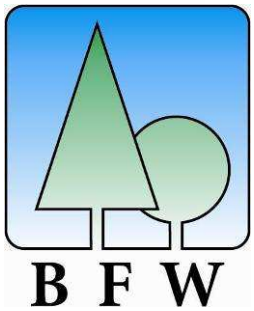


Working package 1 proposal

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: Contract No. 384104

Contents

List of figures

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)1 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 distinguish WP1(plain) and WP2 (Hatched).

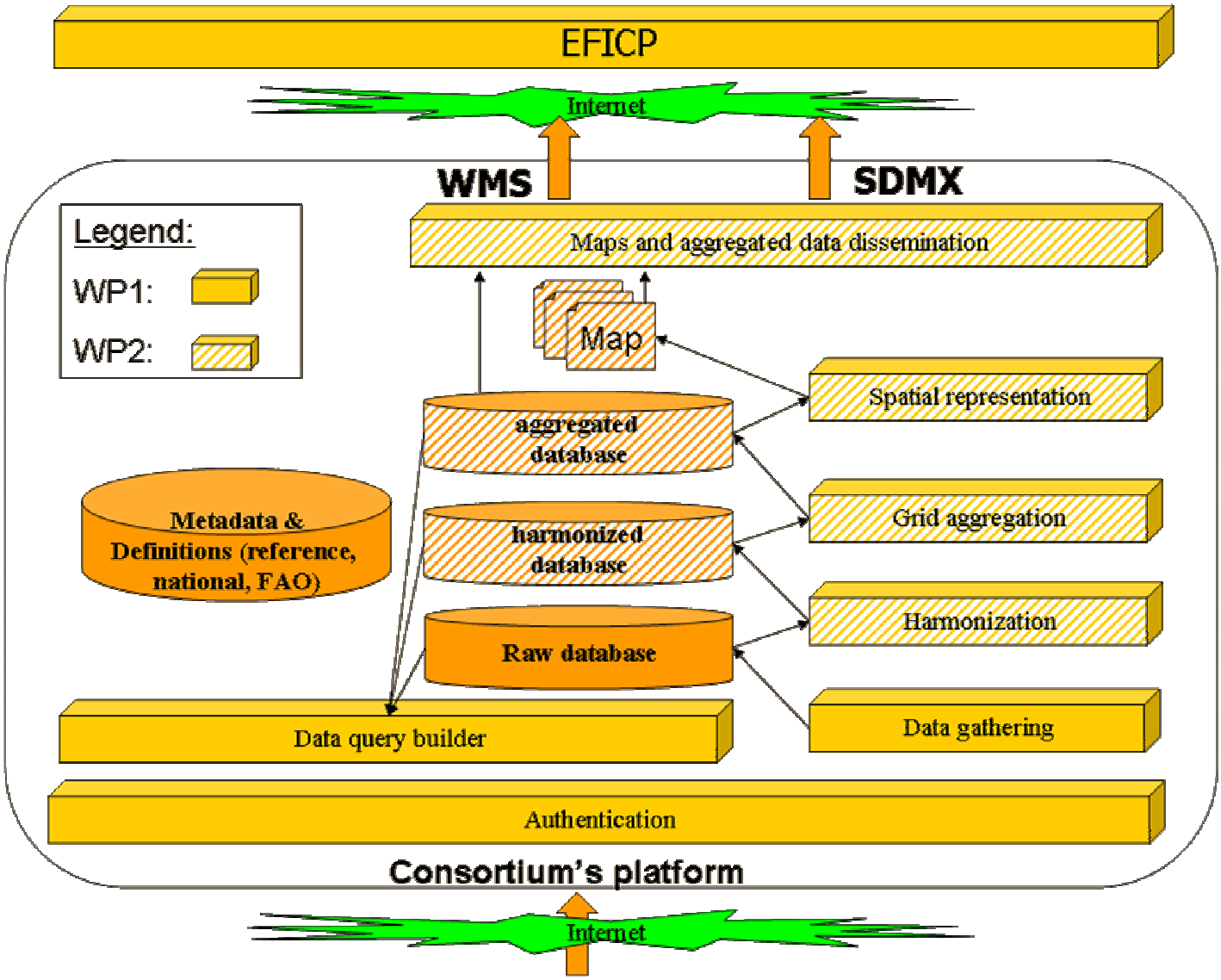


Figure 1: Consortium's platform modules

The chapter 2 provides information for both WP1 and WP2 in order to provide an overview of the whole system
The chapter 3 is limited to the WP1.

2.2 Description of consortium's platform infrastructure

2.2.1 Overview

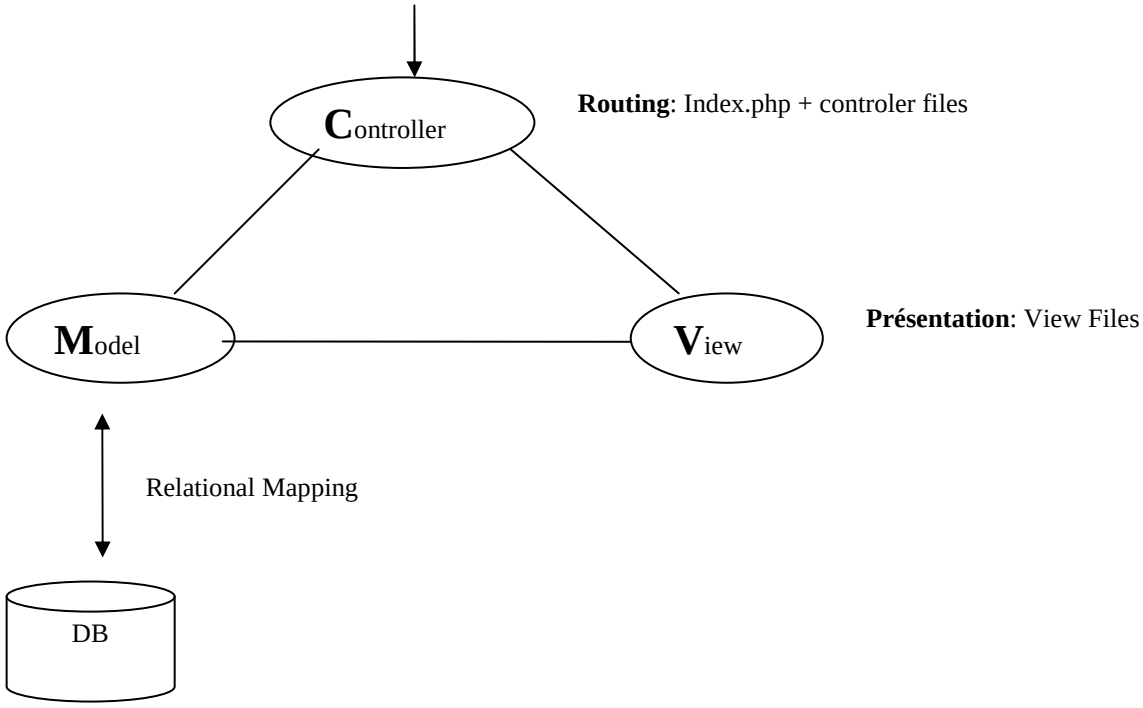
The consortium's platform is mainly developed by the French National Forest Inventory (FNFI) and hosted by the FNFI.
In order to reduce the cost of this platform, numerous modules come from other applications developed by the FNFI.
For each module developed by the FNFI under specific contract, the condition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

2.2.2 Software architecture

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 ease 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 possible to manage geographic types (point, line, polygons, …).

The following browser has been successfully tested in the same software environment:

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

The following figure shows the software infrastructure

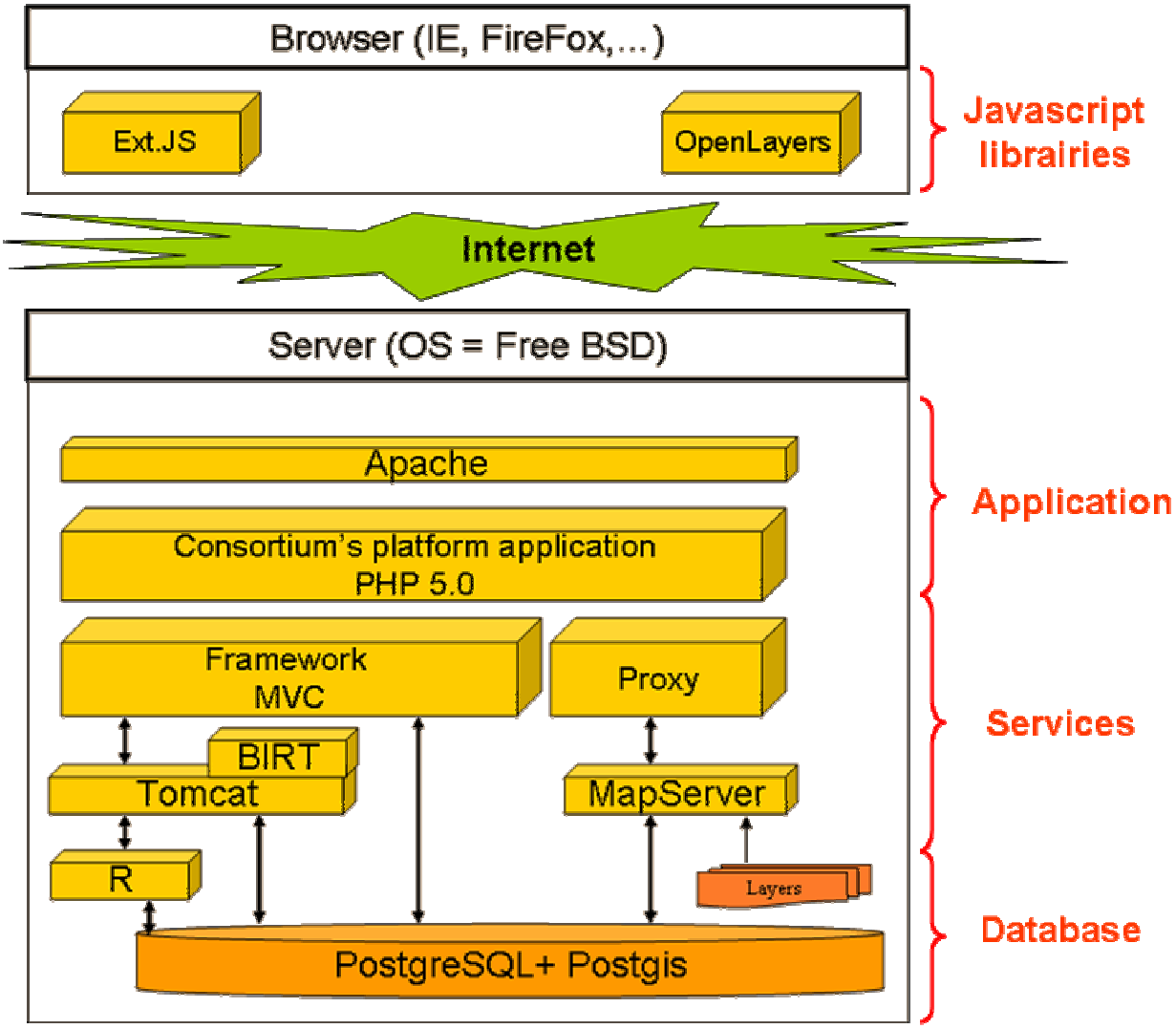


Figure 2: Software architecture

2.2.3 Physical Infrastructure

2.2.3.1 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 belong to FNFI waiting for the hardware service provider intervention.

Thus, only one server will be bought and included in our proposal. The redundant server is already available in the NFI hardware architecture.

2.2.3.2 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.

2.2.3.3 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 managed the security rules between the DMZ and the Internet accesses.
The consortium's platform will be protected by this infrastructure.
The last Internet intrusion tests were carried out by a French firm in 2006. No failure was found.

2.2.4 Maintenance

The maintenance of the system is mainly proposed in the WP2.
The maintenance proposed in the WP1 included only the hosting of the server and is limited to one year after the notification of the contract relative to WP1.

2.2.5 Layers for mapserver

Two categories of layers are used in the Data query builder module.
The first category corresponds to administrative layers to provide general information such as country boundaries, main towns, etc.
The second category corresponds to thematic layers.
All those layers used in the wep mapping modules are provided by the JRC excepted those provided by the consortium under specific contract.
The size of the layers must be compatible with the hard disk size of the server.

2.3 Development organization

The system is developed at the French National Forest Inventory offices (Nogent-sur-Vernission, Fr).
Each module is specified, developed and tested according to the Quality Plan.
Once a module is successfully tested and formally accepted, it should be installed as a working module in the consortium'platform hosted at the FNFI premises.
SubVersion will be used as a automated version management.

3 Consortium's platform: Working package 1

3.1 Main functionalities and goals

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 his 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 to 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 provider a friendly way to upload data files containing raw data. In response to these uploads, the data provider 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

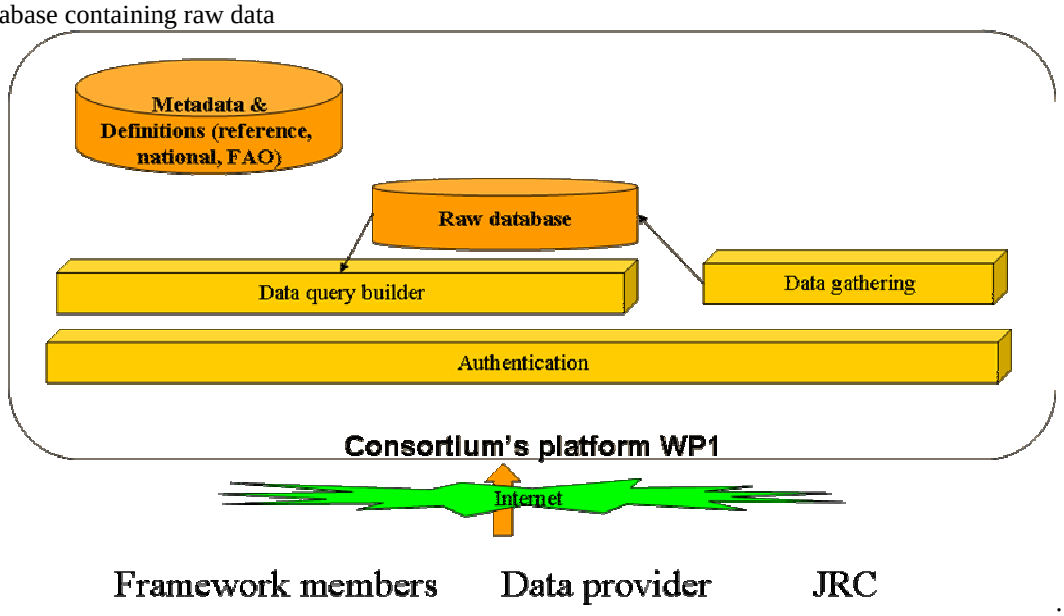


Figure 3: WP1: consortium's platfrom

3.2 Database descriptions

3.2.1 Overview

The purpose of the database is to store raw data and provide authenticated user access to validated raw data..
This database called **Raw Database** has to store data compliant with metadata according to national definitions or reference definition.

A **Metadata database** is required to provide structured information about the data managed in the Raw Database and to offer the possibility to harmonized 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.

The **Web Server Database** is used to configure the web server application.

3.2.2 Metadata database

3.2.2.1 Metadata advantages

Metadata means 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, definition (national, reference), flat files format, database tables description, etc.

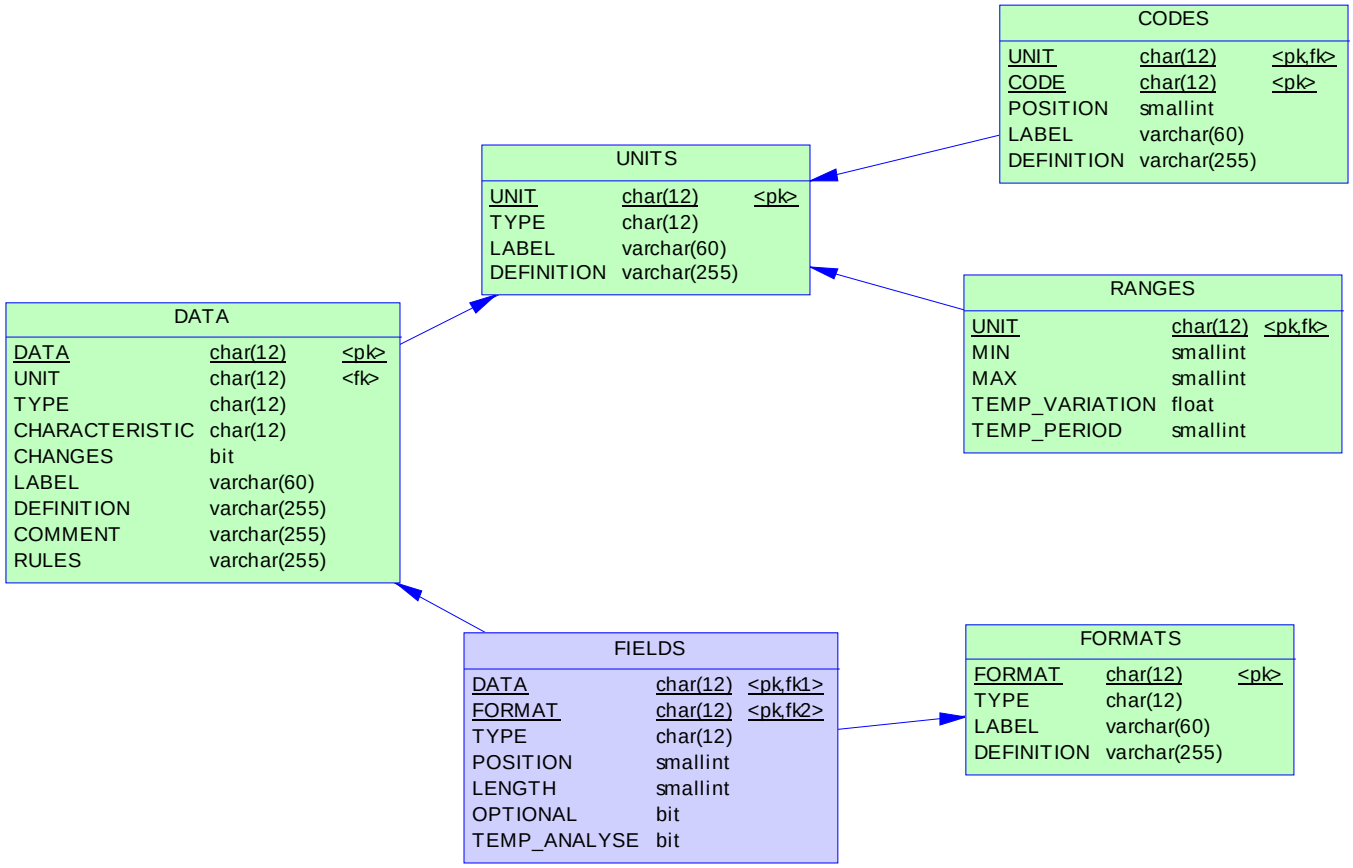
For each **parameter**, descriptive information means: name, type, units, definition, rules, the associated code list, the possible range values, etc. All these characteristics are stored in this database.

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.
All the **flexibility** depends upon the Metadata base structure. As indicated in the following chapter, it will be possible to **monitor new parameter,** to **modify code list** of a parameter without changing the consortium's modules.
All these changes can simply be applied by adding some new records in the Metadata database.

3.2.2.2 Metadata specification

The figure below shows a part of 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.



Legend:
Pk: Primary key
Fk: Foreign key
Each box represents a database table.
Each arrow represents a database foreign key (integrity constraint).

Figure 4: Data model for Metadata

3.2.2.2.1 Data description (tables DATA, UNITS, CODES and RANGES)

The data description is used to model information about all parameters used in the consortium's platform and their likelihood 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 Unit	SPECIES
TYPE	Data type (smallint, char, varchar(..), float, etc.)	Char(2)
CHARACTERISTIC	Characteristic (CODE, MEASURMENT, VALUE)	CODE

Table column	Description	Example
CHANGES	Flag to notify if UNIT has changed since the first entry	NO
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)	

UNITS Table:
The characteristics can be discrete or continuous. The first one (Nominal or ordinal) are expressed with a number of possible values stored in the CODES table. The second one are described using universal measurement units. In case of continuous type, the possible range can be specified in the RANGES table.

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

CODES 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	

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

Table column	Description	Example
UNIT	Unit name	pH (CaCl2)
MIN	Minimum range	2
MAX	Maximum	9
TEMP_RANGE	Maximum range used in temporal analysis	1 (pH unit / 10 yrs)
TEMP_PERIOD	Period to observed in temporal analysis in year	10

3.2.2.2.2 Format description (tables FORMATS and FIELDS)

The format specification describes where and how the parameters are stored.

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

Table column	Description	Example
FORMAT	Parameter name	PLOT
TYPE	Format type (TABLE,FILE)	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

FIELDS Table:
This table contains fields (file format) and column (table format) information.

Table column	Description	Example
DATA	Parameter name	INVENTORY_DATE

Table column	Description	Example
FORMAT	Format name	PLOT
TYPE	Format type (COLUMN, FIELD,…)	COLUMN
LENGTH	Field length	1
POSITION	Field position	5
OPTIONAL	Indicates if this field is **optional or mandatory**	NO

3.2.2.3 A flexible and adaptable system

3.2.2.3.1 How to modify the nomenclature?

These changes can be necessary if more or less details are needed.
In that case, the CHANGES column of the DATA table is filled with 1 (YES)
A new Metadata table called NEW_UNIT is necessary to keep the traceability of this change.
This table indicates that the old unit (UNIT) specified in the DATA table must be replaced by the new unit (NEW_UNIT) if the current date of the plot inventory is older than the date (DATE).

NEW_UNITS		
DATE	datetime	<pk>
UNIT	char(12)	<pk>
NEW_UNIT	char(12)	

The new unit must be added into the UNIT table.
If this new unit is nominal or ordinal, the new code list is inserted into the CODES table.
If this new unit is continuous, the new record range is inserted into the RANGES table.

The following table is filled to establish the relationship between the old code and the new code (if possible).

GROUPS		
UNIT	char(12)	<pk>
CODE	char(12)	<pk>
NEW_UNIT	char(12)	<pk>
NEW_CODE	char(12)	<pk>

3.2.2.4 Metadata adaptation to provide translation between national definition and reference definition (from WP1 to WP2)

For a qualitative variable (unit type = nominal), the metadata database as described previously is adapted to take into account differences between national list of code and reference list of code whenever correspondences exist.
For both quantitative variable and qualitative variable, the **metadata database** schema **needs to be adapted** to take into account that the **definitions are not the same**.
For example, even if a national qualitative variable has the same list of codes than the reference one, if the definition or the threshold of a code is not the same, the mapping between the national and the reference unit won't be possible.

During WP1, the consortium will start to work on the meta database schema to take into account the output of the COST Action E43.
The purpose of this work is to prepare the meta database to the WP2 in order to harmonize Raw data from national definition to reference definition.
The end of this activity will be part of the WP2 when the bridge container will be studied and harmonization will be carried out.

3.2.3 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 an request type.
The first step in transferring data corresponds to a file containing plot coordinates and a plot identifier (plot_id) . This file allows the management of permanent plot by reusing the same plot identifier.

The second step corresponds to a data transfer of :

- plot variables referencing the plot_id
- tree variables referencing the plot_id and tree identifier.

Some variables such as plot coordinates, date of assessment are mandatory.

In order to reuse the generic application called “Base de données naturalistes (BDN)” and the integration module of the biosoil application, the BDN schema is adapted to introduce tables used during the data upload.
The BDN application doesn’t know the data to manage. This application discovers all the metadata for a dedicated protocol in order to prepare all the screen forms and queries.
If we refer to the framework needs, the qualitative and quantitative variables requested for the provision of a service will be identified in each JRC request and inserted into the metadata database.
As the metadata database and the raw database are generics, it will be possible to add new parameters without modifying the raw database structure.
To do that, two generic tables (sample_compl for plot variables and occurrence_compl for tree variables) are already available in the BDN schema. In those tables, each line should have corresponded to a column in the plot and tree table if the schema had not been generic.
The schema below represents the main entities and their associations. This schema will be adapted during the specification phase.

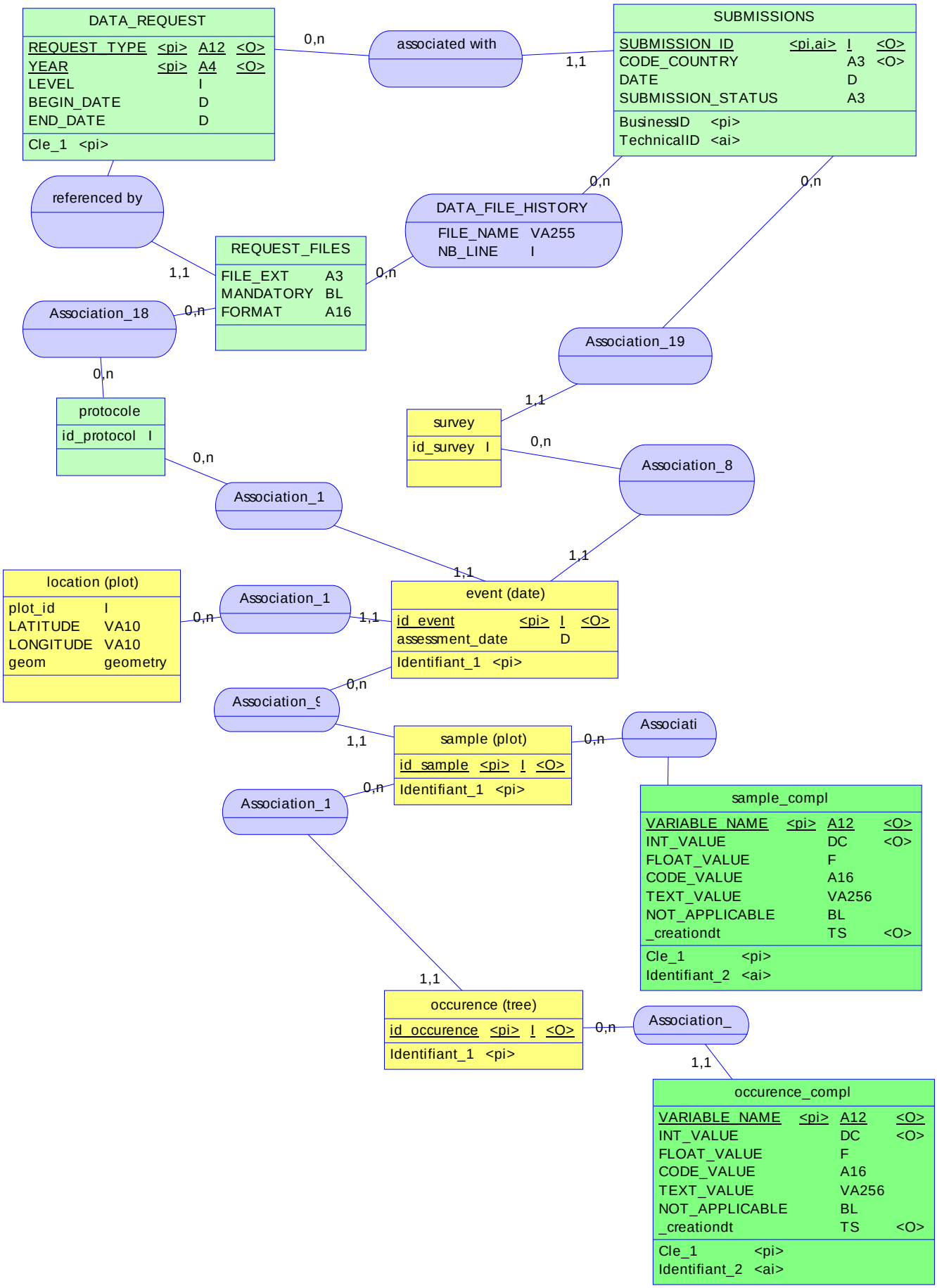


Figure 5: raw database principles

3.3 Main modules

3.3.1 Modules description

All modules are written in PHP language or JAVA language. Each module is composed of classes.
The database interface is written in ANSI SQL standard language (SQL 2).
Each module is self documented and a full description of all classes can be automatically generated using PHPDOC or JAVADOC.

3.3.2 Authentication module

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

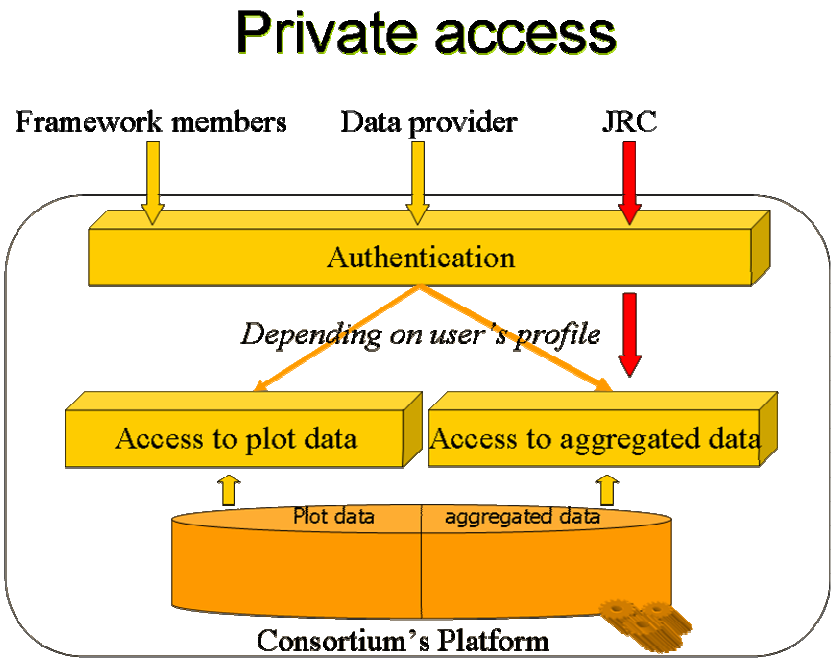


Figure 6: authentication module

3.3.3 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,….

This module is based on the biosoil integration module. This module must be adapted to take into account the BDN schema, permanent plots, incremental upload, …

The main part of the validation checks is based on the **Metadata database** and the **Validation Database**.
The files to upload are documented in the interface specifications. 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.

These mechanisms will be detailed in the software specifications.

The validation checks are composed of two types of checks:

- The ***compliance checks*** verify the accordance of 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 unit, data values, ranges, etc.). If the data are not fulfilling the compliance checks, then they are marked as errors and are rejected.

- The **conformity checks** verify if the data are consistent with other data using cross-checks. 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.

A **dedicated database called "validation database"** is necessary to store all the **parameter cross checks**. This database is composed of the following tables:

- 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.
- The **CHECK_ERROR** table contains the results of the checks.

The various processing checks validate the data provided and ensure that only reliable and comparable data are made available for data dissemination.
Depending on the type of checks, an unsuccessful check can be considered as an **error (stopping the check process)** or just a **warning**.
In case of fatal error (compliance check only), the whole data file is rejected.
In case of warnings, the data provider is aware. The file can be corrected and resent.

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

Reports on the status of the transmitted data by data provider are generated automatically in response to an upload operation. The report is produced instantly through the web application.
The report contains administrative aspects:

- Data provider id,
- submission date,

and all technical aspects that can help the data provider to follow up its check operations:

- compliance results
- conformity results (only when compliance checks are successful)

The data provider would have the opportunity to download the data transmitted status report in order to correct the wrong data. Thus, the same files could be submitted several times, without any restrictions
When the data are validated (submit action), the data provider won't be able to upload another time the same files (same request type, same year)

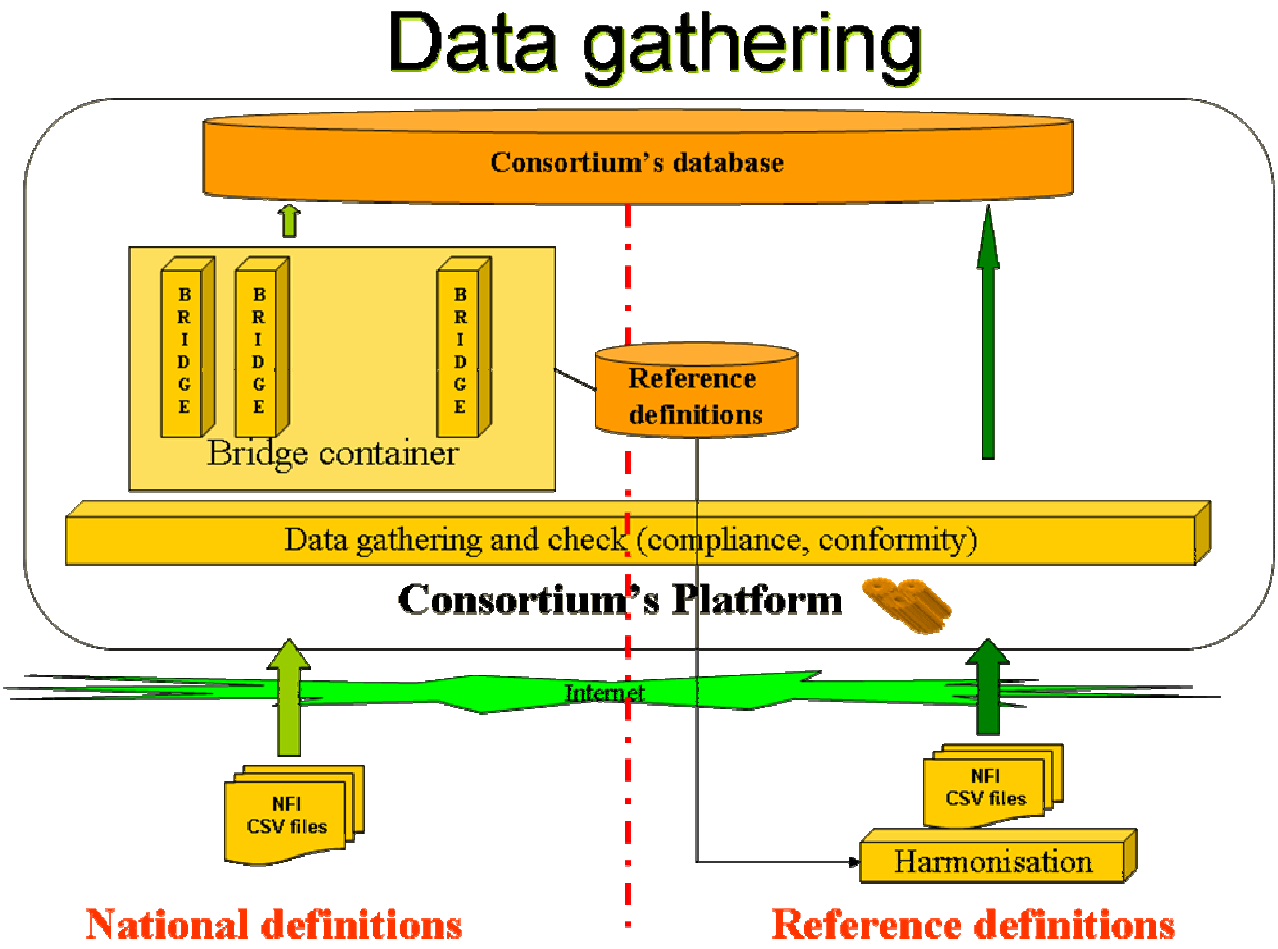


Figure 7: data gathering module

3.3.4 Data query builder

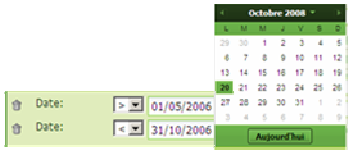
This module allows a user to carry out its own request.

This module is based on the application called "Base de données naturalistes" (BDN) developed by the FNFI and Camptocamp (web mapping part) for the ONF (Office national des Forêts). The BDN license called CECILL is the French translation of the GPL license. Thus, this application can be adapted and used in the consortium's platform. This module must be adapted to take into account the upload module, user's restriction, the web map context (European level instead of French boundaries, etc)

On the left side of the screen, for each theme among location, temporal criteria, plot data, tree data, the user can choose the columns to be screened. He can also restricted the query by adding criteria and by selecting a specific value.
The query can also be restricted using the web mapping module in order to limit the query to a polygon drawn by the user.

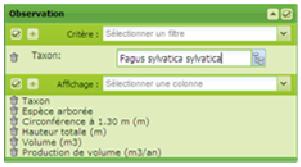
Users can build their own request:

When ? Date, period



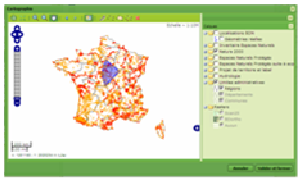
What ? Plot criteria or columns,

Tree criteria or columns



Where ? Nut codes

Polygon drawn by user



When the query is ready, the user can run it.
The result will be limited taking into account the user's profile and the status of the data (validated or not).
Then, the table is dynamically constructed. The user can move from page to page.
The results can be exported in CSV format or SHAPE format.

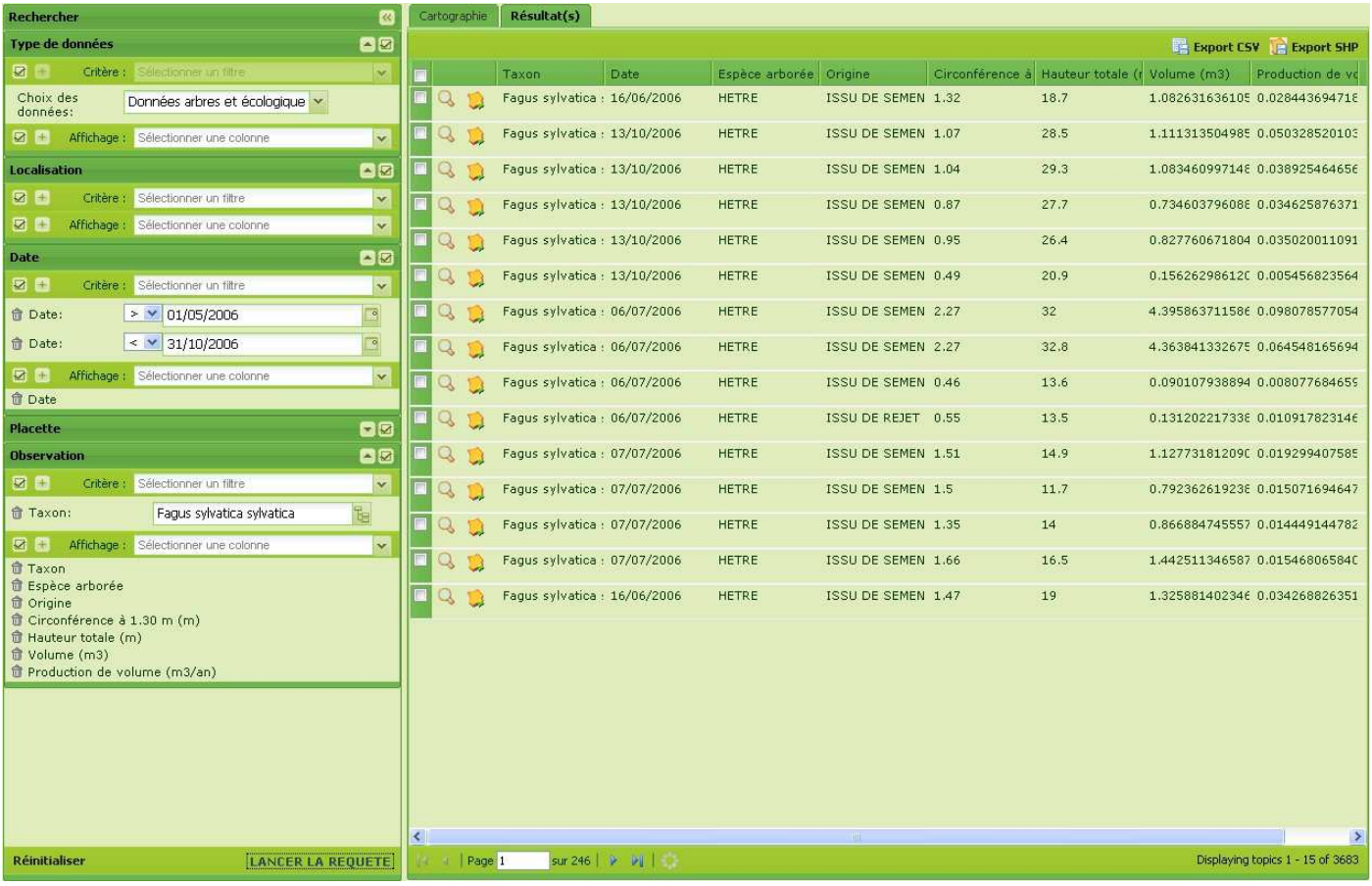


Figure 8: data query builder screenshot

The user can also ask to map representation corresponding to the table content.
In this map each line is represented by blue point
This module is useful to overview all the data or to focus on specific topic or specific specie.

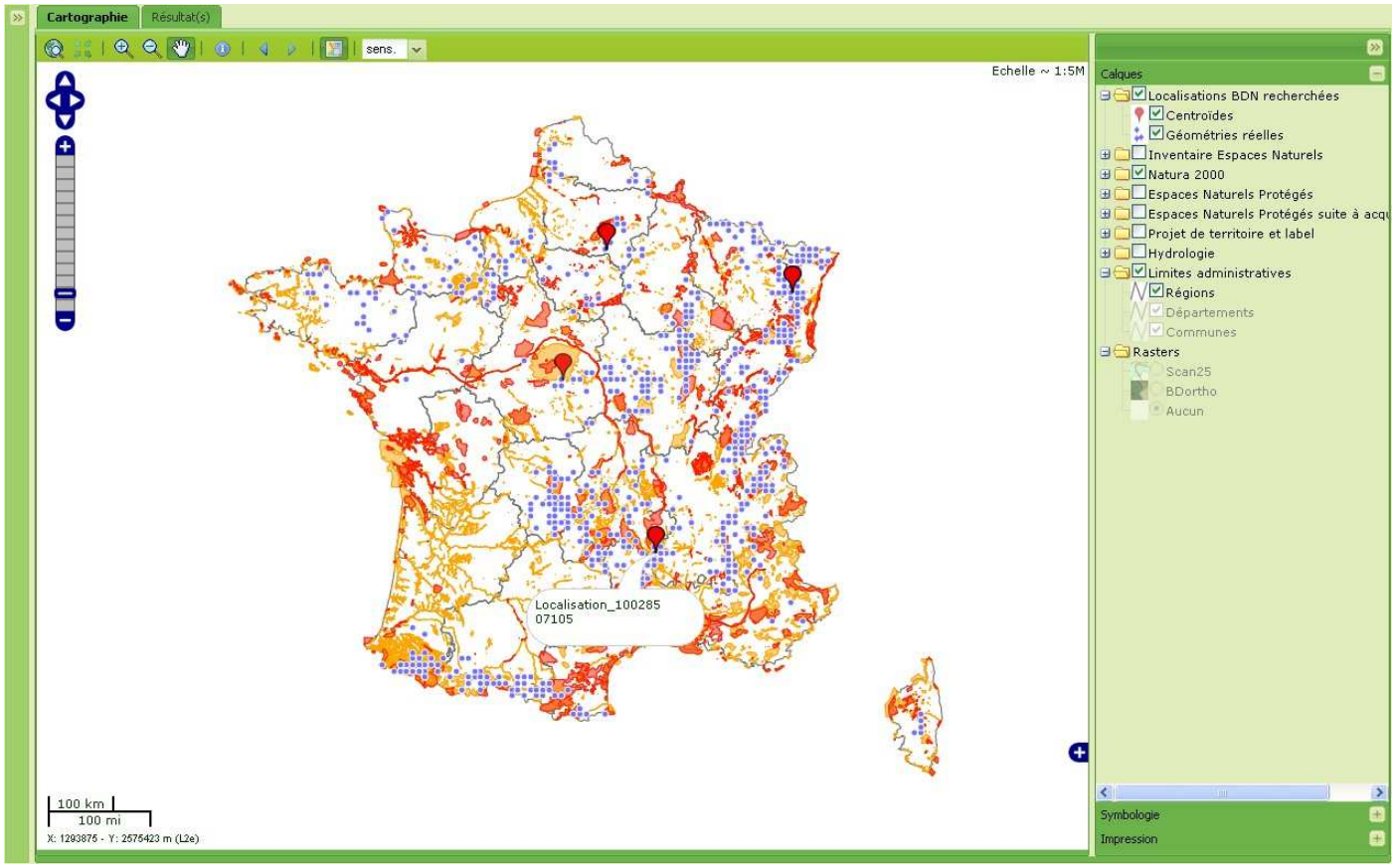


Figure 9: data query builder web map

4 Work plan

4.1 Staff description

4.1.1 Overall manager

Function (FNFI): DIRECTOR OF THE FRENCH NATIONAL FOREST INVENTORY
Name: Claude VIDAL
Task in the project: Project manager
Professional Experience: 33 years
Task in the project: Organisation, Plan, follow up, Meeting, Management, etc…

4.1.2 Thematic and scientific expert

Function (WSL): LEADER OF THE RESEARCH GROUP INVENTORY DESIGN AND PLANNING
Name: Adrian LANZ
Task in the project: thematic scientific officer
Professional Experience: 20 years
Task in the project: metadata and harmonisation tools

4.1.3 Project manager

Function (NFI): HEAD OF THE INFORMATION SYSTEM DEPARTMENT
Name: Jean-Luc COUSIN
Professional Experience: 15 years
Task in the project: Global follow up, Meeting, review, expertise in Information System

Function (NFI): PROJECT MANAGER
Name: Benoit PESTY
Professional Experience: 10 years
Task in the project: Follow up, Meeting, Database and system design

4.1.4 Project team

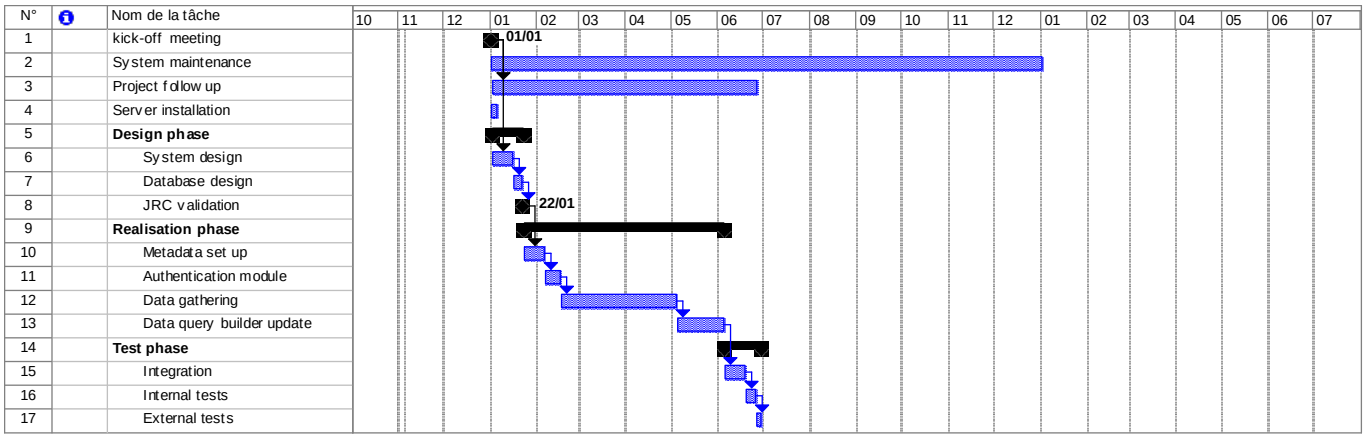
Function (NFI): DEVELOPPER
Name: Sylvain Galopin
Professional Experience: 2 years
Task in the project: Data query builder, bridge container

Function (NFI): DEVELOPPER
Name: Florent Bourcier
Professional Experience: 2 years
Task in the project: User Interface

4.2 Work repartition

See § financial proposal

4.3 Time frame



The project duration is one year after the notification.
The first six months correspond to the specification and development phases. The last six months correspond to the maintenance phase.

4.4 Economical frame

WP1				
Price component	**Personnel type equivalence**	**Price per day**	**Total days**	**Total price**
Labour				
FNFI Overall manager	Project manager	850	5	4250
FNFI senior project manager	Thematic scientific officer and data analyst	750	10	7500
FNFI project manager	Thematic scientific officer and data analyst	750	28	21000
FNFI system administrator	Programmer	400	5	2000
FNFI developper	Programmer	400	118	47200
FNFI Gis operator	GIS operator	400	7	2800
Expert	Thematic scientific officer and data analyst	750	2	1500
Subtotal Labour (1)				86250
Other				
Server				5000
Tapes X5				600
Subtotal Other (2)				5600
TOTAL (1)+(2) = PRICE WP 1				**91850**

This proposal is valid until the 05th of May 2009