECK_ID	Identifier of the check	20012
STEP	Validation step (Compliance, Conformity Uniformity)	Conformity
NAME	Short name of the check	PH_PROBABLE_VALUE
LABEL	Error label associated with the check	The Ph value should be between 2 and 9.
DESCRIPTION	Explain in plain English the meaning of the control	The expected value for a soil Ph is between 2 and 9.
STATEMENT	Contains the verification rule in SQL	INSERT INTO CHECK_ERROR …
IMPORTANCE	Error or Warning	Warning

4.1.8 Spatial information

SPATIAL_LOCATION Table:
This table contains spatial location information (the boundaries of the countries).

Table column	Description	Example
UNIT	Unit name	DEPARTMENT
CODE	Value	1
GEOM	The geometry in binary	1126525411add35…

4.2 Raw database

The raw database stores data coming from data providers.

The files containing the data are uploaded during a **submission phase**.

Each submission is identified by a country code and a JRC request.

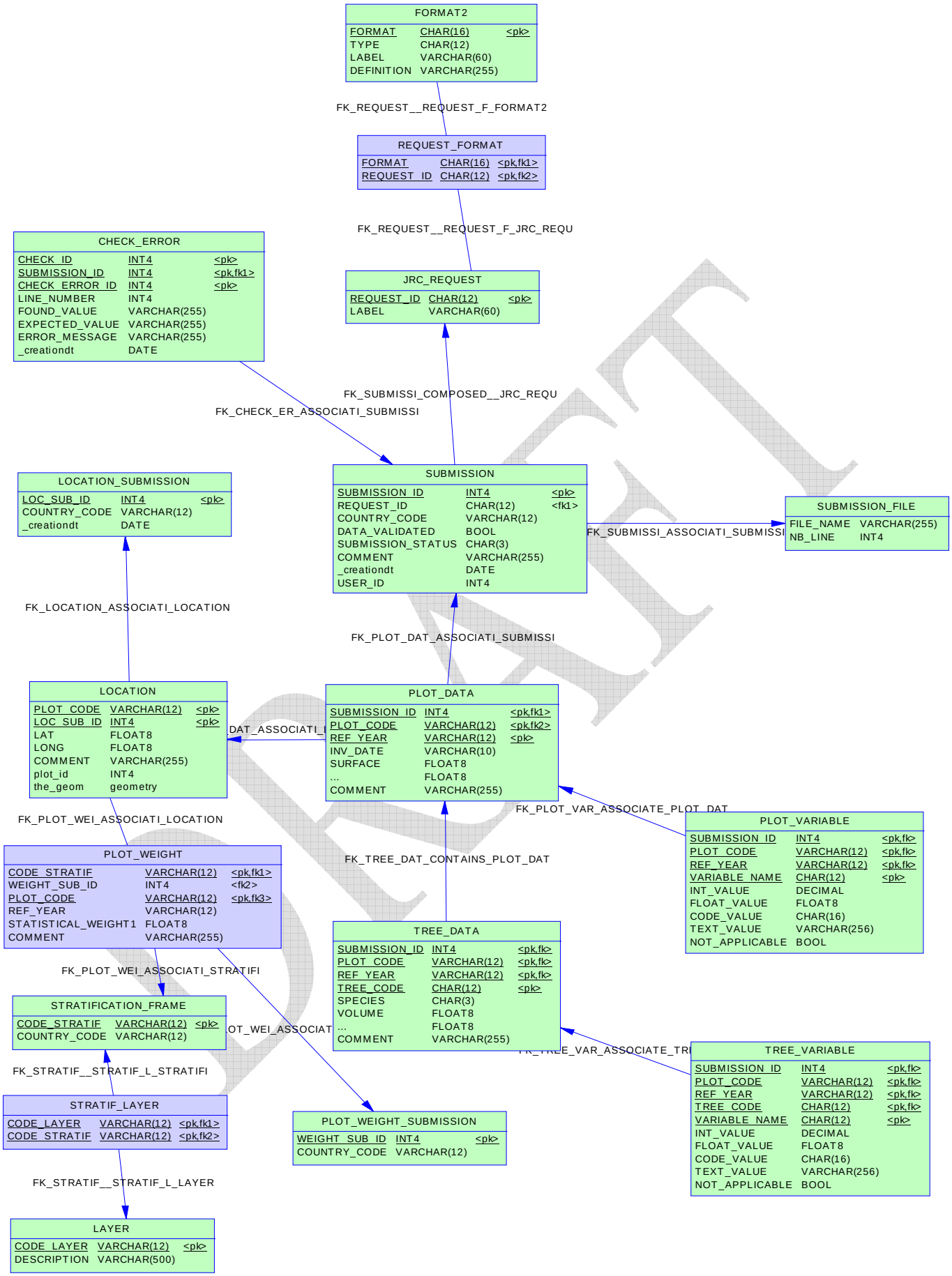
The first step in transferring data corresponds to a file containing plot coordinates and a plot identifier (PLOT_CODE,).
This file allows the management of permanent plots by reusing the same plot identifier.

The second step corresponds to a data transfer of:

- plot variables referencing the PLOT_CODE,
- tree variables referencing the PLOT_CODE, and TREE_CODE.

Some variables such as the date of assessment are mandatory.

For the statistical data processing, another step will allow uploading some information on the statistical weight of each plot.



Raw database principles

JRC_REQUEST Table:
This table describes a request from the JRC (a list of data that will be useful to answer a JRC request).

A JRC request is linked to a list of FORMAT (the list of CSV files asked).

SUBMISSION Table:
This table will contain the status of all the submissions done by the countries. A submission is always linked to a JRC Request.

Each sending of a CSV file creates a new submission entry.
When a submission is "reset" by the user, the data is removed for the TREE and PLOT tables and the entry in the SUBMISSION table is set to a status "CANCELLED".

SUBMISSION_FILE Table:
This table is used to keep the path of the directory where are stored the files relative to the submission.

LOCATION Table:
This table contains the geographical position of the plots.

LOCATION_SUBMISSION Table:
The location data is not linked to a JRC Request, the plots are permanent and can be reused. We store the history of the location submissions in this separate table.

PLOT Table:
This table contains data about the forest plots.

The list of fields of this table will be updated when new mandatory data will be asked by the JRC (a new JRC Request will potentially require to alter this table to add some columns).

An entry in this table is related to a submission. It can be deleted if the submission is cancelled.

For a given plot location and a given reference year, the table can potentially contain more than one line (for example one line with the volume and another line, from another submission, with the area). If there is more than one entry, we should have only one that is not null for a given data.

TREE Table:
This table contains data related to a tree.

The list of fields of this table will be updated when new mandatory data will be asked by the JRC (a new JRC Request will potentially require to alter this table to add some columns).

PLOT_VARIABLE Table:
This table will be used to store complementary data about the plots. Each line of this table will contain one plot information (line based storage instead of column based). The complementary data will be described in the metadata database and can be linked to a specific country.

The information contained in the PLOT_VARIABLE and TREE_VARIABLE tables will be used to calculate harmonized data.

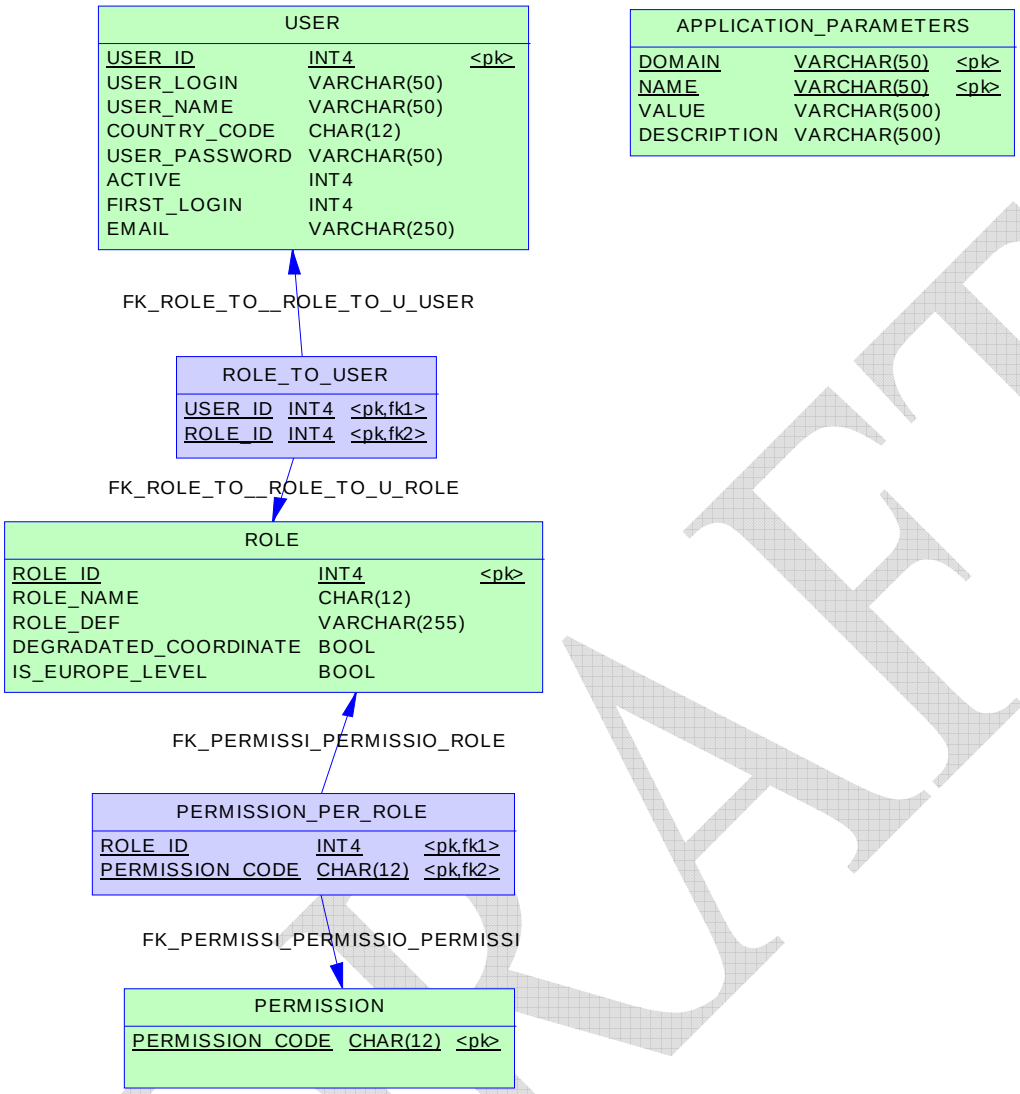
TREE_VARIABLE Table:
See PLOT_VARIABLE.

CHECK_ERROR Table:
This table will contain the results of the checks. It is populated by the execution of the SQL statements contained in the CHECK table or by the "hardcoded" checks (data compliance).

PLOT_WEIGHT Table:
This table will contain the statistical weight (area) of the plot in a given stratification frame. This will be used during WP2.

4.3 Web Server database

This database will contain the tables needed to run and manage the web site.



USER Table:
This table will contain the list of users of the web site with their login and password.

ROLE Table:
This table will list the different roles of the users.

PERMISSION Table:
This table will list the different permissions (actions or modules of the web site). This will allow to affect some access rights to the roles.

APPLICATION_PARAMETERS Table:
This table will contain some general parameters of the application (like the server address, the STMP mail server address, …). It will allow us to manage a production server separated from the test server with the same code.

5 Use cases

5.1 How to define a data?

The definition of a data describes the data (definition, unit, etc). Thus, this definition is also useful in any treatment such as compliance checks when importing the data, harmonization, etc...
Creation of a data step by step:

- Create a new entry into the DATA table to define the new parameter (for example: SURFACE)
- If needed, create a new UNIT to be linked with the data (for example: M²).
- If needed, define the range or the list of codes of the unit.

5.2 How to define a field format for storage?

The definition of a field gives information on where the data is stored in the database.
Creation of a field step by step:

- Add a new entry into the FIELD table to specify where the data will be stored in the database.
- Complete the TABLE_FIELD entry.

5.3 How to define a field format for presentation?

The purpose of the field format is to build the form to be displayed to the user.
Creation of a field format step by step:

- Add a new entry into the FIELD table to specify where the data will be stored in the database.
- Complete the FORM_FIELD entry.

5.4 How to configure a JRC Request?

A request made by the JRC is a set of file FORMAT that has to be submitted in order to answer a question at the European level. To define a request we need to:

- Add an entry into the JRC_REQUEST table.
- Define the needed files (FORMAT).
- For each CSV file, describe the list of needed FIELDs.
- Describe the COMPLEMENTARY files potentially used by some countries.