EFORWOOD

Sustainability Impact Assessment of the Forestry - Wood Chain

Project no. 518128

EFORWOOD

Tools for Sustainability Impact Assessment

Instrument: IP

Thematic Priority: 6.3 Global Change and Ecosystems

Deliverable D1.2.6
D1.2.6 Report on data quality

Due date of deliverable: Month 47
Actual submission date: Month 54

Start date of project: 011105
Duration: 4 years

Organization name of lead contractor for this deliverable:
Johann Heinrich von Thünen-Institut (vTI), European Forest Institute (EFI)

Final version

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)

Dissemination Level

PU	Public	X
PP	Restricted to other program participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

EFORWOOD
Sustainability Impact Assessment of the Forestry - Wood Chain

Deliverable D1.2.6
D1.2.6 Report on data quality

AUTHORS:

Holger Weimar (vTI)
Jörg Schweinle (vTI)
Wendelin Werhahn-Mees (EFI)
Diana Vötter (EFI)
Martina Roubalova (IFER)

Johann Heinrich von Thünen-Institut, vTI, Hamburg, Germany
European Forest Institute, EFI, Joensuu, Finland
Institute of Forest Ecosystem Research, IFER, Jílové u Prahy, Czech Republic

Acronyms

BWCS	Baden-Wuerttemberg Case Study
C	Carbon
cf	Conversion factor
DBC	Database Client
DCP	Data collection protocol
EU FWC	European Forestry Wood Chain
FWC	Forestry Wood Chain
GUI	Graphical User Interface
ha	Hectare
IBCS	Iberian Case Study
ID	Identification number
M2, M3, M4, M5	Modules 2, 3, 4 and 5 in EFORWOOD project
SCCS	Scandinavian Case Study
SIA	Sustainability Impact Assessment
ToSIA	Tool for Sustainability Impact Assessment

Executive Summary

Quality, validity and reliability of the Sustainability Impact Assessment (SIA) of the forestry wood chain (FWC) very much depends on the quality and validity of the data the SIA is based on. This report describes the concept of data quality control and validation of the ToSIA database. Since data requirements of ToSIA and the EFORWOOD concept for data supply and provision are the basis for data quality control and validation, a comprehensive description is given of the Data Collection Protocol, data sources, structure of the ToSIA database and the data import routines.

Data quality control and validation in EFORWOOD is based on checks of completeness and consistency of sustainability indicator values as well as tests of consistency of the data relevant for the calculation of material flows by ToSIA. The completeness check is performed in order to safeguard that all data needed for a complete SIA have been entered into the ToSIA database. The consistency check however is performed by two checking routines. First, in order to identify incorrect data entries constraints that need to be met by each indicator and its subindicators were defined and executed. Second, for indicator values that go below or exceed a probable range, thresholds that should not be exceeded were defined. Indicator values beyond the thresholds were reviewed and if necessary corrected.

Since the SIA is performed for the FWCs of three European regions and the EU as a whole, the correct computation of material flows within the FWCs in the different regions is crucial. Hence the product shares, conversion factors and split ratios (one-to-many and many-to-one), which are needed to compute the material flows by ToSIA, were checked for missing values as well as by definition of constraints.

The checking routines were conducted and sent back to the responsible data providers several times in order to recheck the marked data and correct the values if necessary. With these tasks of data quality control and validation the data of the FWCs in EFORWOOD could be verified or, if necessary, improved significantly.

Key words: data validation, indicator values, conversion factors, split ratios, product shares, ToSIA, Database Client

Table of Contents

1. Introduction

Within the EFORWOOD project the software tool ToSIA has been developed to be able to analyse sustainability impacts of the forest-based sector. The processing of ToSIA is based on data regarding sustainability impacts of processes and products of the forestry-wood chain (FWC). The result of ToSIA is strongly influenced by the reliability and completeness of data, which is used by ToSIA. As a consequence, data quality control and validation is an important part of the data collection within EFORWOOD. The purpose of this deliverable is to describe the measures that were taken to safeguard quality, validity and reliability of SIA of the FWC.

The data is provided by experts of the EFORWOOD project and collected in a database using the EFORWOOD Database Client. Data sources are e.g. official statistics, research data, modeling results. This constitutes a challenge to ensure that interpretation of data collection guidelines, as well as assumptions and calculation routines are consistent. Basically, three different types of data can be distinguished: Firstly, data that define the topology of the Forestry Wood Chain (FWC); secondly, data which refer to the sustainability indicators, and thirdly, data related to the material flow.

Apart from the test chains with a limited number of processes at an early stage of the project, three different Case Studies and the European ~~Forestry-Wood Chain~~ Forestry Wood Chain (EU FWC) were developed. The Case Studies of the EFORWOOD project are; (i) the Baden-Württemberg Case Study with 153 processes, 73 products and about 36 000 indicator values collected, (ii) the Scandinavian Case Study with 158 processes, 55 products and approximately 31 000 indicator values collected and (iii) the Iberian Case Study where 82 process and 150 products were defined, and about 18 000 indicator values entered in the database. However, by far the most complex FWC is the EU FWC. 10 157 processes and 488 products were defined, and about 6 211 800 indicator values reported for two time steps (2015, 2025) in two different reference futures (A1[1] and B2[2]) and three different Natura2000 scenario levels[3] on top of both reference futures.

To be able to provide consistent data, testing routines for the validation of indicator values as well as material flow calculation have been developed to check inconsistencies in the reported data. The results of these routines were sent to the responsible contact person. The contact persons double checked their data entries and reported back to the data validation team. After the contacted project partners replied, the procedure was repeated again to ensure that the previously detected inconsistencies were solved.

[1] A1: globalisation, high economic growth, low environmental awareness
[2] B2: regionalisation, moderate economic growth, more oriented towards environment
[3] Increase in protected area by: level 0: no change, level 1: 10%, level 2: 15%, level 3: 25%

IFER in collaboration with the European Forest Institute (EFI) mainly carried out the data completeness checking. The European Forest Institute verified the data related to the material flow calculation. The Johann Heinrich von Thünen-Institut (vTI) conducted the validation of the relative indicator values.

2. Data requirements of ToSIA

The Sustainability Impact Assessment (SIA) of the forest-based sector in EFORWOOD builds on a conceptual representation of FWCs as chains of value adding processes (Lindner et al., 2010). In ToSIA a FWC can be described as a dynamic structure linking production processes with input and output products. The structure of a FWC is created based on specific information regarding the topology of the chain. Basic components of a FWC are processes, products and shares. In a process, the material and/or energy is transformed according to its attributes. The products are mass-based inputs and outputs of the processes. Furthermore, via the products the processes are linked. The number of the input products as well as the output products of a process varies according to the specific requirements of the process. Input and output product shares define the material flow along the FWC. Figure 1 illustrates the basic structure of the topology.

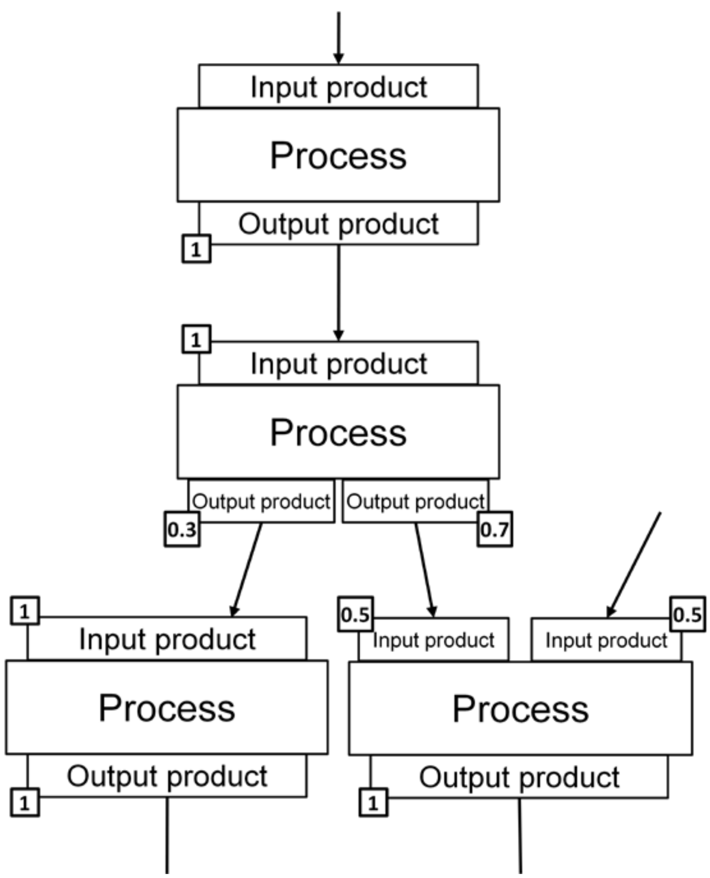

Figure 1. FWC topology; each process has at least one input and output product, an output product of a process is the input product of the successor process, product shares indicate how the material flow is distributed

As described above the sustainability impact assessment (SIA) of ToSIA is based on three basic types of data: data related to the chain topology, data to calculate the material flow, and indicator data. In Table 1 the different types of data are listed.

Table 1. Types of data relevant for SIA

Data types	**Attributes**	**Comment**
Chain topology data	Process	Defined by ID, name, country, module, stage, reporting unit, definitions, assumptions
	Product	ID, name, unit
	Links	Linkage from product to product
Material flow data	Product shares (in-, output)	Divides the material flow (in carbon) of each process into the input or output products
	Conversion factors	Convert the material flow in different units
	Split ratios	Divide one product to many (output) or merge many products to one (input)
Indicator data	Indicator	Indicator values are given always as indicator unit per unit of material flow, with information on their representativeness, additional notes, used algorithm, data source

The topology of a specific FWC is designed according to the standards of process modeling in the chain editor of the Database Client (DBC). The DBC is described in detail in D1.2.5 and in the Client Manual. These standards include the description of a process as a sequence of operations or events, which possibly absorb time, space, expertise or other physical and/or non-physical resources. A process consists of one or more input products and of one or more output products. The output will be different from the input according to the specific processing. Each product and process has an individual identification number (process ID or product ID). The products and processes can be easily reused in other FWCs. If such an ID is re-used in one or several chains, all characteristics and data are the same, and any change of data in one place will also change the data in the respective other places where the ID is used also.

The collected data are stored in the EFORWOOD database. The database can be accessed using the online application Database Client (Figure 2). The DBC structures the data in XML-files. For the processing of ToSIA two XML files, a chain-file and a process–file, have to be exported from the DBC and uploaded by ToSIA. The chain-file contains the information regarding the topology of the chain, i.e. processes, products, links between the processes and the split ratios. The process-file carries information of the processes, i.e. indicator values, conversion factors, product shares, process and indicator definitions.

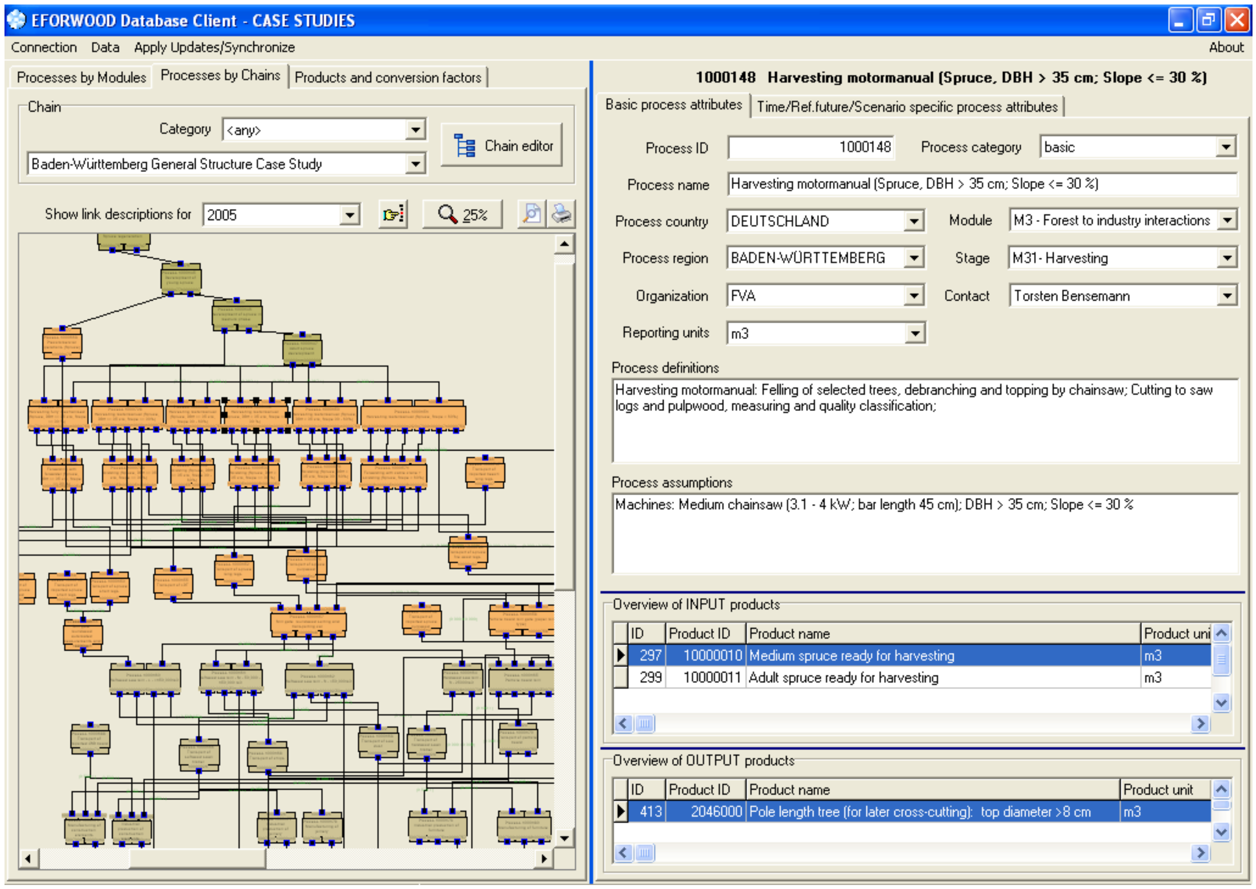

Figure 2. The interface of the EFORWOOD Database Client; the topology of the selected FWC is illustrated on the left side, the data given for the selected processes is illustrated on the right side

When the data has been exported from the DBC, both files are browsed and loaded one after another into ToSIA. When loading the data, ToSIA builds up the topology of the chain. The order and linkages between the processes/products are created. When the loading is complete, ToSIA indicates if the loading was successful. If ToSIA was not able to load the data, the user will be informed with a red error. Otherwise a green message appears stating: "The loading was successful". Errors when loading the XML-files are often caused by:

- erroneous topology e.g. linking wrong products, unlinked processes, or
- missing process data e.g. process region, reporting unit and conversion factors.

The errors have to be corrected and the process of loading has to be repeated. Once the loading was successful ToSIA automatically generates files, giving additional information on the data quality. Detailed explanations on these files are given in chapter 4.3 of this document.

3. Data supply and provision

Due to the complexity of the EFORWOOD project, multiple data sources were used e.g. statistics, models and expert guesses. Numerous experts collected data for the different parts of the forest value chain (modules) and the sustainability indicators. All data was gathered in the online Database Client. Two different methods of entering data into the data base can be distinguished:

(i) manual data entry via the Database Client (DBC), or
(ii) data file import.

Figure 3 gives an overview on the general data flow in the EFORWOOD project.

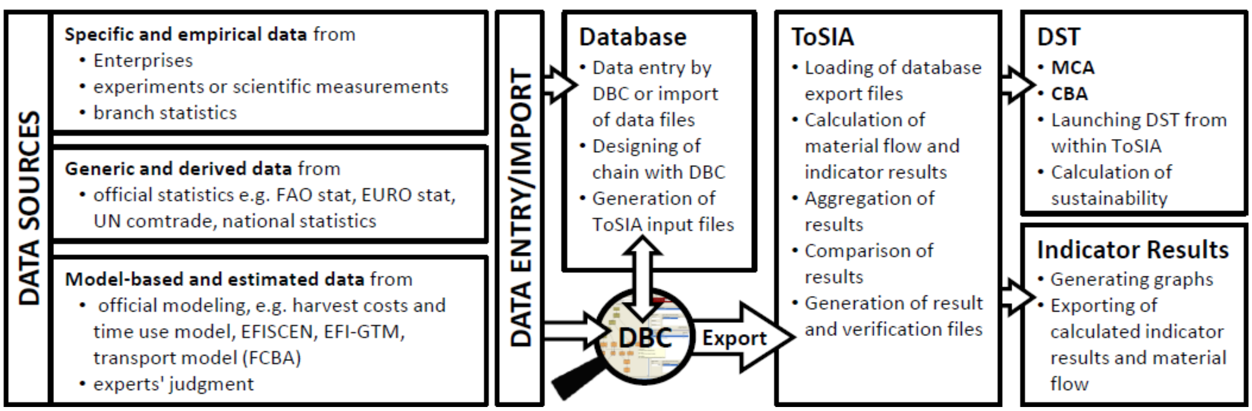

Figure 3. General data flow in EFORWOOD The data are derived from multiple sources and imported into the database using the Database Client (DBC) or directly via the hosts of the database. From the DBC the data is imported into ToSIA, the calculated indicator and material flow results can be either analyzed directly (indicator results) or further processed by using two different Decision Support Tools (DST) integrated into ToSIA - the Multi-criteria analysis (MCA) or the Cost-benefit analysis (CBA).

The development of the DBC in the EFORWOOD project had two main reasons; firstly the large number of data providers, and secondly a tool was needed to design and visualize the FWCs. Especially in the EU FWC, where partners from all over Europe and various experts for the different parts of the FWC entered data simultaneously into the data base, an online application was a necessity. The chain editor in the DBC has several functionalities: It enables to construct the chain, define processes and products and connect products by links. Without the visualization it is very difficult to imagine how the processes are linked and how the material flows are. Furthermore, the identification of a problem of the topology of the FWC or of process data is much easier.

3.1. Data collection protocol

The extensive data collection for the Sustainability Impact Assessment (SIA) of the forestry wood chain is based on the so-called Data Collection Protocol (DCP) (PD0.0.16). For the set of EFORWOOD sustainability indicators, selected by the members of the EFORWOOD consortium as well as external stakeholders, the DCP gives detailed information about the data needed for the sustainability assessment of the FWC.

The EFORWOOD sustainability indicator set is based on the Impact Assessment guidelines of the European Commission, the Sustainable Development Indicators of the European Union (2005), the Indicators of Sustainable Development of the Commission on Sustainable Development of the United Nations (CSD) (2006), the Improved Pan-European Indicators for Sustainable Forest Management of the Ministerial Conference on the Protection of Forests in Europe (MCPFE) (2002) and the European Union Rural Indicators (PAIS). These different sets were reviewed with respect to the needs of SIA of the FWC. Based on the review the EFORWOOD set was developed in an iterative process involving stakeholders and the EFORWOOD consortium. Indicator relevance and scale, data availability and spatial scale, technical feasibility and scale as well as cost of indicator application were the four criteria for indicator selection of the first draft set of indicators. In a stepwise process the first draft was revised and refined until the EFORWOOD consortium agreed on a final set (D1.1.1).

This final set consists of indicators that refer to the whole forestry wood chain and of indicators for individual parts of it. A subset of fourteen indicators of the general indicator list addresses issues especially relevant for policy-makers and the public. This subset is referred to as LEAD+ indicators. The general indicators list contains of eight indicators addressing economic, seven indicators addressing social and seven indicators addressing environmental sustainability. Five indicators specific for sustainability impact assessment of forests complete the list of the EFORWOOD sustainability indicators.

Table 2. Set of EFORWOOD sustainability indicators (PD 0.0.16)

Economic	Gross value added Production cost Trade balance Resource use Investment, research & development Total production Productivity Innovation
Social	Employment Wages and salaries

	Occupational safety and health Education and training Corporate social responsibility Quality of employment Consumer behavior and attitudes
Ecological	Energy generation and use Greenhouse gas emissions and carbon stock Transport Water use Soil condition Water and air pollution Generation of waste
Forest specific	Forest sector enterprise structure Provision of public forest services Forest resources Forest biodiversity Forest damage

The list is a balanced approach to assess economical, social and ecological aspects of sustainability. However, in order to do the SIA, a comprehensive and precise manual for the collection of indicator values is required. But the DCP is not only about this. It is a blue print for the database client as well as guideline for the data validation. Hence, the idea behind the DCP is to give guidance to data collecting experts and to streamline the whole data collection. Data collectors are able to find answers to their frequently asked questions regarding indicator definitions, measurement units, system boundaries, data sources and means to procure and calculate values on indicators, module specific recommendations and key definitions. Furthermore, experts for each indicator are listed as contact persons in the DCP to ensure further support of the data collectors. The information is provided in a common format for each indicator. This common format is structured as follows:

1. Name of the indicator
2. Name of subclasses of the indicator

   Indicators are subdivided into subclasses. Each subclass describes and quantifies a certain aspect of the indicator (e.g. the indicator employment is subdivided into the subclasses “number of person’s employment”, “classified by gender” and “employment on enterprise sites”.

3. Measurement unit

   All indicators and subclasses have defined measurement units e.g. the measurement unit for the production cost indicator is €.

4. Reporting unit

The indicators are calculated per unit of input material flow (the so called reporting unit). In ToSIA, indicators are linked to the material input flow of the process in the selected FWC to calculate the indicator value. E.g., the production cost indicator (subclass labour cost) is calculated for the process “transportation of pellets to home scale use”. The input material flow to this process is tons of pellets. The measurement unit of the indicator is €. Hence, the labour cost of transportation is 2.7 €/ton of pellets. The reporting unit in our example is tons of pellets.

5. System boundary

   The system boundary defines the data to be collected for each indicator regarding technology, space and time. ~~E. g.~~ For example indicator “energy generation and use” has a wide technical system boundary. All indicator values shall cover all renewable energy that is generated in a process and all (renewable and non-renewable) energy that is consumed in a process. Additionally, the energy use of the supply chains (e.g. exploration, transport of fossil energy carriers like oil, natural and coal) to the FWC is also accounted for. In contrast, the indicator “greenhouse gas emissions and carbon stock” has a narrow technical system boundary, since greenhouse gas emissions of the production of machinery; ancillary materials (lubricants) and chemicals used and consumed in the FWC are not accounted for.

6. Possible data source

   In order to accelerate the data collection process known data sources are provided for each indicator and its subclasses.

7. Calculation mode and conversion factors

   In case the information to be uploaded to the database is not ready available in the required format, some calculation routines as well as conversion factors are provided. Conversion factors are used in ToSIA for conversions between different products and units. ~~E. g.~~ For exampleconversion factors have to be available for the conversion of round wood into sawn wood into particleboard into a chair, etc and vice versa. Conversion factors are as well needed to convert product units (t dry mater, m^3, etc.) into reporting units of the processes, to internal reference units (tons of carbon) or to €.

8. Module specifications and recommendations

   In case there are specific recommendations regarding a certain module of the forestry wood chain, they can be found here.

9. Key definitions

   For terms that need to be explained in order to ensure input of correct data into the database, a comprehensive definition is provided e.g. for the term renewable energies.

10. Space for examples

   Finally, detailed explanations are given on how to derive, calculate, or estimate indicator values. Depending on the complexity of the topic, the examples given are rather extensive. During the data collection process it turned out that this part of the DCP is the most important. Given examples are as precise as possible in order to minimize misconceptions and mistakes during data collection and indicator value calculation.

11. Expert contact

    For each indicator and its subclasses a person to be contacted in case of questions is provided in the DCP.

As mentioned before the DCP is not only the backbone of data collection but blueprint for the data client as well. The data client's user interface matches nomenclature and requirements regarding units and conversion factors of the EFORWOOD sustainability indicators. Data collectors are facing the familiar DCP structure when entering indicator values into the database.

3.2. *Data sources*

Many times there are data at hand, but the issue is rather to exercise priorities in procuring them. This is why the DCP provides possible data sources for the indicators to be quantified. In general there are three different types of data available (Table 3).

Table 3. The data collected in the EFORWOOD project derived from various sources, three data types can be distinguished

Type	Source
Specific and empirical	- follow up routines from enterprises - data from experiments or scientific measurements - branch statistics
Generic and derived	- national official statistics - international statistical databases, e. g. EUROSTAT, FAOSTAT, UN Comtrade, UNECE - weighting or scaling factors relevant for adaption of generic data to specific data for the actual case, e.g. average data of costs per cutting form (final felling/thinning) is adapted to the case in question with the aid of case specific shares of cutting forms
Model-based and estimated	- modeling, e.g. harvest costs and time use model, EFISCEN, EFI-GTM, transport model - experts' judgment

In order to ensure a transparent documentation of the data source and data type provided, the Database Client requires a specification of the data source and data type when entering data into the database. All information provided in the database concerning data sources and data types is handed over as meta-data to ToSIA. This is to inform users of ToSIA about the sources and types of data the SIA is performed with.

3.3. Basic structure of the database

The structure of the EFORWOOD database reflects informational content and logical relationships as formulated by respective EFORWOOD modules. The general FWC as defined by EFORWOOD is structured into four hierarchical levels as illustrated in Figure 4.

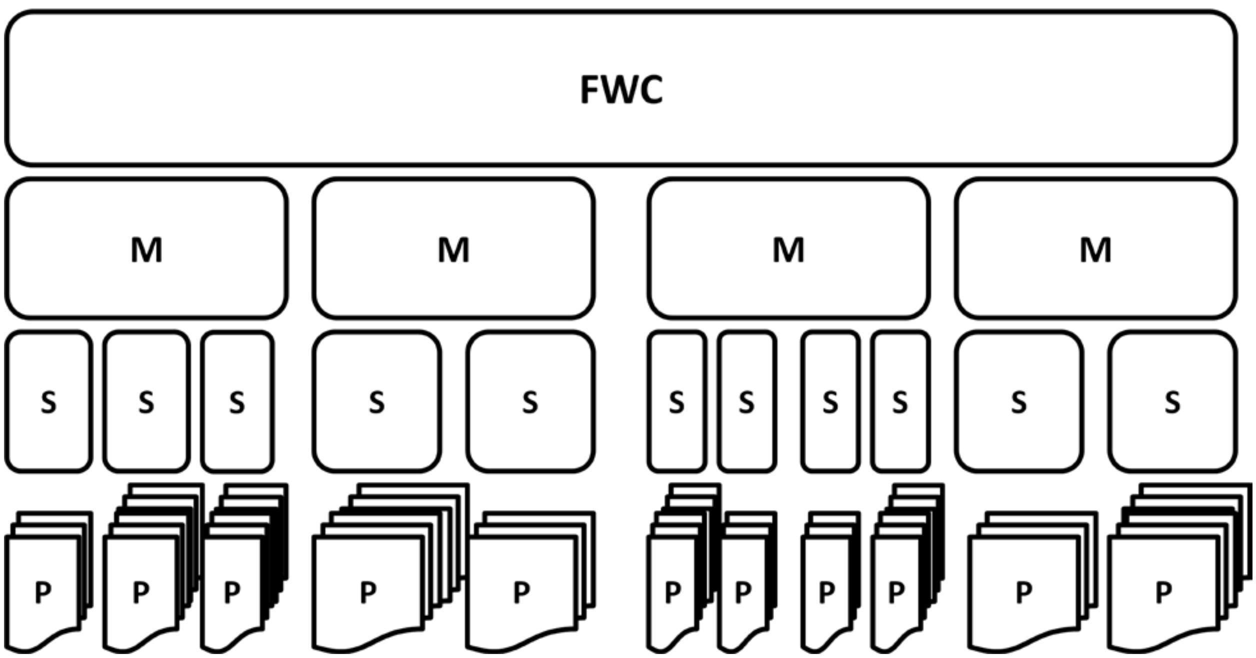

Figure 4. The four hierarchical levels are: the Forestry-Wood Chain (FWC), the four modules (M): Forest management, Forest to industry interaction, Manufacturing, Industry to consumer interaction, the stages (S) and the processes (P), depending on the FWC the number of stages and process per module varies

First level is a FWC itself. Second level breaks the chain into four separate modules: Forest Management, Forest to industry interaction, Manufacturing and processing and the Industry to consumer interaction. Every module consists of several stages, which are natural steps in a FWC flow. Processes represent the fourth and most detailed level of the database structure, e.g. timber harvesting with harvester or sawmilling.

4. Consistency of data

The data validation efforts for the data collected in the EFORWOOD project can be distinguished into three separate methods:

- The data pre-checking – inconsistencies detected when importing the data (no systematic validation routine)
- The completeness check – check if indicator data is missing
- The primary validation – validation of indicator data
- The secondary validation – validation of data related to the material flow

Figure 5 illustrates the data flow and the different steps of data checking and validation.

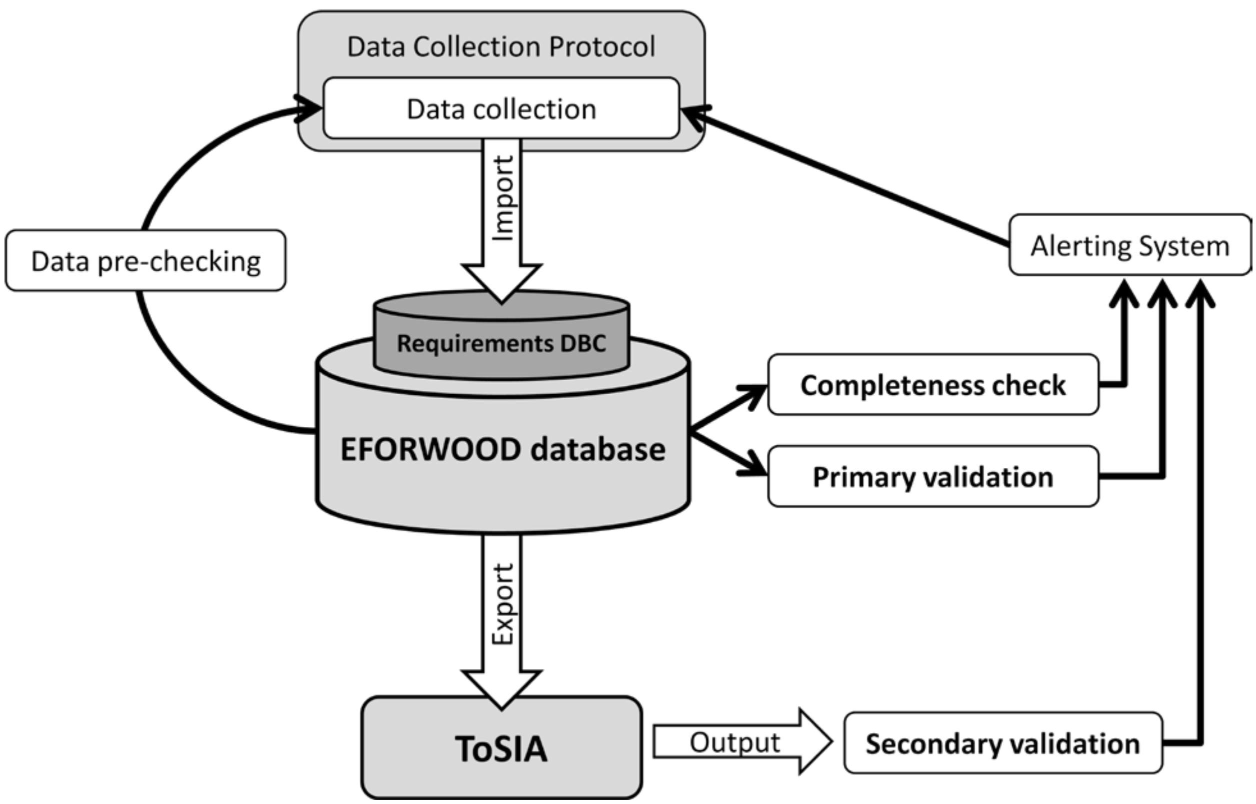

Figure 5. The EFORWOOD database stores the information collected by the partners with the clear guideline of the Data Collection Protocol (DCP). The database provides the necessary input data to run ToSIA. Data quality checks are done in several steps: the first data check is performed when data are submitted to the database by the data provider (pre-checking). Then, both the completeness of the data and individual values (indicator data and material flow data) are checked, primary and secondary validation.

The data pre-checking was conducted by the data provider, when entering the data into the database and the data did not meet the requirements of DBC. The validation of the data was organized as an additional task within the project which should support the data collectors. The results or reports of the different steps of the data validation were sent back to the responsible data providers. The contacted partners were asked to give feedback as soon as the data had been rechecked. In the following chapters, the different methods of data validation are explained further.

4.1. Data completeness checking

For data completeness checking the total number of values that should be entered into the database is calculated as number of processes multiplied by the number of indicators multiplied by the number of relevant time steps / scenarios. As the processes used in each FWC are very diverse, not all indicators were applicable to all processes. Hence, those indicators not applicable to a certain process were flagged as “not applicable”, e.g. indicators specific to Module 2 are not applicable to processes in Modules 3, 4, and 5. Indicators “not applicable” to certain processes were not expected to have indicator values related to these processes in the database. Indicator values expected to be in the database but not provided until the end of data collection because of their un-availability to be obtained or other reasons were identified and flagged as “not feasible”. The main reason for missing values is the non-availability of data. Table 4 gives an overview on the number of indicator values delivered, “not applicable” and “not feasible”.

Table 4. Overview of collected main indicator values in EU FWC: Absolute numbers for 2005 and main indicator (excluding sub indicators)

	Absolute numbers			
Module	Delivered Values	Not applicable	Not feasible	Sum
M2	3 418	1 796		5 214
M3	6 303	12 365	142	18 810
M4	22 033	22 212	3 078	47 323
M5	17 533	53 326	1 810	72 669
Total	49 287	89 699	5 030	144 016
Total in %	34.2%	62.3%	3.5%	100.0%

4.2. *Primary validation - Consistency check of indicator data*

To achieve reliable results of ToSIA it is crucial that the indicator values gathered by a lot of different data providers are collected and modified in a similar, comparable and traceable way. Although different interpretations of the meaning of an indicator or a subindicator should be avoided by using the Data Collection Protocol (DCP), it might occur that some of the data is erroneous. Beside a different interpretation of an indicator definition, errors can also occur during data input into the data base client or data import from an external source to the database automatically. Albeit the cause of erroneous data, it leads to hardly interpretable and misleading results of ToSIA. In order to avoid or at least to reduce these kinds of sources of errors a consistency check of the indicator values of the EFORWOOD database has been developed.

For this check of the indicator values, automated testing routines were developed. In contrast to the plausibility check of the indicator values of the single chains (see D1.4.6) this could not be done manually as the quantity of data was too high.

For the primary validation the evaluation of the indicator data was subdivided in two parts. The first part of the evaluation consists of the definition of constraints of indicator values within one process. The second part can be described as a check for outliers of indicator values of one process in relation to a group of processes. The comparability of these processes within a process group mainly refers to a similarity of input and output products, which consequently requires similar processes.

The results of the testing routines were sent to the responsible partners who provided the data. It should be mentioned that especially the results of the outlier tests does not imply that the data is false. ~~E.g.~~ For example, a process, which has been tested in relation to other processes, might have specific attributes that lead to specific indicator values. However, albeit some of the results of the evaluation are not wrong and do not have to be corrected, the comparison of indicator values of different processes provided by different partner raised a higher awareness of these interactions. The testing routines of the regional case studies have been undertaken and sent out four times, the European case study has been tested and sent out twice, this due to limited time at the end of the project. However, it could be observed that the data providing partners took the testing routines of the Regional cases into account when inputting data for the European Case, as the number of results of the database of the European Case was relatively small. So, it can generally be stated that improvements or modifications of the data based on evaluation results affect consecutive evaluation efforts. In the following two subchapters the two parts of the evaluation of the indicator values will be discussed in more detail.

4.2.1. Constraints – validation within one process

The checking of constraints of indicator values was developed using relations between the different indicators and subindicators within one process. The constraints are either based on the logic of the indicator and/or subindicator structure or expert opinion. For the programming of the testing routine it was necessary to express the constraints as algorithms. The constraints are defined in a way that the result should match an expected value, exceed specific thresholds, or result below a given value. In Tables 5 – 7 , three examples are listed to explain this approach further.

Table 5. Example 1, matching expected values

	Indicator	Unit
Value 1	19.1.1 Greenhouse gas emissions from machinery	kg
Value 2	19.1.2 Greenhouse gas emissions from wood combustion	kg
Expected result (Value 3)	19.1 - Greenhouse gas emissions	kg
Constraint	Value 1 + Value 2 = Expected result (Value 3)	kg

In example 1 the expected value is represented by value of indicator 19.1 “Greenhouse gas emissions”. This expected value should be matched by the sum of the subindicators 19.1.1 and 19.1.2.

Table 6. Example 2, Indicator value should exceed a specific threshold (or comparison of two values)

	Indicator	Unit
Value 1	2.1.3 - Average cost - labour costs	EURO
Threshold (Value 2)	11.1 - Wages and salaries - total	EURO
Constraint	Value 1 > Threshold (Value 2)	EURO

The constraint in Table 6 was built on an economic relation. It says that average labour costs (indicator 2.1.3) must not be higher than wages and salaries (indicator 11.1) of employees.

Table 7. Example 3, indicator value should be below a given threshold

	Indicator	Unit
Value 1	24.1.1 - Water pollution - organic substances (biochemical oxygen demand, BOD)	kg BOD5

Threshold (Value 2)	1	kg BOD5
Constraint	Value 1 < Threshold (Value 2)	kg BOD5

In example 3 (Table 7) it was assumed that the value of indicator 24.1.1 "Water pollution - organic substances (biochemical oxygen demand)" should not be higher than 1 kg BOD5 per reporting unit. For some processes, especially for production processes of the pulp and paper industry, this indicator might exceed the given threshold. If these special cases occurred, the results of the constraints had to be checked carefully before sending to the responsible data providers.

4.2.2. Outlier – comparison of several processes

Besides testing of constraints, which focused on the relation of relative indicator values within one process, the identification of extreme values aimed at a comparison of indicator values of several processes. For this purpose the processes of the regional cases on the one hand and of the EU FWC on the other hand were stratified in a first step according to the structure of modules and stages of the database and, in a second step by logical context, as in one stage the processes often were still too heterogeneous. One aspect for the stratification was the similarity of input and output products of the processes. Another aspect was the position within the FWC, which should be at a comparable same level (e.g. module or stage). As the objective of the three regional cases and hence the topology is different, the stratification led to a grouping with quite heterogeneous processes. In contrast, the grouping of the processes in the EUFWC was much more obvious to arrange as the European chain is based on 27 countries with a similar structure.

Additionally to a grouping of similar or comparable processes for the outlier test, the indicator values also had to refer to the same reporting unit but not to absolute units (e.g. hectares) or relative units (e.g. percentage). Using the description of the measurement units in the Data Collection Protocol the applicable indicators were selected. A list of the indicators and subindicators, which have been chosen for demonstration and which could be used for the outlier test, is stated in the Annex. Based on this list the majority of the indicator values could be used for outlier testing. Table 8 illustrates, that more than two third of the indicators were tested. In the Iberian Case Study this value amounts to more than 80% of the demonstration indicators.

Table 8. Total number of indicators and share of tested indicators in the different case studies

Number of…	BWCS	SCCS	IBCS	EU FWC
… indicators and subindicators in chain in total	89	89	67	73

… indicators and subindicators used for outlier testing	61	63	55	49
Share of tested indicators and subindicators	68,5%	70,1%	82,1%	67,1%

For the identification of outlier within a process group thresholds were defined which should not be exceeded. The definition of the thresholds is based on the mean value of an indicator of a process group. The minimum threshold was set at 10 per cent of the average. The maximum threshold was defined as the fivefold of the average. All indicator values outside the defined range were listed in a result table and sent to the responsible partner. Figure 6 illustrates the result of an outlier check.

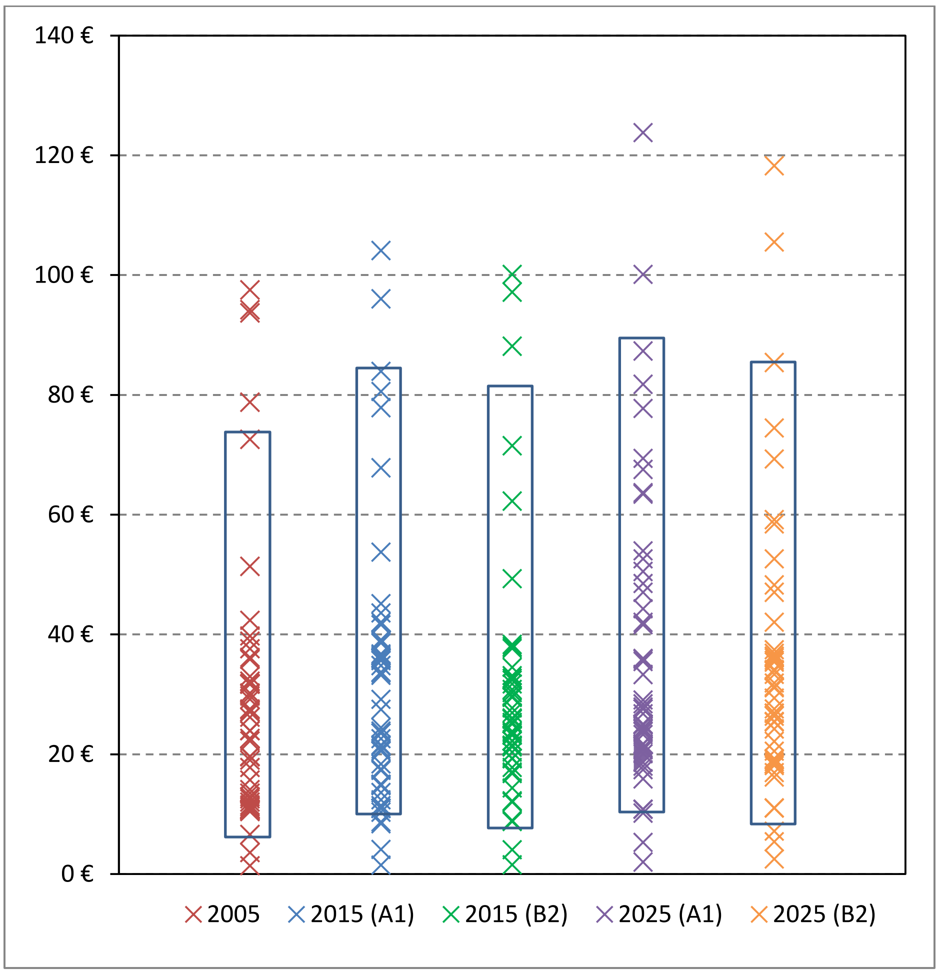

Figure 6. Schematic example of an outlier testing result for one indicator in a process group

The crosses represent the values distributed by the respective time validities. Based on these values an average was calculated for the different time validity. The boxes represent the range within the thresholds. Subsequently, the values outlying are defined as extreme values. The extreme values were reported back to the data providers with a request to check and if necessary to correct them.

4.3. Secondary validation - Consistency checking of material flow data

The material flow calculation of ToSIA is basically initialized by the indicators 22.1 “Forest and other Wooded Land Area“ and 3.1.1 “Import of wood and product derived from wood in the forest”. The data needed to calculate the material flow along the FWC in ToSIA are: product shares, conversion factors and split ratios (one-to-many and many-to-one). If one of them is missing or false, the material flow is wrong as well. Hence, consistency checking routines where implemented directly in the programming of ToSIA. The reports regarding the consistency of the material flow are generated automatically and saved as CSV-files in the same directory as the ToSIA program. The files can be imported into Excel. As shown in Table 9 the reports of the material flow consistency checking are automatically generated and named from ToSIA.

Table 9. Files exported from ToSIA

File	Automatic naming	Delimiter
Output product share	(name of FWC)_tosia_output_shares.csv	#
Split ratios	(name of FWC) _tosia_output_split_ratios.csv	#
ToSIA data report	(name of FWC) _tosia_data_report.csv	#
Postlog	(name given saving run) _tosia_postlog.csv	#

As the reports are created by ToSIA any user can use them for checking his/her indata and if basic modeling constraints (like all output products of a process sum up to maximal 100%) have been respected. During the EFORWOOD project, these consistency checks were first filtered by validity of the complaint and then sent to the responsible data providers, in order to correct the data. Subsequently, the different testing routines are explained in more detail.

4.3.1. ToSIA data report

ToSIA computes the validation of the figures needed for the flow calculations internally. Several testing routines are implemented. All errors or deviations of the established routines are reported in the “ToSIA data report”. This report has five different testing routines:

- Missing conversion factors

Whenever a needed conversion factor is missing, it will be reported. This accounts for the conversion factors from product unit to tons of carbon and from product unit to reporting unit if the units differ. The missing values are replaced by the default value "1".

- Missing split ratios
  If a split ratio (one-to-many) has not been entered, the missing value will be faked by ToSIA and set 1. In the consequence, the sum of the default value always exceeds one (see also chapter 2) and is included in the data report. Whenever the sum of the split ratios ≠ 1, it is always reported.
- Sum of the split ratios ≠ 1
  Whenever the sum of the split ratios is ≠ 1, it is included in the data report.
- Sum of the product shares ≠ 1
  Whenever the sum of the product shares of a process is ≠ 1, a report is written. In some cases the sum of the product shares can be less than one, e.g. if some process waste (such as dust emissions) is unaccounted for. A product share >1 is always a mistake, as the material flow would be enlarged.
- Sum of the output flows differs from 1
  This particular report indicates whenever the relative output flow of a process differs from one. The messages of this data reports should be seen as additional information on the material flow calculations. In some cases the messages indicate errors in the data – e.g. if split ratios have not been correctly entered into the database then the results may show that 100% of a product is distributed to each of several succeeding processes, resulting in an incorrect multiplication of the amount of that product. In some cases the messages indicate certain situations that can really occur in the chains – e.g. one output product (waste) ends in the chain. Therefore, the messages in the ToSIA data reports require interpretation by the ToSIA users.

The ToSIA data report is also displayed in ToSIA (data verification). The report is structured by the time validity (scenario), process name and ID, the name of the contact person (data provider) and last but not least the report itself. This information proved to be sufficient to find the error in the data set.

4.3.2. Input output checking

Every process has a certain input and output material flow, which can be stated in carbon or in the unit of the different products. In case no product ends somewhere in the FWC (e.g. waste which is not further processed or burned), the input flow in carbon should sum up to the same amount as the continuing output flow in carbon. Whenever products of a process are not linked or further processed, the output flow diminishes by the carbon amount of the product. Other factors could erroneously affect the carbon output flow of a process:

- *The conversion factors*. The flow calculation is very sensitive to variations of the conversion factors (cf) - a small relative change (in the cf value) can have a big quantitative impact.
- *The share of output products and/or product shares.* If these factors are adjusted other than one, the material flow is manipulated (in- or decreased).

A report is exported after each saved ToSIA run, stating how much carbon entered and how much left a process. This CSV-file is named Postlog (Table 9). On the basis of this report, errors manipulating the material flow can be detected.

4.3.3. Conversion factors

Conversion factors are a crucial part of the correct calculation of the material flow. They convert the different products along the FWC into comparable units. Most of those products are 100% wood based, a few products consist of a mixture of different materials as MDF boards or furniture. These products were not tested with this standard routine. Two variables influence the conversion factors of wood based products:

(i) the tree species or product (e.g. paper) and
(ii) the moisture content.

Each tree species has a specific dry weight. As a general approximation it was assumed in the EFORWOOD project that 50% of the dry weight of a product is carbon. Moisture content of wood products decreases from harvesting processes until use. On the basis of those assumptions constraints were developed to verify conversion factors inserted in the database and used in the material flow calculations.

Assumptions:

- 0.5 of the dry weight of the timber is carbon
- the weight in tons of 1 m^3 oven-dry wood product is twice the conversion factor (cf) to tons of carbon
- the conversion between the same units is always 1

Calculation of moisture content:

- weight of the product (P_w) minus the dry weight (P_d) (twice the carbon content) divided by the dry weight of the product multiplied by 100, equals the moisture content of the product in % (P_n)

Moisture content of P_n in %:

$$\frac{(P_w - P_d)}{P_d} * 100$$

Based on the constrains and assumptions explained above, an Excel template was developed and applied for all conversion factor sets of the four EFORWOOD forest value chains. The conversion factors per case study where downloaded from the EFORWOOD database and copied into the template. Then, conversion factors are verified according to the assumptions stated in this chapter. Thereof algorithms and data constrains were developed. All conversion factors deriving from the thresholds defined were highlighted red and marked with: “check”. According to the other testing routines, the results of the verifications were sent to the responsible partners for review. This procedure was repeated several times until all inconsistencies were solved.

5. Conclusion

Within the EFOWOOD project it was a major task to provide sound and qualitatively high level data. Starting with the development of the Data Collection Protocol (DCP) several tasks were implemented in order to fulfill the data requirements of ToSIA in an appropriate way.

Albeit the DCP gave guidance to the data collectors for the harmonized provision of indicator values, it could not foresee all possible questions to answer in this single document. As there was still room for interpretation a subsequent test of the provided data was necessary. These testing routines developed did not only improve the indicator data as such but also raised question to the interpretations of the DCP which in turn led to an improved guideline for the data collection.

For the provision of data regarding the material flow, very limited time was left in the project. Due to this, the consistency check of the material flow data gave guidance for the interpretation and improved the understanding of the material flow calculation in ToSIA.

All checking routines were conducted and sent back to the responsible data providers several times in order to recheck the marked data and correct the values if necessary. These tasks were necessary to assure the validity and reliability of the provided data.

6. References

Berg, S.; et al. (2008). EFORWOOD Deliverable PD0.0.16: Manual for data collection for Regional and European cases (2008). Background document for EFORWOOD Training, Working document for Task Force Final version 2 July 2008 – UPDATE 3 September 2008, Skogforsk, Uppsala.

Commission for Sustainable Development (2006). Indicators of Sustainable Development. (http://www.un.org/esa/sustdev/natlinfo/indicators/isdms2001/table_4.htm).

European Commission (2005). Sustainable Development Indicators for the European Union. Luxemburg: Eurostat.

IFER (2010). EFORWOOD Deliverable D.1.2.5: Database of case studies and EU FWC and summary report of database development, IFER, Jilove u Prahy.

IFER (2010). EFORWOOD Deliverable D.1.2.5 Annex: EFORWOOD Database Client, User Manual, IFER, Jilove u Prahy.

Lindner, M.; Suominen, T.; Palosuo, T.; Garcia-Gonzalo, J.; Verweij, P.; Zudin, S.; Päivinen, R. (2010): ToSIA – A Tool for Sustainability Impact Assessment of Forest-Wood-Chains." Ecological Modelling, in press.

Ministerial Conference on the Protection of Forests in Europe (2002). Improved Pan-European Indicators for Sustainable Forest Management. MCPFE Expert Level Meeting 7-8 October 2002, Vienna, Austria.

Rametsteiner, E.; Pülzl, H.; Puustjärvi, E. (2006). EFORWOOD Deliverable 1.1.1: Draft FWC indicator set: Detailed review of existing sustainability indicator concepts and sustainability indicator sets of relevance for the FWC, review of potential indicators for selection and their assessment, BOKU, Vienna.

Rametsteiner, E., Berg, S., Laurijssen, J., Le-Net, E., Lindner, M., Peuhkuri, L., Prokofiewa, I., Schweinle, J., Vötter, D., Carnus, J.-M., Edwards, D., Jactel, H., Raulund-Rasmussen, K. and Tomé, M. (2008). EFORWOOD Project Deliverable 1.1.6: Revised FWC-sustainability indicator set document, BOKU, Vienna.

Werhahn-Mees, W.; Vötter, D.; Suominen, T.; Lindner, M. (2010). EFORWOOD Deliverable D1.4.6: Documentation of ToSIA developments, EFI, Joensuu.

7. ANNEX

List of tested indicators for outlier check

Indicator	**Unit per reporting unit**
1.1 - Gross value added (at factor cost)	EURO
2.1 - Production cost	EURO
2.1.1 - Average cost - raw materials from FWC	EURO
2.1.2 - Average cost - raw materials from outside FWC	EURO
2.1.3 - Average cost - labour costs	EURO
2.1.4 - Average cost - energy costs	EURO
2.1.5 - Other productive costs	EURO
2.1.6 - Non-productive costs	EURO
6.1 - Investment (gross fixed capital formation) in total[2]	EURO
6.1.1 - machinery and equipment[2]	EURO
6.1.2 - vehicles[2]	EURO
10.1 - Employment - number of persons employed	number of persons in full time equivalent in reference year
11.1 - Wages and salaries - total	EURO
12.1 - Occupational accidents	absolute number
12.1.1 - Occupational accidents (non-fatal) - absolute numbers p.r.u.	absolute number
12.1.2 - Occupational accidents (fatal) - absolute numbers p.r.u.	absolute number
15.1 - Persons employed part-time and employees with a contract of limited duration (annual average) in total[1,2]	total number of employees
17.1. - Apparent consumption of wood per capita[3]	m3
18.1 - On-site energy generation from renewables	kWh
18.1.1.1 - On-site heat generation from renewables - residues from process - inputs	MJ
18.1.1.2 - On-site heat generation from renewables - other wood biomass	MJ
18.1.1.3 - On-site heat generation from renewables - non-wood based renewable heat	MJ
18.1.2.1 - On-site electicity generation from renewables - residues from process	kWh
18.1.2.2 - On-site electicity generation from renewables - other wood biomass	kWh

18.1.2.3 - On-site electicity generation from renewables - non-wood based renewable electicity	kWh
18.1.3.1 - On-site fuel generation from renewables excluding fuel used for mill site heat and electricity generation and excluding fuel that is used as a product further in the FW3 - residues from process	MJ
18.1.3.2 - On-site fuel generation from renewables excluding fuel used for mill site heat and electricity generation and excluding fuel that is used as a product further in the FW3 - other wood biomass	MJ
18.1.3.3 - On-site fuel generation from renewables excluding fuel used for mill site heat and electricity generation and excluding fuel that is used as a product further in the FW3 - Non-wood based renewable fuel production	MJ
18.2 - Energy use	kWh
18.2.1.1 - Energy use - Heat from renewable sources	MJ
18.2.1.2 - Energy use - Heat from fossil sources	MJ
18.2.2.1 - Energy use - Direct fuel use - renewable fuel	MJ
18.2.2.2 - Energy use - Direct fuel use - fossil fuel	MJ
18.2.3.1 - Electricity use - from 100% renewable sources	kWh
18.2.3.2 - Electricity use - from 100% fossil sources	kWh
18.2.3.3 - Electricity use - from the grid	kWh
19.1 - Greenhouse gas emissions	kg CO2 eq.
19.1.1 Greenhouse gas emissions from machinery	kg CO2 eq.
19.1.2 Greenhouse gas emissions from wood combustion	kg CO2 eq.
19.2 - Carbon stock	kg CO2 eq.
19.2.1 - Carbon stock in woody living biomass (above ground)	kg CO2 eq.
19.2.2 - Carbon stock in woody living biomass (below ground)	kg CO2 eq.
19.2.3 - Carbon stock in woody dead wood	kg CO2 eq.
19.2.4 - Carbon stock in soils of forest	kg CO2 eq.
21.1 - Water use (freshwater intake by industry) [relevant for industry][1,2,3]	m3
21.2 - Water use (of the forest ecosystem)[1,2]	m3
21.2.1 - Water use (of the forest ecosystem) - Evapotranspiration from the system[1,2]	m3
21.2.2 - Water use (of the forest ecosystem) - Groundwater recharge1[1,2]	m3
22.2.2 Growing stock on forests available for wood supply	m3
22.4.1 - Balance of increments and fellings: Net annual increment	m3
23.1.6 - site nutrient budget averaged over total rotation period (N, P, K, Ca, Mg)[1]	%
24.1.1 - Water pollution - organic substances (biochemical oxygen demand)[1,2,3]	kg BOD5
24.1.2 - Water pollution - nutrients (nitrogen, phosphorus) as Nitrogen or TKN (Total KJELDAHL Nitrogen)[1,2,3]	kg TKN
24.2.1 - Non-greenhouse gas emissions into air - CO[1,2,3]	kg
24.2.2 - Non-greenhouse gas emissions into air - NOx[1,2,3]	kg
24.2.3 - Non-greenhouse gas emissions into air - SO2[1,2,3]	kg
24.2.4 - Non-greenhouse gas emissions into air - NMVOC[1,2,3]	kg
25.2.1 Volume of standing deadwood[1,2,4]	m3
25.2.2 Volume of lying deadwood[1,2,4]	m3

27.1 - Generation of waste in total	kg
27.1.1 - Not classified as hazardous waste	kg
27.1.2 - Hazardous waste	kg
27.2.1 - Waste to material recycling	kg
27.2.2 - Waste to incineration	kg
27.2.3 - Waste to landfill	kg

[1]) Baden-Württemberg Case Study

[2]) Scandinavian Case Study

[3]) Iberian Case Study

[4]) European Forestry Wood Chain