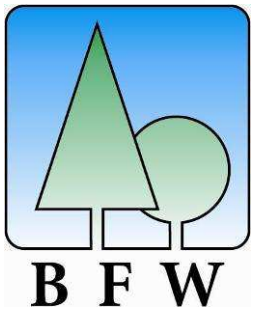


Specific Contract 2 (SC2) proposal (version 2)

Enrichment of the EFDAC E-forest platform (SC2.1)

"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

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) 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 EFDAC E-forest platform

2.1 EFDAC E-forest platform overview

In order to provide the required services, a platform called “EFDAC E-forest 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 EFDAC E-forest platform.
The first working package corresponds to the building of the EFDAC E-forest platform to perform raw data upload, storage and queries based on a central metadata base.
The second working package corresponds to an enrichment of the platform to run specific processes. As requested by the JRC, this Specific Contract is separated into two separated proposals. This proposal (called SC2.1) concerns only the following modules: aggregation, spatial representation and a further proposal (called SC2.2) will concern the harmonization and the dissemination to EFICP. The previous proposal concerned the whole Specific Contract 2 that means SC2.1 plus SC2.2.
To avoid losing the main purpose of this Specific Contract that is data harmonization, the specifications remain unchanged but when it is necessary, limitations are given to specify functions out of scope.
The picture below shows the main modules by distinguishing SC1 (plain) and SC2 separated in SC2.1 and SC2.2 (Hatched).

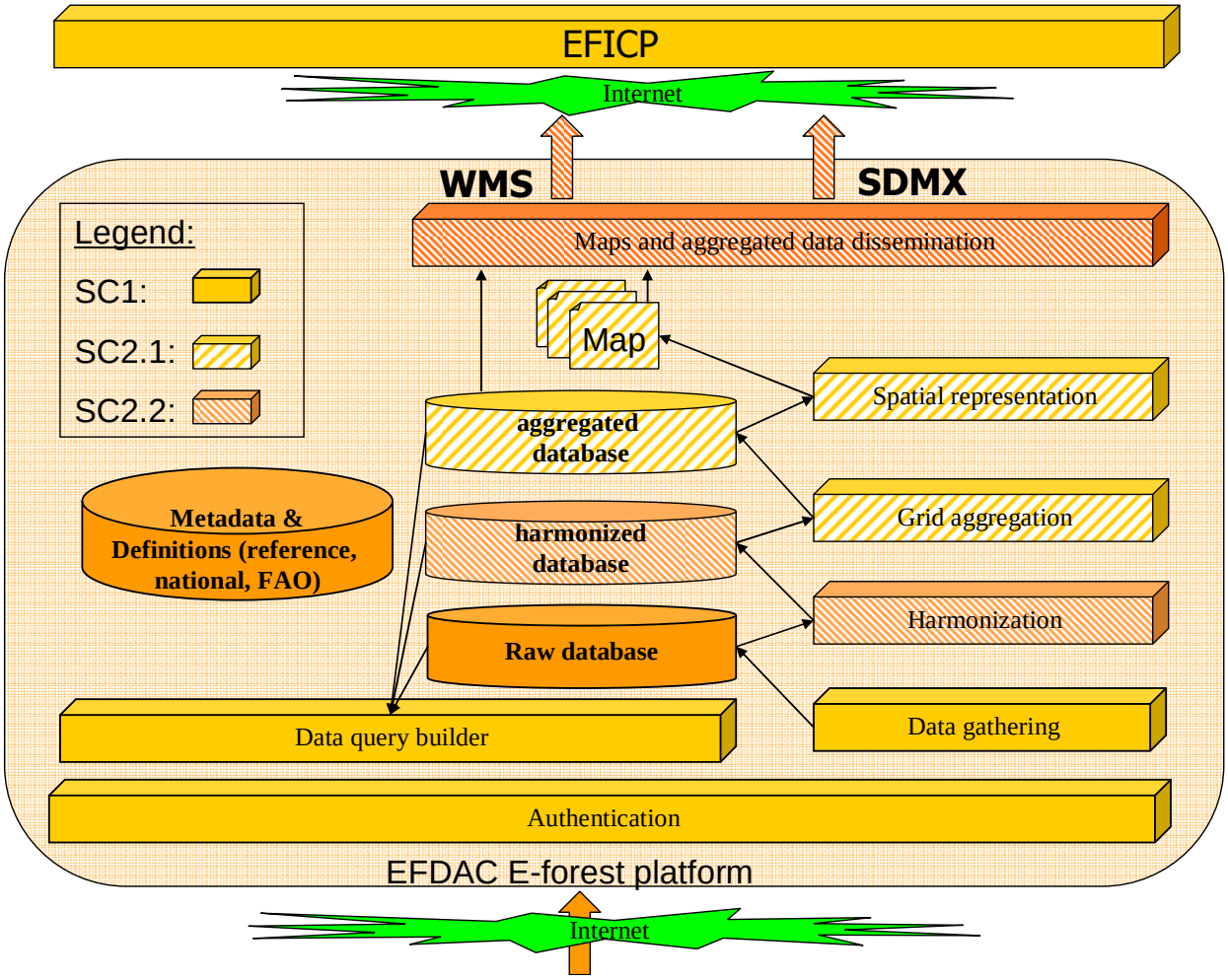


Figure 1: EFDAC E-forest platform modules

The chapter 2 provides information for both SC1 and SC2 in order to provide an overview of the whole system
The chapter 3 is limited to the SC2.

2.2 Description of EFDAC E-forest platform infrastructure

Refer to SC1 proposal

2.2.1 Maintenance

The maintenance of the platform defined in this proposal is limited to **one year** after the contract signature.
The maintenance that cover all the framework contract will be provided in a further proposal in response to another specific request coming from JRC.

2.2.1.1 Hardware maintenance

The hardware maintenance is provided and supported by the JRC.

2.2.1.2 Web server maintenance

The standard maintenance tasks consist of:

- Performing back-ups of database and web server modules in case of changes.
- Updating the necessary packages to run the web server application.
- Updating the web server modules.
- Assistance in case of software dysfunction.

2.3 Development organization

Refer to SC1 proposal

2.4 System transfer to the JRC

As the consortium platform is considered as a part of the EFDAC system, the fully functional system will be installed at the JRC at the end of the contract.
The place of delivery is the Land Management Unit of the JRC, Ispra.
All the raw data will be destroyed. Only the aggregated data will be transferred to the JRC.
This system transfer is out of scope of this proposal because it is postponed to the SC2.2.

3 EFDAC E-forest platform: Specific Contract 2

3.1 Context

The Working package 2 (SC2) considers that the EFDAC E-forest platform as defined in the working package 1 is already partially developed. That means that the module called data gathering is operational.
The SC1 was mainly dedicated to authentication, data gathering and data queries. The main purposes of the SC1 were to offer a user friendly interface to carry out these operations.
The SC2 corresponds to an enrichment of the platform in order to carry out the main jobs for which the platform exists: the data **harmonization** (out of scope of this proposal) and the data **aggregation** in order to **provide to the JRC services** or **data** at a European level.
These processes are mainly core processes without user interface.
These processes are based on the result of the COST action E43.

3.2 Main functionalities and goals

3.2.1 Previous proposal

The goals of the new functionalities are to provide a tool to produce statistics or maps at a European level.
With the SC1 platform, the uploaded data were compliant with national definitions.
Using the reference definitions provided by the COST action E43, the metadata database is enriched to take into account that these reference definitions are based on core quantitative variables with theirs associated threshold and core qualitative variables with their key modalities.
Thus, the process for harmonization can be developed to compute the raw data and transform them into harmonized data using the central metadata database. This process represents the cornerstone of the EFDAC E-forest platform.
When the data are harmonized, the platform provides a process to aggregate them using different geographical units (e.g.; NUTS level or a reference grid.)
The grid data can be transformed into a map using linear regression or a more sophisticated smoothing method. This map has to be considered more like a graphical thematic representation instead of a spatial representation.
As the EFDAC E-forest platform is considered as a part of the EFDAC umbrella, communications are made possible between the EFDAC E-forest platform and the EFICP in order to disseminate data.

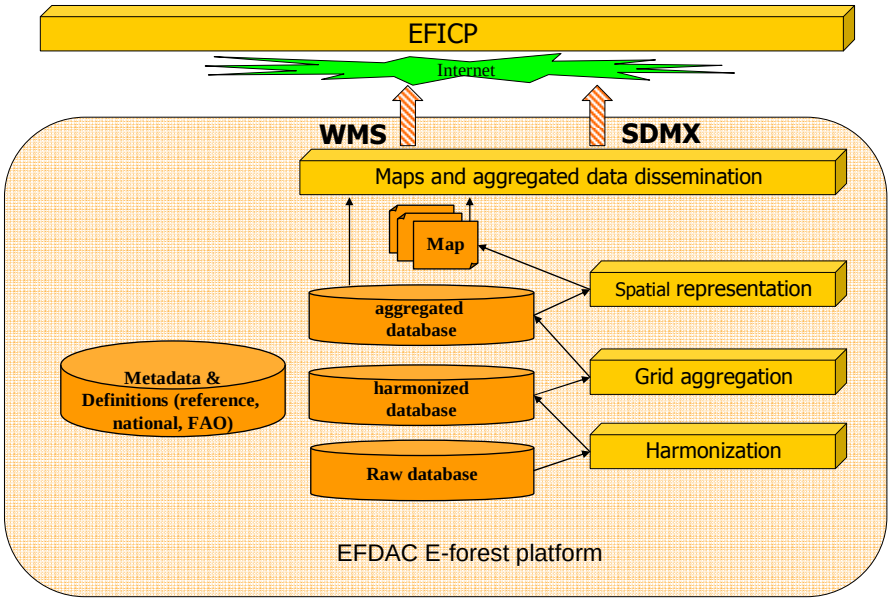


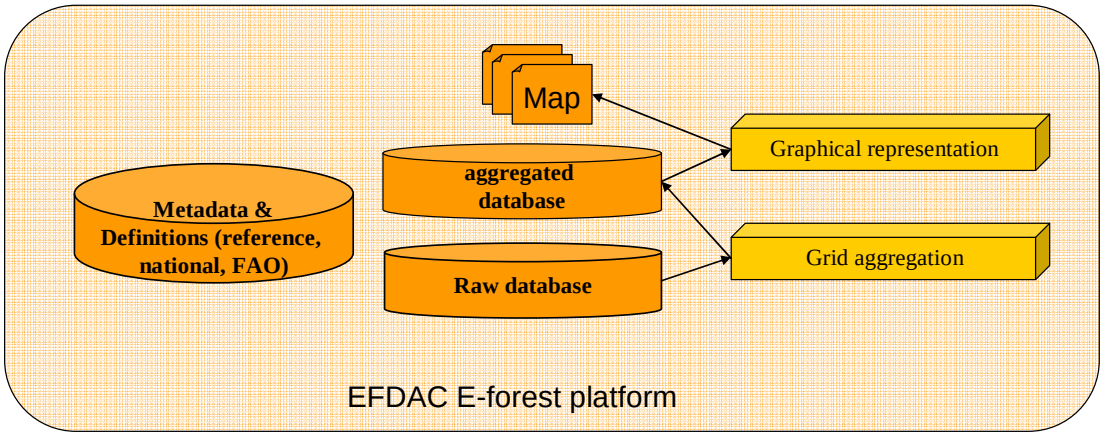
Figure 2: Enrichment of the EFDAC E-forest platform within SC2

The figure 2 shows the new functions and databases.
On the left side the three main databases store data coming from data providers: raw database, harmonized database, aggregated database. The meta database is used to describe all the data managed in all the databases.
On the right side, the main processes are defined: harmonization, aggregation and graphical representation.

The arrows represent the relationship between the databases and the processes.

3.2.2 Current proposal

Because of the harmonisation and data dissemination modules are postponed to further proposal, the figure 3 shows the new functions concerned by this proposal



3.3 Database descriptions

3.3.1 Overview

To simplify the reading of the diagram, two new databases are added to store **harmonized data** and **aggregated data**.
During the specification phase, the three databases for raw data, harmonized data and aggregated data will be studied.
One output of this study will be the conceptual model in which the three databases may be merged if the data categories (raw, harmonized and aggregated) can be labeled using a status.

The **metadata database** will be **enriched** to take into account the fact that the definitions are based on core quantitative variables with theirs associated threshold and core qualitative variables with their modalities.

Note: The harmonisation module being postponed to further proposal, the update of the metadata database is out of the scope of this proposal. The following chapter is kept for information only.

3.3.2 Metadata database

The enrichment of the metadata database will be carried out in three steps:
1) the metadata database structure will be improved to take into account that a definition can be based on more than one core variable and hierarchical representations must be implemented.
2) the metadata database will be filled by some of the definitions gathered by the COST action team.
3) the metadata database will be tested using the harmonized process.
This process can be iterative until the database structure is optimum.

3.3.2.1 Modification of the metadata database structure

The list of core variables with their associated thresholds needed for the reference definition of forest are summarized in the following table:

Reference definition	Core variable	Associated threshold
Forest	tree and tree height	5 m
	minimum size	0.5 ha
	minimum width	20 m
	tree crown cover	10 %
	Land use	Excluding predominantly urban and agricultural land uses

A possible adaptation of the metadata database could be:

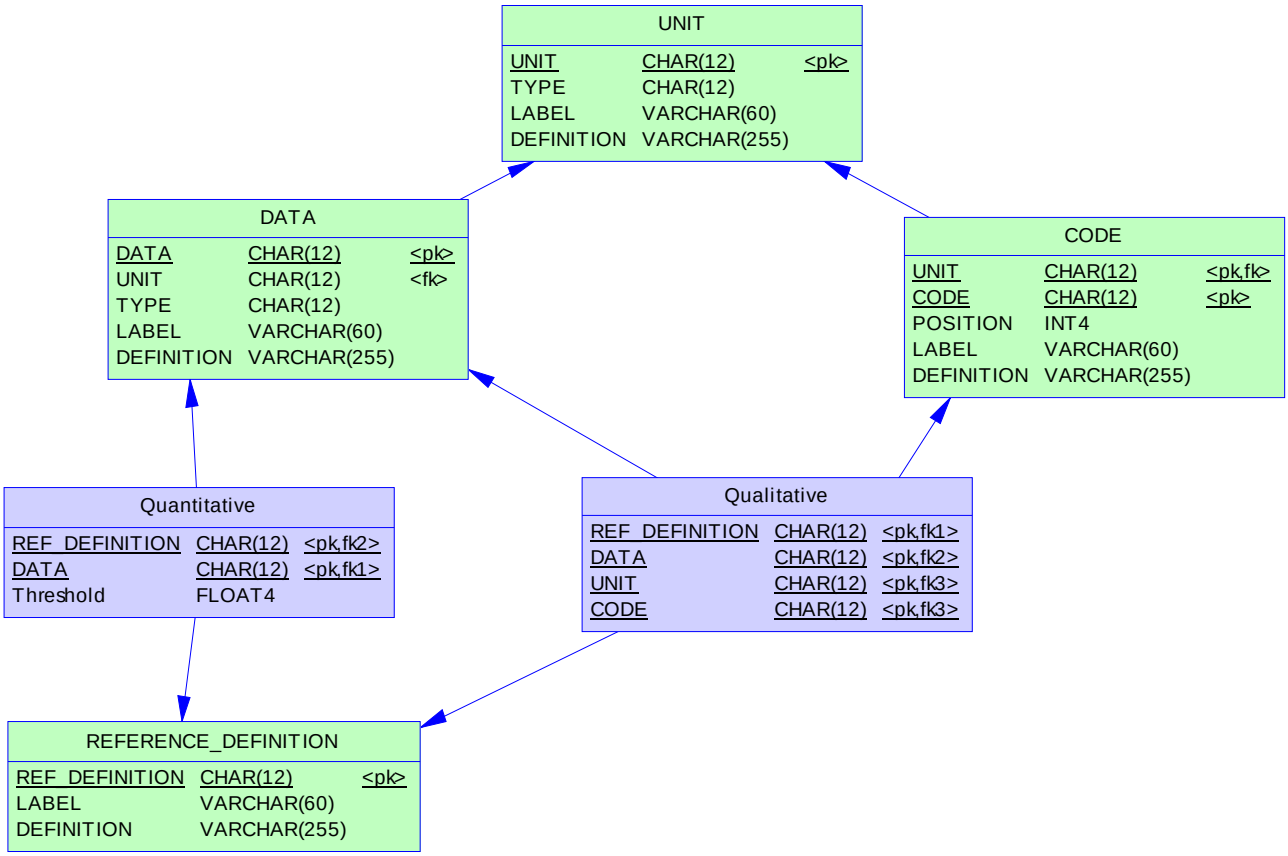


Figure 3: Enrichment of the metadata database

Where the “reference_definition” table contains the definition and the “qualitative” and “quantitative” tables contains the core variables with their codes or threshold.

If we refer to the COST action proposal, a hierarchical partitioning approach divides the whole tree into separate elements (see Figure 4. The same principle used for storing the flora taxonomic referential was implemented in the BDN application.
For each node, the system provides information such as depth, parent, children, etc. The IFN develops a special representation in order to retrieve all the nodes of a specific branch in one query without carrying out queries recursively.

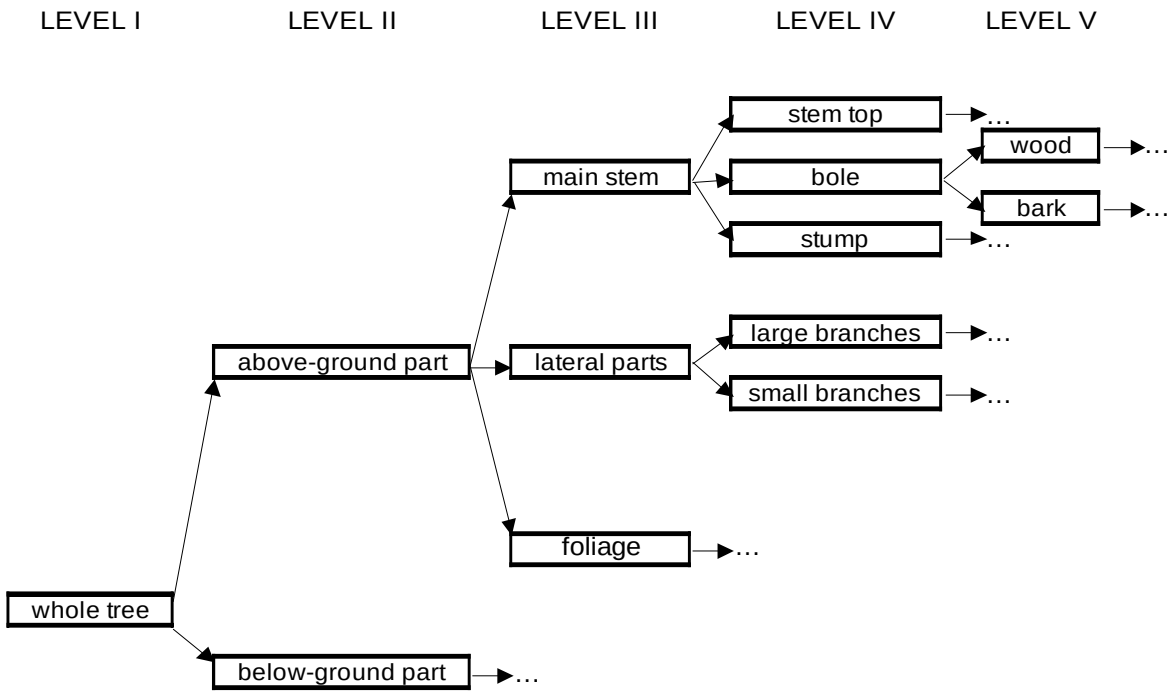


Figure 4: Tree partitioning

3.4 Main modules

3.4.1 Harmonization module

Note: The harmonisation module is postponed to further proposal, the following chapter is kept for information only.

National Forest inventories use different definitions for national purposes. Even basic definitions such as forest area, tree, are not based on a unique definition. The COST action E43 introduced the concept of reference definitions to provide a way to harmonize data at a European level. The reference definitions are based on core variables to make them clear and objective. The quantitative core variables are associated with a defined threshold. The qualitative variables are based on modalities. With these analytical definitions, the reference definitions and national definitions are described in the metadata database. Thus, it becomes possible to compare the thresholds of the quantitative variables and modalities of qualitative variables. If the units of the qualitative variable are different between national and reference definitions, the metadata database provides the structure to set the mapping between the two lists of codes (see table groups on the SC1 proposal).
To carry out the process of harmonization, a structured approach has been proposed by the COST action. This methodology is based on bridging functions.
A bridging function corresponds to recalculations based on existing data or statistical treatments to convert from existing definitions to reference definitions. Sometimes, data are missing to carry out the process of harmonization, sometimes data are available in surplus. In the first case, the bridging function is called “extensive”. In the second case, the bridging function is called “reductive”.
These recalculations are generally done at plot level and new estimates are then calculated using these new sampling results. In some cases these recalculations can be done at tree level (volume functions for example) and new estimates are calculated summing up these basic recalculations. A third way of working using modeling functions at result level can also be used. In this case there no recalculations nor at plot nor at tree level but only a modelisation applied at result level. The precision of results using this last way of working is much weaker.

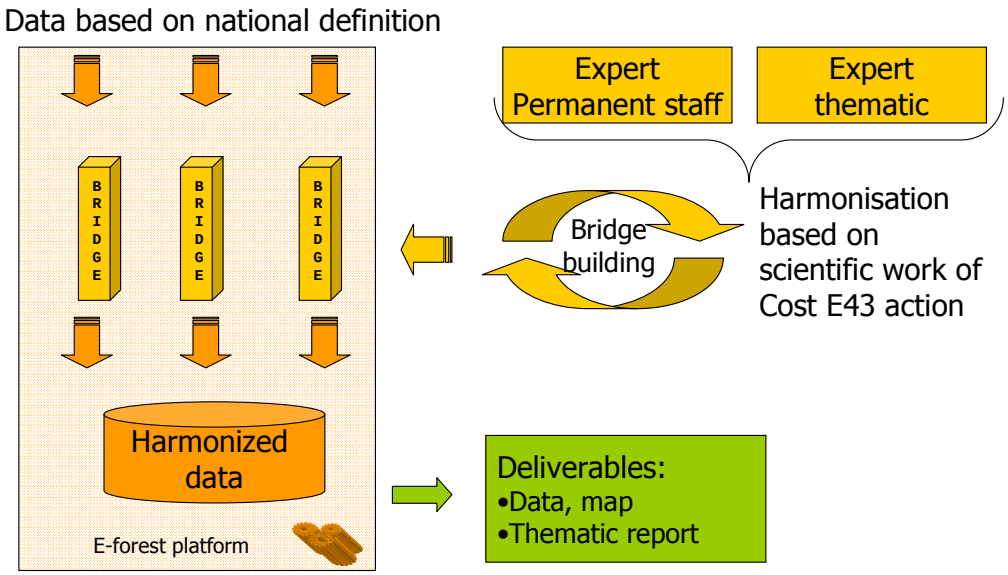
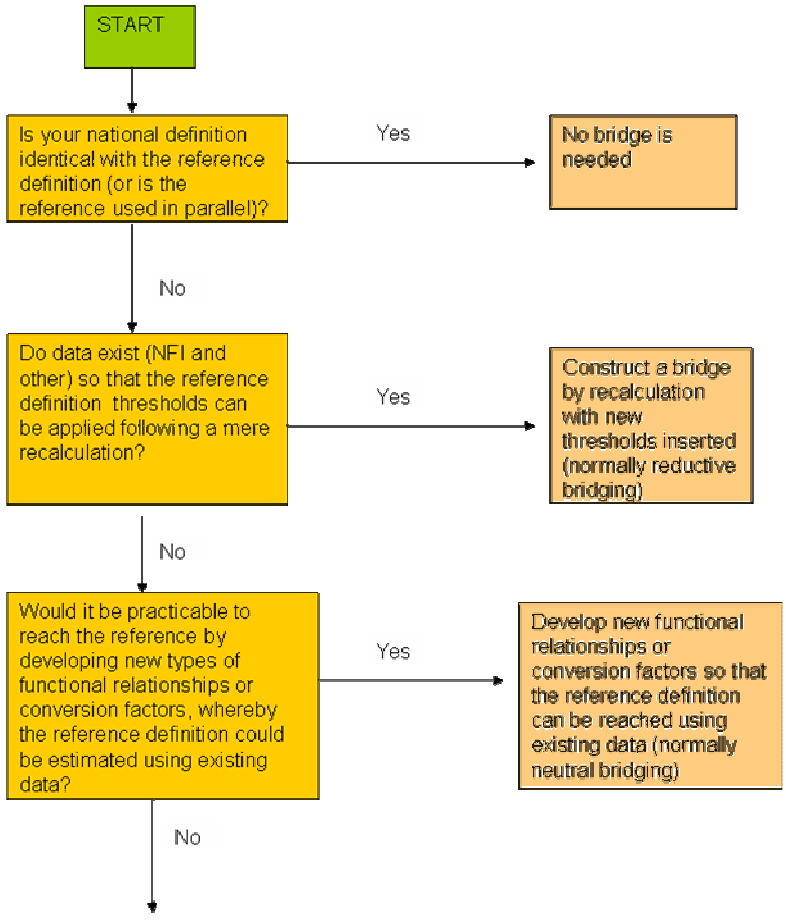


Figure 5: Harmonization process overview

The COST actions E43 provides an algorithm to define the strategy on how to implement the bridge to be built. This algorithm is the basement of the harmonization process.



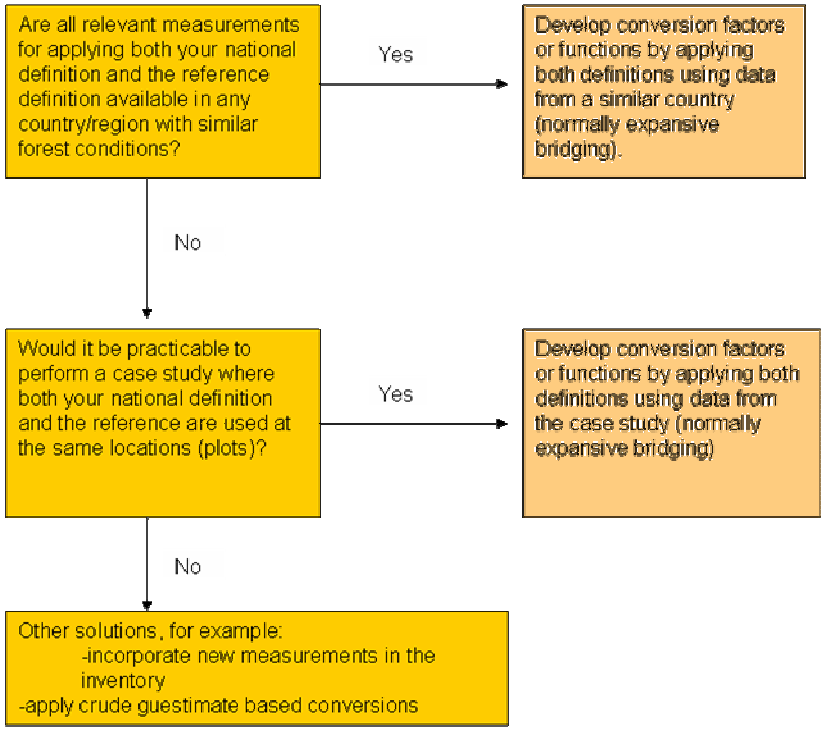


Figure 6: Bridge strategy

One of the difficulties in this algorithm is to develop it as an automatic process. Many of the questions in it can only be answered by a thematic expert.
However, our proposal includes the development of the user interface in which the expert is assisted. This interface allows the user to compare automatically the national definition and the reference definition.
When the strategy is chosen, the next step consists in developing the bridge.
In order to have a process that is able to launch the new data harmonisations, each bridge must respect a specific interface. Thus the application can automatically discover the new bridge and run the specific treatment. This principle is based on object oriented development

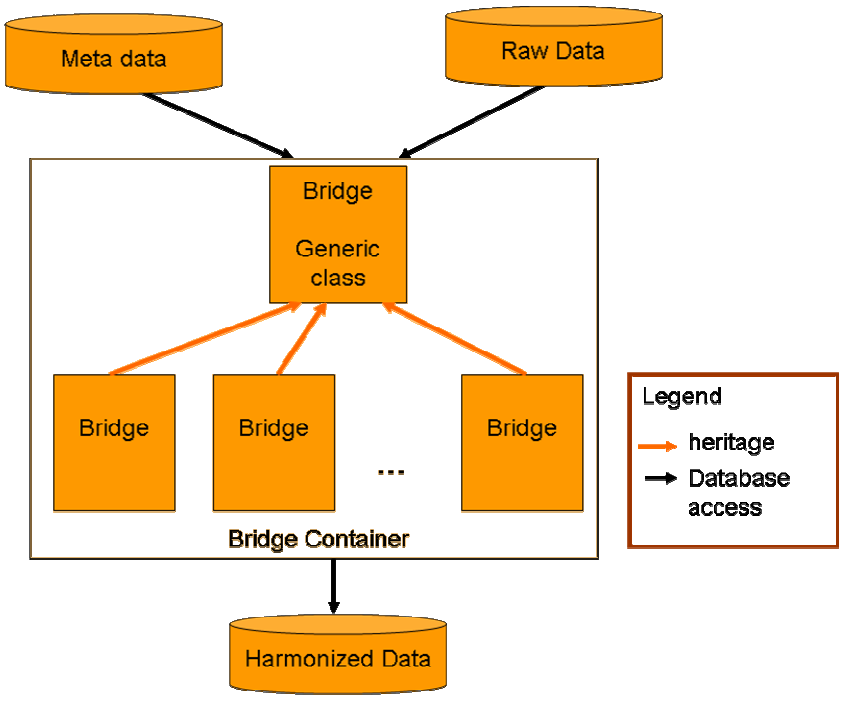


Figure 7: Bridge container

This process is **accessible** only for the **permanent staff** of the consortium in charge of data harmonization. Depending on JRC request, the data provider can also run this process for harmonizing its own data and have access to the results using the data query builder.
At the end of the bridge execution, the harmonized data are stored in the harmonized database.
We will propose adding a **workflow** to ask the data provider for a formal validation before dissemination of its harmonized data.

3.4.2 Aggregation Module

Note: The harmonisation module being postponed, the aggregation module will work on a copy of the raw data.

Systems for referencing NFI plots are mainly based on plot coordinates.
As proposed in the INSPIRE rules, aggregation can be carried out using harmonized multi-resolution grid with a common point of origin and standardized location and size of grid cells.
The recommended cell-size can be one of the following: 5x5 km, 10x10km, 25x25 km, 50x50 km.
Another possibility of aggregation is the use of NUTS codes. In that case, the NUTS code will be inferior or equal to level 3
The following figure shows the data transformation.

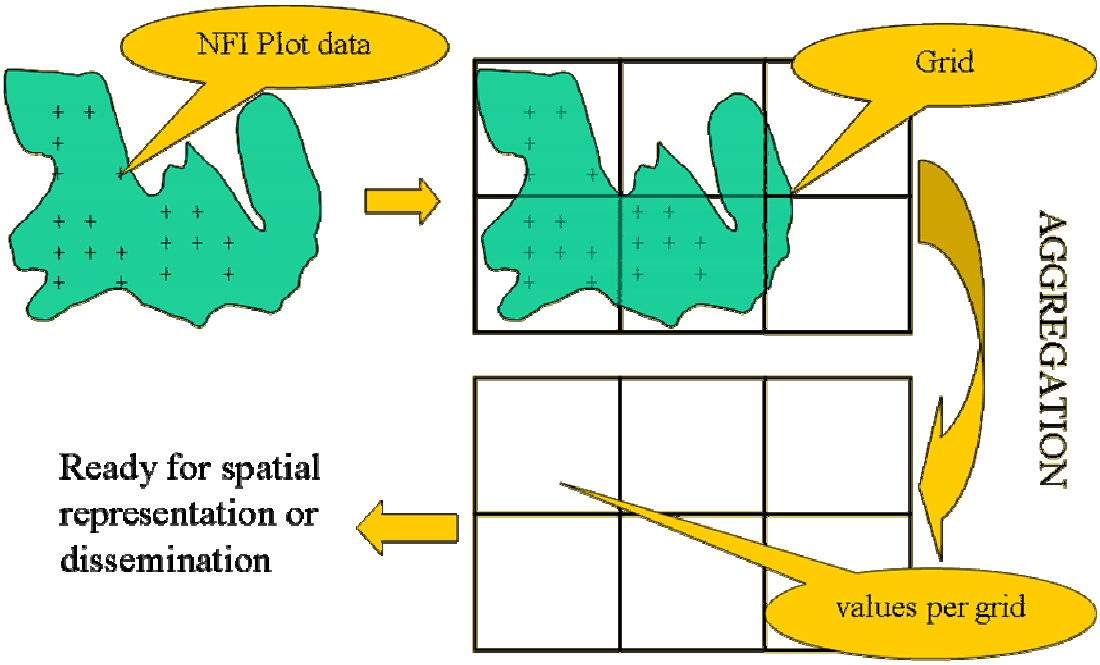


Figure 8: spatial aggregation

This module will be accessible through the user interface. Thus it will be possible to set up the main parameters such as country, grid reference or NUTS level and run the process.
Most of the treatments will be carried out by developing SQL procedures using the spatial cartridge of PostgreSQL called PostGIS.
The access to this module will be limited to permanent staff only.

3.4.3 Spatial representation

The spatial representation concerned by this proposal is mainly based on interpolation.
One of the most commonly used techniques for interpolation of plot or grid is inverse distance weighted (IDW) interpolation for qualitative variable only. Inverse distance weighted methods are based on the assumption that the interpolating surface should be influenced most by the nearby plots or grids and less by the more distant plots or grids
Due to interpolation, the boundaries of the generated polygons are estimated taking into account each grid cell value and the distance between each cell. Thus, the result is a map in a raster format that can be used as a spatial representation. This map should be considered as a graphical representation to illustrate a phenomena.
For qualitative variable it is foreseen in this proposal to represent only presence/absence.

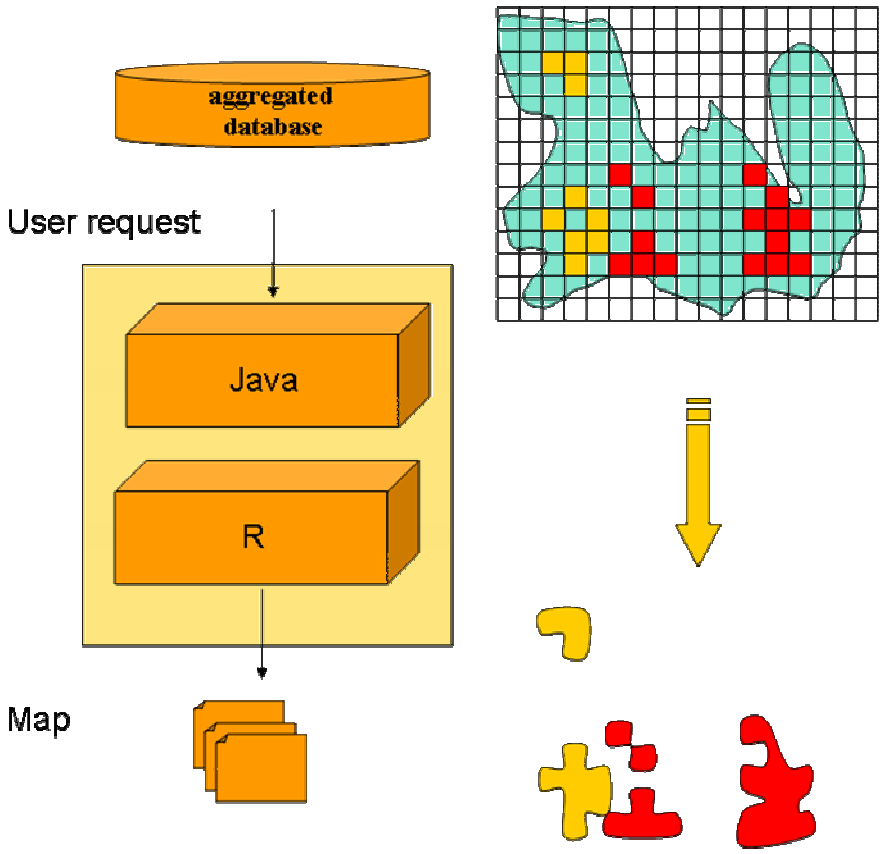


Figure 9: Spatial representation

The spatial representation process will be developed using R. R is a language and environment for statistical computing and graphics (http://www.R-project.org). Thus, R provides all the necessary algorithms to implement such transformations.

3.4.4 Data dissemination

Note: The data dissemination module is postponed to further proposal, the following chapter is kept for information only. A simple web page (web mapping view) allowing to show the result of the aggregation module will be developed to compensate for the absence of this module.

3.4.4.1 Reference definition and metadata consultation

A consultation module is developed to disseminate the analytical reference definitions and national definitions. This module is accessible to all the authenticate users. This module consists in a user friendly interface to carry out requests (forms) and access to the generated results (tables).

3.4.4.2 Dissemination to EFICP

The EFICP has been developed to provide a system for internet-based information and communication in the European forest sector and will allow for the search, access, view, analysis, and download of a wide range of forest related information, including geographical and statistical information.
The EFICP has been designed to exchange data with a wide range of external systems and services, adopting INSPIRE guidelines and implementing rules for the geospatial data exchanges. EFICP also provide a SDMX client to exchange statistical data.
Under the EDFAC umbrella, the EFDAC E-forest platform will provide information to the EFICP. In fact, the EFDAC E-forest platform acts as an intermediary system between the data providers and the EFICP.
For the map dissemination (WMS), the EFDAC E-forest platform will use a mapserver.
If possible, the platform will also expose a WFS service.
For the statistical data, the JRC will provide a SDMX client. Our proposal considers that the SDMX client needs only to be configured in order to carry out a mapping between the aggregated database and the data requested by the EFICP. Our proposal does not include the implementation of the SDMX protocol.
Based on the metadata and a web form, the administrator of the EFDAC E-forest platform can generate a file containing the metadata attached to a map available in the mapserver. This file conformed to the ISO19115 - ISO19139 standard will be uploaded manually into the EFICP system.

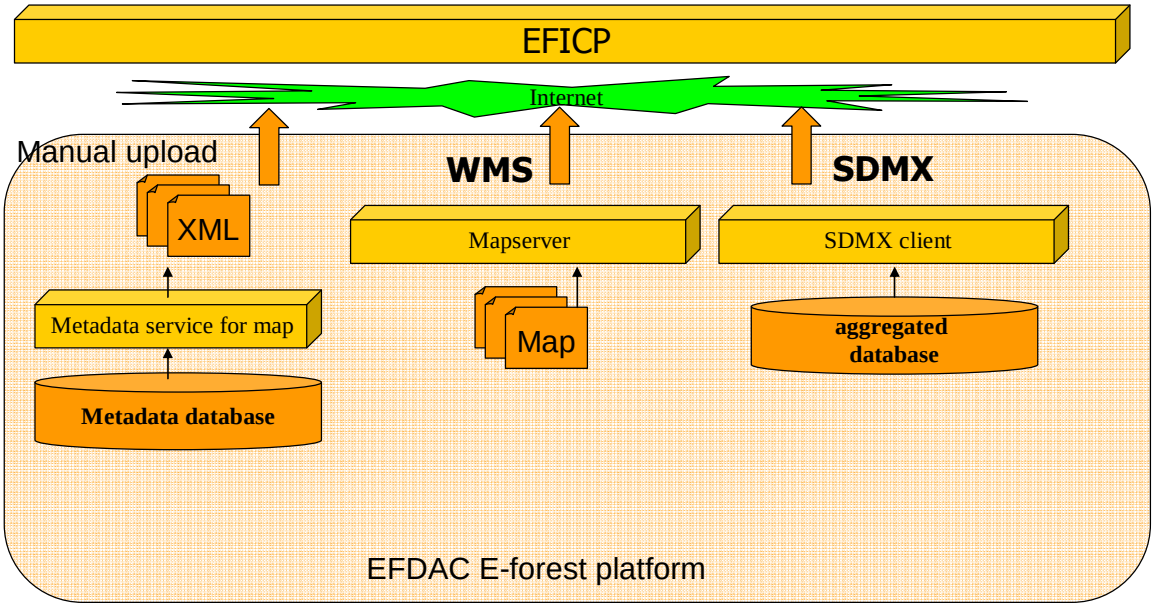


Figure 10: Data dissemination to EFICP

3.4.5 Data query builder enrichment

The data query builder module is adapted to take into account the new databases.
Thus it will be possible to carry out queries in all the databases (raw database, harmonized database and the aggregated database)
This module will be adapted to offer the possibility for data provider to validate the end of the harmonization process. The validation will allow the data dissemination
The map generated by spatial representation will also be available in this module.

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

Function (F & L): SENIOR RESEARCHER AND PROJECT MANAGER AT THE DEPARTMENT OF FORESTRY AND WOOD PRODUCTS (DANISH NFI)°
Name: Annemarie BASTRUP-BIRK
Task in the project: thematic scientific officer
Professional Experience: 22 years
Task in the project: review and link with SC3

Function ():
Name:
Professional Experience:
Task in the project: Metadata database design or review, bridge design or review

4.1.3 Other experts

Function (NFI): HEAD OF THE GIS DEPARTMENT
Name: Marianne Duprez
Professional Experience: 10 years
Task in the project: Expertise in GIS, Web mapping, Spatial representation

4.1.4 Project managers

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

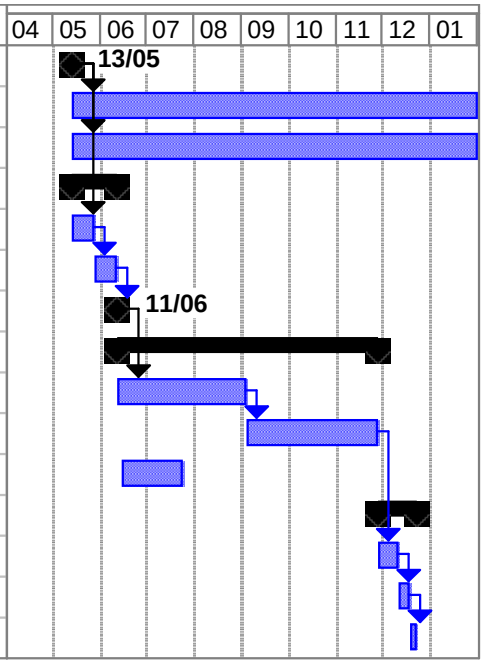
4.1.5 Project team

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 Time frame

N°	Nom de la tâche
1	kick-off meeting
2	System maintenance
3	Project follow up
4	**Design phase**
5	System design
6	Database design
7	JRC validation
8	**Realisation phase**
9	Agregation module
10	Spatial representation
11	Data query builder update
12	**Test phase**
13	Integration
14	Internal tests
15	External tests



The project duration corresponds to **one year** after the contract notification and signature. This Specific Contract suppose that the SC1 is partially developed (at least the data gathering module) and operational.

4.3 Financial proposal

SC2				
Price component	**Personnel type equivalence**	**Price per day**	**Total days**	**Total price**
Labour				
Project manager	Project manager	850	5	4250
FNFI senior project manager	Thematic scientific officer and data analyst	750	15	11250
FNFI project manager	Thematic scientific officer and data analyst	750	24	18000
FNFI system administrator	Programmer	400	5	2000
FNFI developper	Programmer	400	152	60800
FNFI Gis operator	GIS operator	400	20	8000
Expert	Thematic scientific officer and data analyst	750	7	5250
Subtotal Labour (1)				109550
Other				
Subtotal Other (2)				0
TOTAL (1)+(2) = PRICE				**109550**

This proposal is valid until the 30th of June 2009