



ANTTI SIRKKA

Managing Product Life Cycle Data
Using Automatic Identification



ACADEMIC DISSERTATION

To be presented, with the permission of
the board of the School of Information Sciences of the University of Tampere,
for public discussion in the Auditorium Pinni B 1096,
Kanslerinrinne 1, Tampere, on October 5th, 2012, at 12 o'clock.

UNIVERSITY OF TAMPERE

ACADEMIC DISSERTATION
University of Tampere
School of Information Sciences
Finland

Copyright ©2012 Tampere University Press and the author

Distribution
Bookshop TAJU
P.O. Box 617
33014 University of Tampere
Finland

Tel. +358 40 190 9800
Fax +358 3 3551 7685
taju@uta.fi
www.uta.fi/taju
<http://granum.uta.fi>

Cover design by
Mikko Reinikka

Acta Universitatis Tamperensis 1767
ISBN 978-951-44-8920-4 (print)
ISSN-L 1455-1616
ISSN 1455-1616

Acta Electronica Universitatis Tamperensis 1240
ISBN 978-951-44-8921-1 (pdf)
ISSN 1456-954X
<http://acta.uta.fi>

**Managing Product Life Cycle Data
Using Automatic Identification**

Supervisors: Associate Professor Marko Junkkari
University of Tampere

Professor Hannu Kangassalo
University of Tampere

Opponent: Professor Hannu Jaakkola
Tampere University of Technology

Reviewers: Professor Asko Riitahuhta
Tampere University of Technology

Professor Jaak Henno
Tallinn University of Technology

Acknowledgements

First, I want to thank my supervisors, Professor Hannu Kangassalo, who encouraged me to start this work, and Associate Professor Marko Junkkari for support and guidance throughout this work. Without their support this dissertation would never have been started and, most certainly, it would not exist today.

I am greatly indebted to my co-authors Marko Junkkari, Ville Puntanen, Anders Björk, Martin Erlandsson, Janne Häkli, Kaarle Jaakkola, Åsa Nilsson and Kaj Nummila. I also wish to thank Timo Niemi, Kati Iltanen and all the colleagues in the Data Management and Conceptual Modelling Seminar for their support during my research.

This study was carried out in the School of Information Sciences, University of Tampere, and in Tieto Finland, in 2006-2012. The School offered a great deal of support for writing the dissertation, and I wish to thank the entire personnel of the School. Tieto Finland has provided a good environment for working and writing the dissertation. I am grateful for the support provided by the entire company, with special thanks extended to Ville Puntanen, Jari Mäkipää, Teemu Hämäläinen and Juha Anttila for making this dissertation possible.

This work was supported financially by Tieto Finland, Tekes within the Cloud Software Program and the European Commission within the Sixth Framework Programme, contract No 34732. I would also like to thank all partners in these projects who have contributed to the work.

In addition, I wish to express my gratitude to my family and friends who have supported and encouraged me throughout my studies. Finally, and most importantly, loving thanks to Auli for everything.

Tampere, October 2012

Antti Sirkka

Abstract

Managing the life cycle of products is becoming more and more important. Organizations are facing increasing pressure from consumers and legislators to accurately measure and manage the environmental impact of products. However, the complexities of today's supply chains pose a challenge for gathering accurate data throughout the life cycle of the product

The life cycle of a product can be defined as a network of entities responsible for the procurement, manufacturing and distribution of the product. In order to enable tracing through the dynamic supply chain, the products must be identified. The development of automatic identification enables us to identify each object in the supply chain and trace it through the complex and dynamic supply chain where each organization manages a part of the chain. Thanks to traceability, we can connect the information about the products' movements with the information about processes. In other words, we can allocate the properties of the processes to the actual product instances involved in each process.

To be able to store the life cycle information of products, we must have a model that enables the allocation of life cycle information to the traced product throughout the supply chain. This dissertation defines such a model (*traceability graph*) that can be used to allocate life cycle information from processes to individual products. Further, the model enables multidimensional analyzes of data associated with the life cycle information of products and their components. The dissertation also specifies a solution for collecting, storing and sharing life cycle information about the product throughout its life cycle, enabling consumers to make educated choices based on accurate information regarding products they are purchasing. The method enables supply chain stakeholders to exchange life cycle information by utilizing the EPCGlobal Network architecture.

The case example used in this dissertation is environmental impact information. In recent times, consumers and legislators have become increasingly interested in the environmental impacts of products throughout their life cycle. The biggest challenge with measuring the environmental impact is the fact that supply chains are complex and dynamic. A manufacturer can use various subcontractors and supply various end manufacturers or retailers in different countries. So far, the most common method of calculating the environmental impact of a product has been to measure the resources used, emissions and production for a certain period of time and then calculate the average environmental impact of the product. This work provides methods to monitor environmental performance even at a product level.

List of used abbreviations

RFID	Radio frequency identification
OLTP	Online transaction processing
OLAP	Online analytical processing
MOLAP	Multidimensional online analytical processing
ROLAP	Relational online analytical processing
HOLAP	Hybrid online analytical processing
LCI	Life cycle inventory
LCA	Life cycle assessment
ELCD	European reference life cycle database
PDM	Product data management
PEFC	Programme for the endorsement of forest certification
PLM	Product lifecycle management
GHG	Greenhouse gas
WRI	The world resource institute
WBCSD	The world business council for sustainable development
GRI	The global reporting initiative
EPC	Electronic product code
EPCIS	EPC information services
ONS	Object name service
ERP	Enterprise resource planning
MES	Manufacturing execution system

List of Publications

This dissertation is based on the following original research papers:

Paper I:

Antti Sirkka. Modelling traceability in the forestry wood supply chain. ICDE Workshops. 104-105. 2008.

Paper II:

Marko Junkkari and Antti Sirkka. Formal Definition of Traceability Graph. Report 3. School of Information Sciences. University of Tampere. 2011.

Paper III:

Marko Junkkari and Antti Sirkka, Using RFID for tracing cumulated resources and emissions in supply chain, International Journal of Ad Hoc and Ubiquitous Computing, IJAHUC 8(4). 220-229. 2011.

Paper IV:

Antti Sirkka and Marko Junkkari, Multidimensional Analysis of Supply Chain Environmental Performance, Sustainable ICTs and Management Systems for Green Computing. IGI Global. 231-250. 2012.

Paper V:

Anders Björk, Martin Erlandsson, Janne Häkli b, Kaarle Jaakkola, Åsa Nilsson, Kaj Nummila, Ville Puntanen, Antti Sirkka*. Monitoring environmental performance of the forestry supply chain using RFID. Computers in Industry 62(8-9). 830-841. 2011.

*Corresponding Author

Contents

Acknowledgements.....	5
List of Publications	8
1. Introduction	10
2. Motivation / Background.....	14
2.1 Environmental Accounting	14
2.2 Location information.....	17
3. Workflows	18
3.1 Modeling Workflows	18
3.2 Data-Centric Workflows	21
3.3 Provenance in Workflows	23
4. Logical Tracing: traceability graph	24
5. Physical Tracing: Radio Frequency Identification	29
6. Storing and Analyzing Traceability Data	33
7. Sharing the Life Cycle Information.....	39
7.1 EPCglobal Architecture.....	39
7.2 Sharing Life Cycle Data.....	41
7.3 Taxonomication.....	44
8. Papers.....	46
9. Discussion.....	49
10. Conclusions.....	51
References.....	53

1. Introduction

Product lifecycle management (PLM) is a holistic concept that integrates products, processes, methods, models, applications and organizations which participate in environment where the product is developed, used and supported (Grieves 2006; Ameri & Dutta 2005; Stark 2011). This thesis presents a model and methods to collect and share product life cycle information between supply chain stakeholders. The presented system can be seen as a product data management (PDM) system. PDM systems control the basic data about a product and are used to integrate and manage supply chain processes (Stark 2011).

In general, a supply chain is an ecosystem of organizations involved in providing a product to consumers. Processes use raw materials to create finished products for consumers. In this thesis, the term 'supply chain' is used to specify the processes that participate, directly or indirectly, in the production and transfer of a product from suppliers to consumers.

A supply chain consists of processes following each other in a partial order. This means that the result of a process (output flow) is used as a raw material – or component – in another process (input flow). Processes have properties, such as resources and emissions, associated with the result products of processes. These properties are accumulated in a supply chain, i.e. from the data-oriented perspective, information on the preceding processes is associated with the products and process. For example, when calculating the energy used for a product, the route of the product must be taken into account.

In recent times, consumers and legislators have become increasingly interested in the environmental impact of products throughout their supply chain. The biggest challenge with measuring the environmental impact is the fact that supply chains are complex and dynamic. Moreover, the current environmental accounting systems are not sophisticated enough and they lack standards to exchange information in the supply chain.



Figure 1. Global supply chains.

Parts of products are acquired from various subcontractors and the distribution of the parts and products is managed by various logistics companies. Figure 1 illustrates this. In other words, environmental information should be shared among stakeholders in the supply chain. This also means that supply chains are dynamic, i.e. similar products in the store may feature quite different supply chains. For example, a chair can be manufactured locally or transported from another continent. Moreover, the manufacturer of the chair may have used local raw material or raw material transported from another continent.

So far no general methods for taking into account these kinds of variations have been presented. The environmental impact of a product is typically calculated by measuring the emissions, resource usage and production information on a yearly basis and using this information to calculate the yearly average impact of a product. For example, in the European Reference Life Cycle Database (ELCD), a process data set: *Spruce wood; timber; production mix, at saw mill* (PE International 2002) the average transport distance between the felling site and the saw mill is 144 km even though the transportation distance of logs can vary from a few kilometers to thousands of kilometers.

In order to trace the emissions generated throughout the whole life cycle of the product, we must be able to trace the product through its supply chain. The development of automatic identification (auto-id) enables us to identify each object in the supply chain and trace it through the complex and dynamic supply chain where each organization manages a part of the chain. There are numerous different marking methods for identifying products in a supply chain, such as imprinting, the finger print method, Laser marking, Label marking, Ink jet

marking and transponder marking. The usage of auto-id technologies for measuring the environmental burden of a product has been studied in recent years, for example, in (Data et. al, 2010; Data and Staake, 2008; Staake and Fleisch 2009). The dissertation studies the use of RFID in tracing products in the supply chain.

With traceability, we can connect the information about the products' movements with the information about processes. In other words, we can allocate the properties of the processes – for example, the environmental burden caused by the process – to the actual product instances involved in each process. This work describes a case where every item is marked with an individual tag which enables object-level traceability and life cycle information management. Tracing can be implemented using different frequencies. For example, in some cases the most feasible alternative would be to only mark the packages or containers of similar products with similar supply chains. It is also possible to mark only some of the products/patches and monitor the supply chain by means of estimates. The marking of individual products enables life cycle data management at item level as opposed to the company-level reporting that is predominantly used in reporting today.

The case study used in this dissertation has been selected from manufacturing and the forest industry and focuses on collecting life cycle information – more accurately, the environmental impact of a product - from a forestry wood supply chain. The model and system presented in this dissertation were tested in the Indisputable Key project funded under the European Commission's Sixth Framework Programme. In the research project, the trees were traced from the forest and through the saw mill. The data gathered in the project is used for analyzing the environmental impacts of the forestry wood supply chain with the aim of improving the efficiency of the supply chain (Häkli et. al., 2010).

In order to be able to trace the individual products through complex supply chains and to collect the related life cycle information from the supply chain processes, the following questions need to be answered:

- How to model the life cycle of an individual product that can be transformed during the supply chain processes?
- How to store the life cycle data and allocate the data to the specific objects?
- How to analyze the huge amount of life cycle data that is collected from the traceability system?
- What kind of infrastructure is needed to collect and share the life cycle data between supply chain stakeholders?

In this work, a workflow model – *traceability graph* - is developed (Junkkari and Sirkka, 2011; Paper II) that enables the allocation of life cycle data (e.g. environmental burden) to the traced product throughout the supply chain. The model has the ability to manipulate transformations of the products. A generic workflow model is developed, generalizing the data model for the traceability system of the forestry wood supply chain (Sirkka, 2008; Paper I).

The traceability graph is mapped to a relational data model in (Junkkari and Sirkka, 2011; Paper III) which allows storing the life cycle data associated with processes and products allocating the data to specific objects. The *traceability cube* (Sirkka and Junkkari, 2010; Sirkka and Junkkari, 2012; Paper IV) is a method of querying and analyzing the product life cycle information. Finally, a solution for sharing the life cycle information of a product between supply chain stakeholders is presented (Björk et. al., 2011; Paper V) using a case study.

Next, the motivation of this work is presented by introducing different usage cases of the traceability graph in product life-cycle data management. Then, a short survey on workflow models is presented. In Section 4 the traceability graph is presented in detail. Section 5 describes the usage of radio frequency identification in tracing products through supply chain. In Section 6 the methods for analyzing the life cycle information of products are presented. Section 7 describes how the traceability graph information can be managed and shared in supply chains using an EPCglobal network, which is a computer network that can be used to share product data among supply chain stakeholders. Finally, the papers comprising the dissertation are presented and the conclusions are given in the final sections.

2. Motivation / Background

The traceability graph is a model for tracing, analyzing and querying data. It gives a starting point for the provenance of the life cycle an object. Next, we introduce some practical usage cases and their background for the traceability graph. The examples illustrate how the item-level information of can be used to improve the efficiency of supply chains.

2.1 Environmental Accounting

The most common method used for measuring environmental performance in organizations is to calculate the total environmental impact of the whole company. The dominant methodologies used are the Greenhouse Gas (GHG) Protocol, which is a guideline for organizations to estimate their greenhouse gas emissions, and the Global Reporting Initiative reporting framework, which defines the sustainability reporting framework for the organization. Both methods result in a total environmental burden figure for the whole company. However, these resulting values can't be used to measure the environmental impact of a certain product because the emissions are not allocated to individual products.

The World Resource Institute (WRI) and The World Business Council for Sustainable Development (WBCSD) through the GHG Protocol are developing new standard – Product Life Cycle Accounting and Reporting Standard – that will provide a standardized method to inventory the emissions of an individual product throughout its life cycle. The Global Reporting Initiative (GRI), whose sustainability reporting framework is the most widely used one in the world, has also started to develop GRI sustainability reporting guidelines in its Supply Chain Disclosure project, in order to have sustainability reports cover the whole supply chain's performance.

The most common approach for assessing the environmental impact of a product is the international standard of the life cycle assessment (LCA) ISO 14040 series (International Organization for Standardization, 1997). There are also specifications for the assessment of the greenhouse gas emissions of products: the Publicly Available Specification (PAS) 2050 (Carbon Trust, 2008) that builds on ISO standards for the life cycle assessment by describing the requirements for the assessment of the greenhouse gas emissions, and the ISO 14067 standard for Quantification and Communication of the Carbon footprint of a product.

Life cycle assessment is a standardized method for calculating the environmental impact caused by a product during its life cycle. The goal of LCA is to compare the environmental impact caused by different products so that the customer can choose the least burdensome one.

Life cycle assessment has four main sequential phases. The scope and functional unit of the life cycle assessment are defined in the first phase. The scope defines those life cycle processes, inputs and outputs which are included in the assessment. The functional unit defines what precisely is studied, i.e. it outlines the meaningful entity and provides reference for inputs and outputs of the processes. A functional unit can be, for example, one cubic meter of timber, one mobile phone or one tomato.

In the second phase, which is called life cycle inventory analysis, the input and output flows are defined for each of the processes belonging in the assessment (see Figure 2). There are two flow types: Elementary flows describe the process inputs and outputs – inputs of raw materials and energy resources and outputs of waste and emissions respectively. Product flows are used to describe the flow of products and by-products through the process. If a process produces more than one product, an allocation is also needed. For example, in the forest-wood industry the main product of the sawing process is a board. The sawing process also produces wood chips and saw dust that are considered to be by-products. In this case, the emissions caused by the sawing process can be allocated to a board, wood chips and saw dust using a volume- or value-based allocation method.

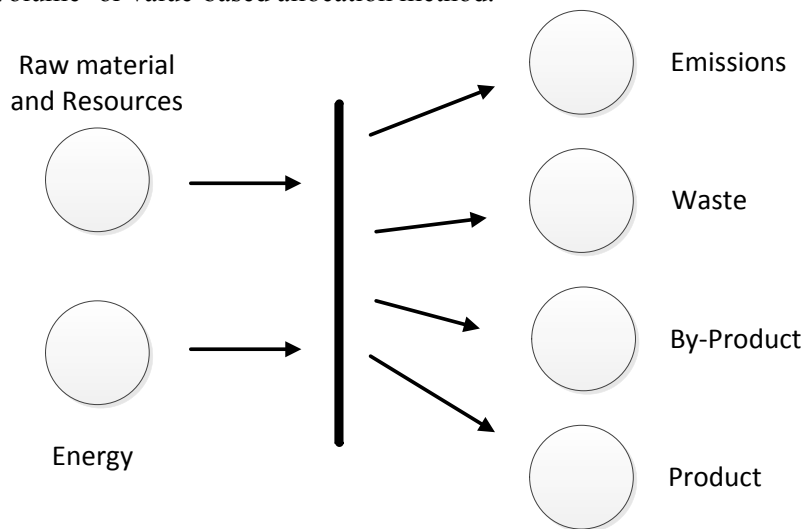


Figure 2. Process and Elementary Flows.

In the third phase of LCA called impact assessment, the results of the life cycle inventory analysis are assigned to the impact categories which include (International Organization for Standardization, 2000):

- Climate change
- Depletion of the stratospheric ozone layer
- Acidification of land and water sources
- Eutrophication
- Formation of photochemical oxidants

- Depletion of fossil energy resources
- Depletion of mineral resources.

For example, emissions into the air, like carbon dioxide and methane, are assigned to the Climate Change category and emissions like sulphur dioxide (SO₂) are assigned to the acidification potential category.

In the final phase, interpretation, conclusions of the assessment are made. The interpretation should include the identification of significant impacts and conclusions based on these impacts. Recommendations for improving the environmental performance should also be made. For example, a study may reveal that the company can rationalize logistics in its raw material acquisition, enabling savings in logistics expenses and reducing the environmental impact.

The biggest problem when measuring the environmental impact caused by a product at the item level is the fact that supply chains are dynamic. A manufacturer of a product can use various suppliers and supply various retailers in different countries. For example, a product that is made from raw material transported from another continent and then transported to a retailer on yet another continent is bound to have different environmental impacts than a product whose supply chain consists of stakeholders all located in the same country. However, the common method of calculating the environmental impact on a product is to measure the resources, raw materials used, emissions and production amounts for a certain period of time – most often, annually - and then calculate the average environmental impact of the product. This method does not take the dynamic nature of the supply chains into account and thus it does not provide sufficient information about an individual product's environmental impact. For example, in the ELCD core database process data set: Spruce wood; timber; production mix, at saw mill (PE International, 2002) the average transport distance between the felling site and the saw mill is 144 km even though the transportation distance of logs can vary from a few kilometers to thousands of kilometers.

If more accurate traceability is needed, the individual products must be identified at the physical level. This enables the tracing and monitoring of the products through the dynamic supply chain. The development of auto identification enables us to identify an object moving in the supply chain. This means that we can connect the movements of the physical objects with their virtual representation in databases. With the help of traceability, we can track the relationships between properties of processes - in this case the environmental burden caused by processes - and actual physical products.

The approach presented in this dissertation supports the phases of life cycle assessment as follows: The scope of the life cycle assessment is modeled by the traceability graph, by expressing those life cycle processes, inputs and outputs which are included in the assessment. The life cycle data of products are collected to the database based on the traceability graph in the Life Cycle Inventory (LCI) phase of the assessment. The impact assessment and interpretation phases are supported by the traceability cube, which makes it possible to use OLAP (Online Analytic Processing) type operations (Chaudhuri and Dayal, 1997) for analyzing and monitoring the life cycle information of the products. The environmental impact data can be shared between supply chain stakeholders using the EPCglobal network presented in Chapter 7.

2.2 Location information

The precise real-time location information and the ability to store and query the route of products enables improvements in logistics and transportation (e.g. Prodonoff, 2008). For example, manufacturers are facing increased pressure from consumers and legislators to know and report the origin of their products – i.e. the raw materials used in the production of the consumer goods. At the moment, it is challenging for producers to gather accurate origin information because supply chains are too complex. For example, forest-industry companies are certificating their products using PEFC¹ chain-of-custody certification (Forests area: 229 million ha, Forest owners: > 475,675). PEFC is a method for tracing wood from forest to the final product to ensure the wood or wood fiber can be traced back to certified forest. The certification allows two different implementation methods: The percentage-based method allows mixing certified and non-certified raw materials; however, a company can only sell the proportion of its production corresponding the proportion of the certified raw material as certified. The physical separation method requires certified and non-certified raw material to be physically separated throughout the supply chain.

As has been indicated, the methods can be integrated and improved using the Traceability Graph. Accurate item-level information of the origin can be used to create a transparent and trustworthy origin certification system. The traceability graph also enables a better and more accurate service for recalling unsafe, defective or hazardous products. By using the precise real-time product location information, the traceability graph allows manufacturers to recall only the products containing unsafe elements. For example, the precise traceability information can be used in the food industry. The traceability graph enables manufacturers, suppliers and resellers to avoid total recalls, by identifying the products that contain a suspicious raw material. With the system, companies are better equipped to retrieve affected products and to protect their reputation and brand value.

¹ <http://www.pefc.org/>

3. Workflows

Various graph-based methods have been developed for describing the functionality (dynamic aspects) of information systems. Early dataflow languages e.g. (Kosinski, 1973) were primarily intended for modeling the actions of computer programs, whereas workflow models (van der Aalst and van Hee, 2002) are focused on modeling physical processes and activities. Historically, the distinction between dataflow and workflow models is vague and some authors (Ellis and Nutt, 1980) use the term “information flow” for describing the flowing of all information from one process to another. Nowadays, workflow diagrams typically distinguish different types of activities, such as sending, transforming and packing of materials. Some modern modeling methods of information systems, e.g. UML (Booch et al., 1999), contain diagrams for dynamic aspects of programs (interaction diagrams) and the modeling of workflows (activity diagram).

In workflows, materials, documents and other pieces of information are transferred from one process to another (van der Aalst and van Hee, 2002; Bonner, 1999). The workflow model of UML is informally defined and its purpose is to map real-world activities to the underlying software solution (or vice versa). There are also a number of commercial applications that have a component for drawing workflow diagrams. Their common feature is that they support the illustration of different types of processes.

There are two main types of workflow models, process- and data-centric. The traditional workflow models are mainly process-centric. Recently data-centric workflow approach has gained popularity. In data-centric workflows the goal is to model initial, intermediate and final data sets, based on the transformations of data sets. In process-centric models the focus is on processes and the timing between them.

3.1 Modeling Workflows

There are formal methods for modeling the functionality of information systems that are not primarily intended for any specific purpose. Petri Nets (Petri, 1966), YAWL (van der Aalst and Hofstede, 2005), Temporal logic (Attie et al., 1993), and Transaction Logic (Bonner, 1999) are further good examples of formalizing workflows. These methods emphasize the timing within processes and supply chains. In general, these are very expressive languages that enable not only the modeling of the functionality of systems but also other aspects of information systems. For example, Transaction Logic is based on F-logic (Kifer and Lausen, 1995) that is a general framework for specifying object-oriented and deductive aspects. Next, the Petri nets are addressed in more detail because it is the most common method for modeling workflows.

For example, Petri nets are a mathematical modeling tool for describing different systems. Petri net is a bipartite graph consisting of places and transitions, which are connected by directed arcs; a Petri Net is a tuple $\langle P, T, F \rangle$ where P is a set of places p_1, p_2, \dots, p_n , T is a set of transitions t_1, t_2, \dots, t_n and F is a multi-set of relations (arcs) between places and/or transitions, i.e.

$$F \subseteq (P \times T) \cup (T \times P).$$

Petri nets describe the behavior of systems with system states and their changes. Places may contain tokens which are used to model the state of the system. State changes are described with transitions. The tokens are consumed when transition fires. The tokens are moved by a transition from input arcs to output arcs. Places are illustrated by circles, transitions with bars and arcs with arrows. Tokens are represented as black dots.

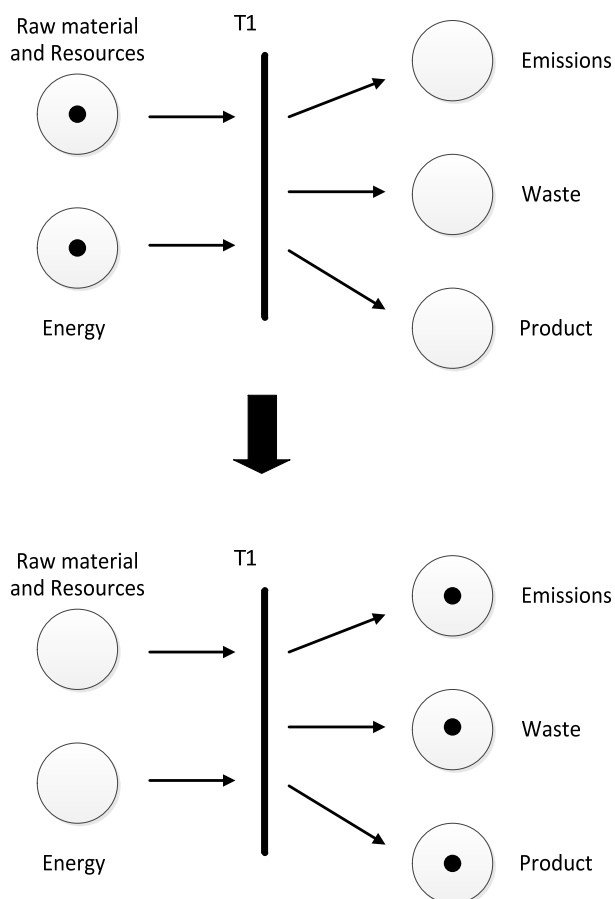


Figure 3. Petri Net example.

The primitives of Petri nets are illustrated in Figure 3 using a simplified manufacturing process where raw material and resources are used to produce a product, generating waste and emissions at the same time. Figure 3 (the upper part) illustrates that energy, resources and raw material are available, and transition T1 is enabled. When transition T1 is triggered, the tokens are moved via output arcs to places shown in Figure 3 (the lower part), where transition T1 is no longer possible.

Originally, Petri Nets were developed for modeling the concurrent behavior of distributed systems and they are traditionally used for defining or describing the functionality of computer programs, but they are also proposed for the exact representation of workflow models (van der Aalst, 1998; van der Aalst and van Hee, 2002) and a wide variety of other application areas (Murata, 1989) like performance evaluations, communication protocols, software design, process modeling (DiCesare et al., 1993) and concurrent programming. There are many extensions to the classical Petri Nets – called high-level Petri Nets – that are used either to simplify the model (e.g. Colored Petri Nets, Hierarchy) or to add properties that cannot be modeled using basic Petri Net (e.g. Timed Petri Nets).

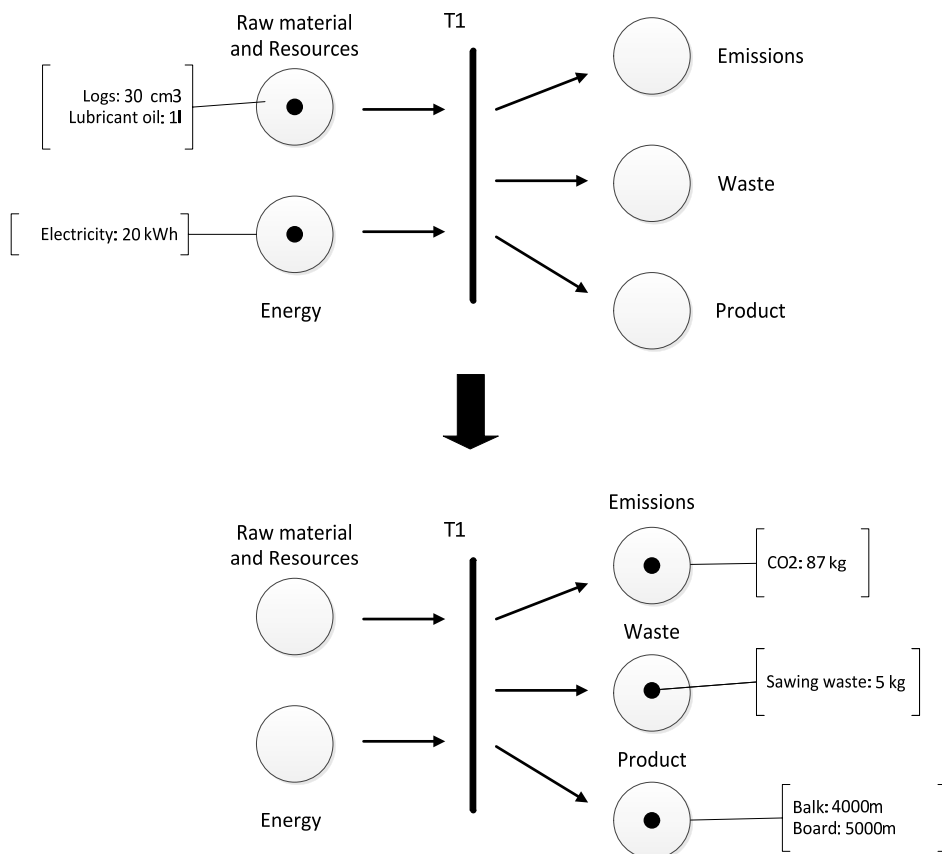


Figure 4. Colored Petri Net.

In a real system, tokens can represent products that have attributes such as volume, price or identification code. However, in the classical Petri Nets, tokens are identical and cannot be distinguished. In a Colored Petri Nets, each token has attributes for presenting properties of processes. For example, in Figure 4 the tokens have the following attributes:

- Raw material and resources (logs: 30 cubic meters; lubricant oil for saw: 5 liters)
- Energy (Electricity: 20 kWh)
- Emission (CO2: 87 kg)
- Waste (Sawing waste: 5 kg)
- Product (Balk: 4000 m; Board: 5000 m)

Another problem when modeling real-world systems with classical Petri nets is that the models usually become too complex resulting in problems when analyzing systems with the workflow management system. The hierarchical extension of classical Petri Nets can be made by decompose complex systems into subsystems.

Timing is another usable extension which is used to model the temporal behavior of a system. One method of introducing time to Petri Nets is to add a notion of temporal constraint to transition (Berthomieu and Diaz, 1991). A constraint defines the lower and upper bounds to the transitions. For example, as illustrated in Figure 5, if raw material token arrives at t_0 and energy token in t_3 then transition T1 can be triggered between t_8 and t_{13} .

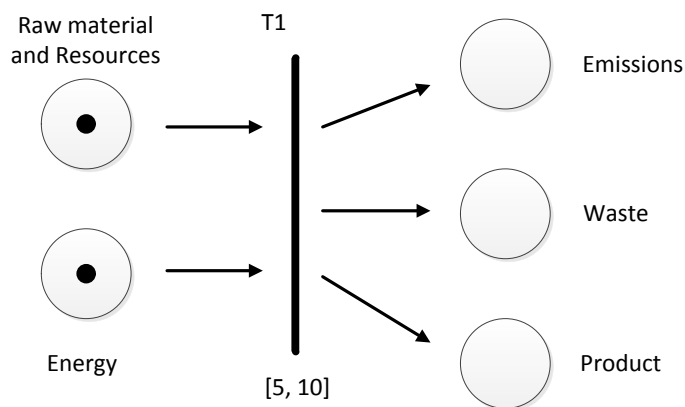


Figure 5. Timed Petri Net.

3.2 Data-Centric Workflows

In data-centric workflows the goal is to visualize initial, intermediate and final data sets, focusing on the transformations of data sets (Akram et. al., 2006). The data sets are parameters to services that consume the input data set and create output data sets. The most usual application area of data-centric workflows is scientific problem-solving.

In scientific problem-solving, data-centric workflows are used to model, design and execute an analysis, capturing series of analytical steps. The primary feature of the scientific workflow methods is to concentrate on process functionality based on the data (Curcin and Ghanem, 2008). Each analytical step entails instructions how to handle the underlying data.

Scientific workflows provide formalization for the scientific analysis process by modeling the data transportation, transformation and analytical steps between distributed computational steps accelerating scientific progress. Scientific workflows are directed graphs where places represent computation steps and arcs represent the transitions between these steps. The main components of the scientific workflow model are workflow engine, services, applications and tokens of data. Services are accessed through interfaces, with applications implementing the functionality of the services. Tokens of data are consumed and produced by services which form the workflow. A workflow engine invokes the services in a predefined order.

Normally, scientific calculation involves huge data sets, geographically distributed over users and resources. A process often creates multiple queries invoking many analytical calculations. There normally is a semantic mismatch between applications, and calculations executed by applications can be time-consuming. These requirements pose a big challenge for scientific workflow systems (Barker and van Hemert, 2007).

One example of scientific workflow systems is Kepler (Altintas et. al., 2004) that is a system for constructing, composing, orchestrating and sharing scientific workflows. In Kepler the workflow components are represented by actors that represent computational steps or data sources. With output and input ports that are used to map connections that transport tokens. The actors and connections implement the workflow model. Kepler is used to model the data flow between computational steps used to achieve an outcome called scientific result.

Kepler includes a library of over 350 actors that can be used to analyze and integrate applications and it supports various data models. Kepler has also been used to combine both real-time data streams and data from archives, enabling the modeling and analysis of streaming sensor network data - for example, meteorological sensors (Barsegian et. al., 2010; Penning et. al., 2007).

Scientific workflows play an important role in Grid computing which combines multiple computer resources to solve a single task, such as geophysics, astronomy or bioinformatics where the amount of data can be petabytes (Foster et. al., 2001; Chervenak et. al., 2001). The main principles in Grid are the non-centralized control of resources, standardized open

protocols and interfaces. A Grid network is a distributed system that enables to share anonymous resources based on their availability.

The model presented in this dissertation can be seen as a data-centric workflow where each supply chain process is a calculation step, from where the life cycle data are allocated to a product. The main aspects of data-centric workflows (Curcin and Ghanem, 2006) include concentrating on process functionality based on the data, i.e. each node has behavior instructions related to the data. From the life cycle data management perspective, this is a secondary feature, as well as timing representation in YAWL, Temporal logic, and Transaction Logic. Instead, the traceability graph emphasizes the handling of aggregation and movement of data between processes.

The life cycle information from each process is shared by all supply chain stakeholders by services in a Grid system. The infrastructure of the system is presented in Section 6 where the system is also characterized based on the Grid system taxonomy.

3.3 Provenance in Workflows

The most important requirement in scientific analysis is the ability to reproduce and reuse results. The *provenance* information enables users of a scientific workflow system to track the information about calculation steps and data used to derive the data products, thus enabling the scientific community to share the results and reproduce the calculation process (Freire et. al., 2008; Simmhan et. al., 2005).

There are two different types of provenance information (Clifford et. al., 2008). *Prospective provenance* contains specification about the calculation steps, i.e. the scientific workflow that needs to be executed. *Retrospective provenance* contains information about the calculation environment used to derive the data product and the calculation steps that were executed. The provenance information must also contain the relationships between data products and calculation steps. The provenance includes inputs of a workflow, the outputs of the workflow, and definitions of the calculation steps and data tokens between the steps.

Various provenance models have been developed (for example S. Cohen et. al., 2006) which all share the same support for retrospective provenance and dependencies between data and processes. The provenance data must contain at least: a unique identifier of any data token, information about the calculation steps and sequence of the input/output operations and provenance information for user and or parameter data. For example, the provenance information collected by the Kepler Scientific Workflow System (Altintas et. al., 2006) includes:

- Contextual information (who, what, where, when and why) about the executed scientific workflow
- Workflow definition, i.e. the actors of the workflow, the connections between the actors and the parameters used.
- Workflow evolution information that specifies the evolution of the workflow definition.
- Input data including the information about the transformations applied to a data item.
- Output data including the information about the transformations applied to a data item.
- Information about intermediate data sets.

(Cohen et. al., 2006) define the provenance as a function $Prov(d)$ with input of the data token d and output of a set of steps and input data tokens on which it depends. $Info(d)$ contains contextual information about the executed calculation step or data token at hand.

$$Prov(d) = \begin{cases} (sid, \{d_1:Prov(d_1), \dots, d_n:Prov(d_n)\}) & d \text{ is calculated data} \\ Info(d) & d \text{ otherwise} \end{cases}$$

where sid is the id of the calculation step, d is the output data token and d_i is the input data token.

In our example, see Figure 3, we process the provenance output data object for Emissions

$$Prov(Emissions) = (T1, \left\{ \begin{array}{l} \text{Raw material and Resources: } Info(\text{Raw material and Resources}), \\ \text{Energy: } Info(\text{Energy}) \end{array} \right\})$$

The provenance data collected from supply chain processes enables the connection between the products and processes. This means that the relationship between processes and products can be traced with provenance information. The example above features the environmental burden caused by processes, and actual product instances. Next, a model for managing the provenance information of products (life cycle data) in the supply chain is developed from the viewpoint of data modeling.

4. Logical Tracing: traceability graph

In data models, the intensional (schema) and extensional (instance) levels are typically distinguished. The intensional level describes the information about concepts and their relationships and the extensional level describes the object and their relationships. In models

of complex structures, such as composed objects, the strict correspondence between these levels is essential because instances (objects) of an object type may structurally differ from each other. The same concerns many applications of workflows: the supply chains of similar objects may vary and different data are associated with objects and their components. For the purpose of managing the structural diversity of composed objects, the integration of the intensional and extensional levels has been proposed (Junkkari, 2005). This allows advanced structural analysis and declarative query formulation thorough transitive relationships (Niemi et. al., 2004). In the model presented in this dissertation, the levels are integrated into each other, i.e. the workflow schemata and instances are not explicitly separated. The same concerns the data associated with workflows. This gives a possibility for forming more flexible workflows.

The traceability graph developed in this work is used to model the supply processes of physical products and resources and emissions associated with the products and their components. The traceability graph includes principal primitives needed to model and manipulate data-centric aspects in workflows in a way that enables tracing products at different granularity levels. The traceability graph is not bound to any existing data or workflow methods. The model can be applied using existing formalisms and systems. First, the supply chain is modeled as a basic workflow model as presented in Figure 6.

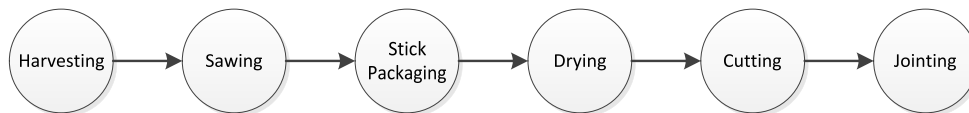


Figure 6. Example Workflow Model.

The workflow model provides a starting point for developing the traceability graph that can be seen as a provenance model, where provenance is collected as a set of nodes – discrete activities through supply chain processes – that describe the workflow of an individual object traveling through a supply chain. Each node has the specific physical location, the time interval how long an object was present in the node and elementary flows – life cycle data shared and collected from the process.

The traceability graph has the ability to manipulate products and their transformations. For example, a product may be composed of many parts, or a product may be manufactured using masses of raw materials. Table 1 summarizes the cases of object transformation. The traceability graph also has the ability to manipulate the properties of processes and to allocate them to products that are handled in that process.

Table 1. Identity manipulation

Transformation	Identity	Description
equivalence	maintain	a batch is transferred from a specific process to following as such
subsetting	maintain	a batch is divided into subsets for refining
supersetting	maintain	several patches are combined together for refining
division	change	objects are splitted into several objects
composition	change	objects are composed to single objects

The traceability graph can be presented using nodes and edges with their properties. A node is used to describe a discrete supply chain process that has a specific physical location and the time interval when an object was present in the node. The node also contains information about object life cycle data from that process, for example, the amount of energy used during the process allocated to an individual object. An edge expresses a product flow between two processes. The supply chain of an object can be described as nodes following each other in a partial order. A manufacturing process is an activity transforming the raw materials, components and energy into products, waste and emissions. The properties of edges and nodes of the traceability graph are presented in Table 2.

Table 2. Properties of edges and nodes.

name	description
object	The subject of tracing. This can be either a single product or a batch depending on the precision of tracing in the underlying supply system
node	A discrete supply chain process, which has a specific physical location, the time interval when an set of product portions was present in a node and the set of attributes associated with the process
product portion	It involves the quantity of products, the identifiers of objects and the ratio of the emissions and resources compared with the total ones in the process node. The ratio is calculated using an application specific-method. It can be based on the portion of mass or machine operation time, for example. Product portions of a process are viewed through the end products of a process.
attribute	It expresses information associated with a process. Input attributes can describe the resources of a process, whereas output attributes can be used for determining the emissions of a process. Each attribute has two values: one for the underlying process and the other for containing the cumulated values from the previous nodes. A cumulated value is calculated based on the ratios of product portions and quantity that is sifted from the previous nodes via edges.
edge	Products are sifted from a node to another via edges. An edge also determines the mapping of objects between two processes as described in Table 1.
sifted product portion	An edge involves those objects that are sifted from a start node to the end node (only some products of a product portion may be selected from other processes). This part of the start node's product portion is called a <i>sifted product portion</i> . In transferring products from a process to another, the attributes must be re-calculated for corresponding to the sifted product

	portion. This is based on the ordinary and <i>derived attributes</i>
derived attributes	These are associated with an edge and determine the amount of an ordinary attribute that is related to the sifted product portion.

In Figure 7 there are eight process types (A, B, ..., H). Process type A has four instances (A1, ..., A4). These nodes have no predecessor, which means that the traced objects have been created in these nodes. The objects are transferred forward in the graph. For example, objects from Nodes A2 and A3 are transferred to Node B2. In this step, objects are not modified but object sets (product portions) of A2 and A3 are unionized to the object set in B2. This also means that the resources and emissions of process nodes A2 and A3 are aggregated to the new set of objects in B2. In the C process nodes and G process nodes, the objects in hand are divided into several objects. A double-headed arrow illustrates this. For example, a physical object is decomposed or divided into parts. The products of process nodes G are components for the process nodes H, i.e. in node H2 objects are composed of the objects that are outcomes from Nodes E1 and E2. A shared start arrow illustrates this.

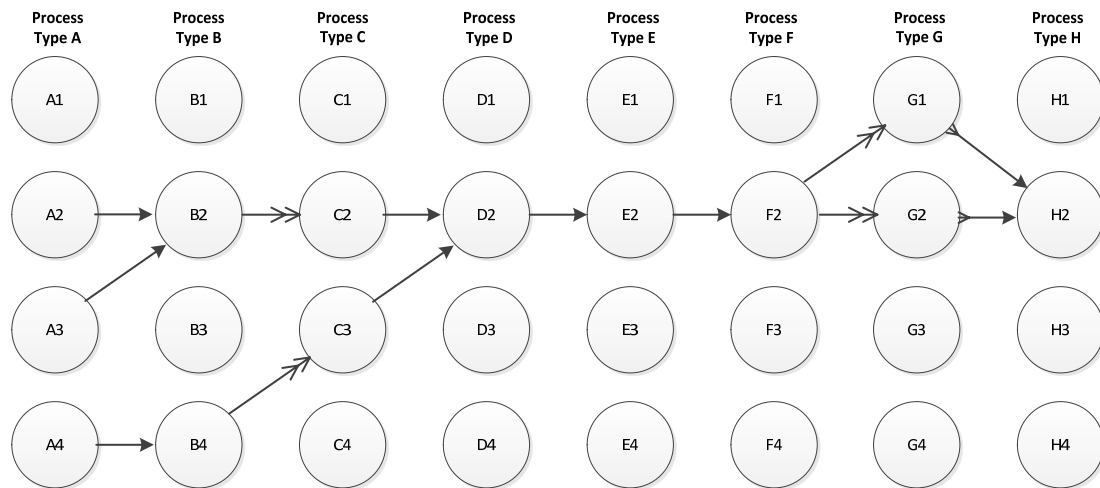


Figure 7. Traceability Graph.

In a traceability graph, nodes can be grouped based on their process types, i.e. similar nodes are instances of a process type. The nodes of the process type may have different properties, such as the time span and physical location of an activity.

The development of smart identification enables us to identify an object moving in the supply chain and this enables us to track the movements of the object throughout its supply chain. This means that we can connect the physical-world objects with their virtual counterparts in databases. With traceability we can track the relationships between properties of processes - in this case the environmental burden caused by processes - and actual product instances. The object transformation – Table 1 – is managed by storing the physical identifiers of related

objects to the database. For example when a log is sawn to boards (*division transformation*), the physical identifier of the object *log* and the physical identifiers of the objects *boards* are stored to the database.

In Figure 8 we present our running example, simplified glued laminated timber production. The processes included in the example are the felling of trees (harvesting), transporting the logs to the saw mill, sawing logs to boards, packaging the board for drying, drying the boards, transporting the boards to the refiner, cutting the boards and, finally, jointing the glued laminated timbers from boards.

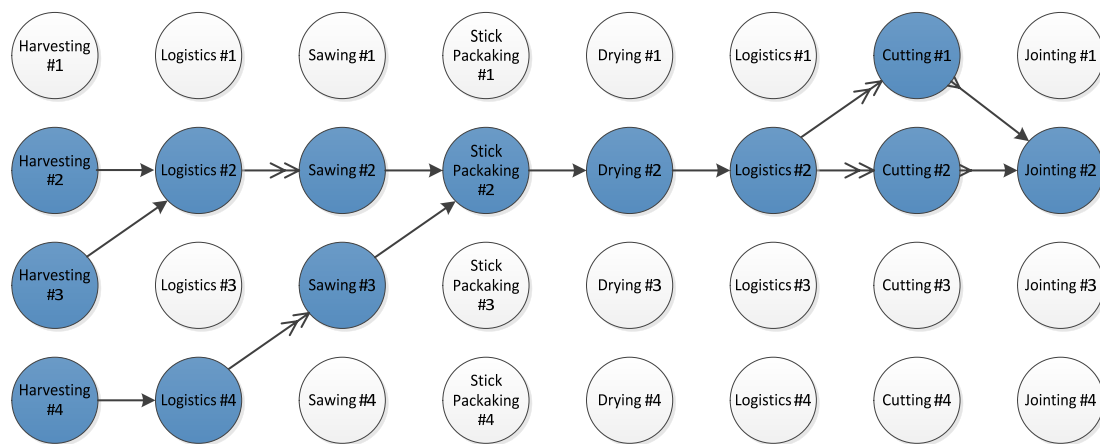


Figure 8. A sample supply chain.

In the sawing nodes, the objects are divided into several objects, i.e. a log is sawn into multiple boards; a double-headed arrow illustrates this. Then, these object move forward in the graph to forthcoming activities. In jointing nodes, objects are composed of objects that are products of the cutting nodes, i.e. a glulam beam is glued from multiple boards; a shared start arrow illustrates this.

Using a traceability graph, the supply chain of the object can be traced, i.e. the route of the object can be monitored. This also means that all the life cycle information related to those processes can be attached to the object. In Figure 8 the colored nodes are distinct activities in which the object, its component or raw material used to produce the object have participated. This subgraph is the supply chain of the specific object in Jointing #2. The main feature of the traceability graph is the support for querying and analyzing of product life cycle information. The graphical representation only illustrates nodes (processes) and the edges between the processes. The type of the edges specifies the manipulation rules for product portions and attributes. A more detailed description is found in (Junkkari and Sirikka, 2011; Paper II)

5. Physical Tracing: Radio Frequency Identification

An identification method is needed in order to be able to identify each object physically in the supply chain. With physical traceability, we can connect the physical information about the products' movements with the information about processes. In this study we focus on the Radio-Frequency Identification (RFID) technology for marking the objects in a supply chain. The RFID technology has the following benefits compared with the traditional identification technologies like bar codes: The system does not require a line of sight and reader and tag communication are not orientation sensitive. The reading is automatic and does not need human intervention. Each item is individually labeled with a high degree of security - a tag is more difficult to counterfeit than a simple barcode.

RFID uses radio waves to transfer identification from a tag that is attached to an object (Stockman 1948). The RFID technology is similar to bar code identification in the sense that an identification code is embedded to an object. With RFID technology the identification process does not require a clear line of sight, and a vast amount of tags can be read simultaneously. RFID can be used in various application areas, each requiring a specific type of RFID tags based on the application requirements (Finkenzeller 2004). RFID tags can be active, passive or semi-passive

- Passive tags do not have their own source of power. The needed power is gained by using the energy of the radio wave from the RFID reader antenna. The read range of passive tags is up to 10 meters. Passive tags are the most inexpensive alternative, and thus they are most commonly used in industrial applications.
- Active tags contain their own power source and transmitter, thus providing a much longer identification distance.
- Semi-Passive tags contain a power source but they do not include a transmitter; nevertheless they provide a longer reading distance than passive tags.

RFID tags can also be divided into four groups based on the used radio frequency.

- Low frequency (125-134 kHz) tags are used for identification purposes; for example, animal tagging, access control and car immobilizers.
- High frequency (13,56 MHz) tags are used, for example, in asset tracking; hospital systems, reusable assets and laundry & library systems.
- Ultra High Frequency (868-956 MHz) tags are used in logistics applications like the tracking of containers and pallets.
- Micro Wave (2,45 GHz) applications are used in vehicle identification, for example, in harbors.

The RFID reader application reads the RFID tag constantly when the tag is within the reader's range, producing a stream of observation tuples $\langle \text{reader}, \text{EPC}, \text{timestamp} \rangle$; reader is an identification code of the RFID reader, timestamp is the time instant of an observation, EPC is an Electronic Product Code that is a universal identifier which provides a unique identity for physical objects. EPC is defined by the EPCglobal Tag Data Specification (EPCglobal 2008b).

The EPC number – illustrated in Figure 9 - is an 64- or 96-bit code that is divided into number sequences. Example: The EPC number header part defines the coding scheme of the EPC number, the manager number defines the company that has authority for the types of products in a supply chain, the object class part defines the category of an object, i.e. product group, and the serial number is the unique code in that identifies the specific object.

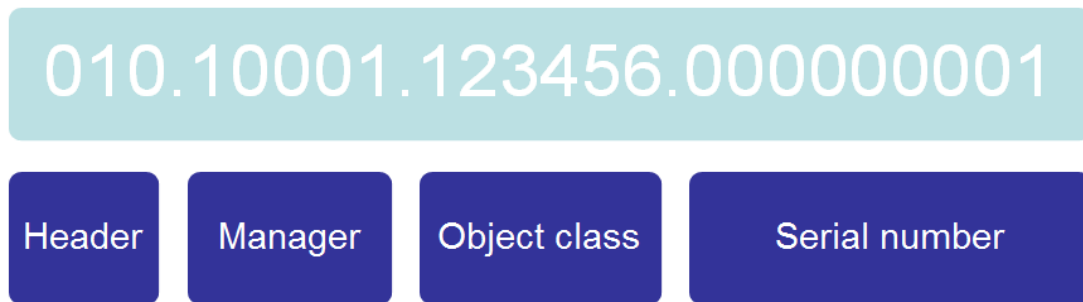


Figure 9. EPC number.

The volume of data generated by storing the complete history of movements of individual objects throughout the supply chain is enormous. To be able to use this information effectively, a data structure that supports effective information retrieval is needed. Aggregation and filtering are common methods used in RFID middleware applications to reduce the amount of RFID data generated. This includes the removal of certain RFID readings based on the reader's identification code or product's electronic identification code. Table 3 presents the commonly used aggregation types (Floerkemeier and Lampe, 2005).

Table 3. Aggregation Types.

Aggregation Type	Description
Entry & Exit	Used to reduce the successful reads of the tag to two readings – In and Out – i.e. when the object appeared in the reader's range and when the object disappeared from the reader's range. The raw RFID data must be cleansed to include only the entry event and exit event of an observation. Cleansed RFID observations are stored in a stay tuple (reader, EPC, time_in, time_out) where time_in and time_out attributes describe the time the item stayed in the location of the reader.
Count	Used to return only the total amount of similar objects. Similarity can be based on some object property.
Passage	Entry & Exit observations may be compressed to a simple passage observation.
Virtual Readers	The RFID readers may be grouped into a function as one “virtual” reader. This enables the application to see them as one reader with a bigger read range.

RFID data may also be compressed by using object containment. This idea is presented in (Gonzalez et. al., 2006b) and (Wang et al., 2010). Usually, the products are moved into groups inside a containment object during some steps in the supply chain. For example, the manufacturer sends products to the distribution centers in containers. If an electronic identification code is used to mark the entire transport batch, the movement information of each individual product need not be tracked and stored every time the container is moved. Actually, (Harrison 2003) presented the same idea, calling it 'symbolic location'. It means that the object can be contained by another object.

The readings made about containment entities like package, container or truck load can be stored to the same tuple $\langle \text{EPC}, \text{location}, \text{time_in}, \text{time_out} \rangle$ where EPC is an identifier of the containment object. To be able to trace the individual products inside the containment object, tuple containment $\langle \text{co_EPC}, \text{EPC}, \text{valid_from}, \text{valid_to} \rangle$ is needed; co_EPC is a code of an containment object and EPC is a code of a contained object. The validity of containment is stored in the attributes valid_from and valid_to. Valid_from shows when the containment starts and valid_to shows when the containment relation ceases to be valid.

The benefit of using the containment tuple is the point that it will compress the amount of saved RFID readings by the proportion of objects in a containment object and movements recorded for each containment object. For example, a container may contain thousands of individual products and it can be moved multiple times during the logistics process from manufacturer to shop floor, which means that the amount of rows in the database will be reduced considerably.

The RFID technology and data management have been researched extensively during the past decade – e.g. RFID data staging to OLTP and OLAP applications (Krompass et. al, 2007) provides a method that responds to different kinds of needs posed by transactional and analytical applications on RFID data, RFID data management (Sudarshan et. al., 2004) and RFID data warehousing (Gonzalez et. al, 2006).

The next Section presents the storing and analyzing of traceability data. To enable the analysis of the data amount resulting from the traceability system, the data is transformed to a *traceability cube*, which is a multidimensional array optimized for handling large amounts of data. In the work, this is integrated with RFID data.

6. Storing and Analyzing Traceability Data

The life cycle information of products from the traceability graph must be stored to a database in order for it to be usable for monitoring and analysis. In this dissertation, the information is mapped to a relational database, see Figure 10. The main relations needed for storing the information are *Node*, *Object* and *Route*. The *Object* relation is used to store the information about object transformation – the relationships between objects, *Node* relation stores the process-specific information and *Route* relation is used to connect objects with nodes, thus creating the supply chain of a specific product. More detailed information can be found in (Sirkka and Junkkari, 2011; Paper III).

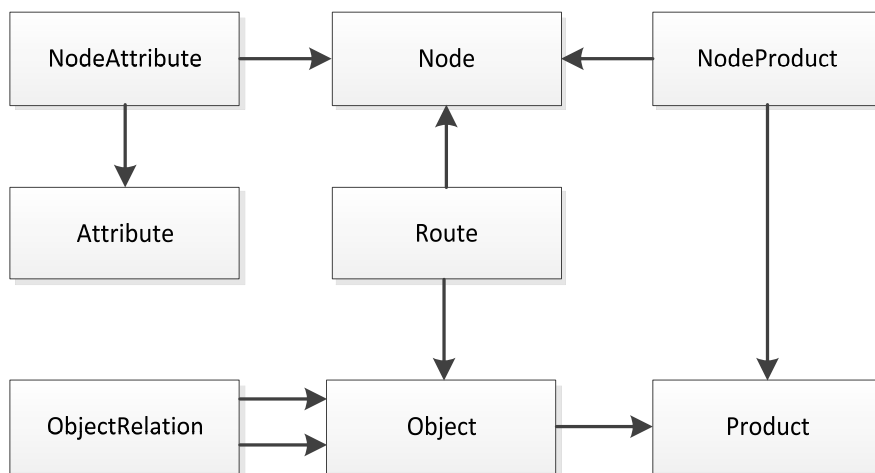


Figure 10. Database schema for the traceability graph.

The data cube can be utilized for analyzing the life cycle information stored in the database. The *traceability cube* enables OLAP type operations for analyzing the life cycle information of the products. OLAP is a method used for describing the analysis of the complex data in data warehouses. The OLAP Council has defined three main functions that are provided by OLAP systems: multidimensional views of data, ability to perform complex calculations and intelligent handling of time, i.e. time intelligence.

The OLAP systems can be divided into three main categories. In the relational OLAP type, the data and dimensions are stored as relational tables in the database. The multidimensional OLAP – used in this paper – stores the data in the optimized multi-dimensional array called OLAP Cube. The OLAP Cube is a data structure that allows the fast analysis of data from multiple perspectives (dimensions). The cube consists of facts that are called measures, and dimensions that categorize the facts. The hybrid OLAP mixes the ROLAP and MOLAP types, i.e. it can use relational tables for some data and cubes for more aggregated data. (Chaudhuri, 1997)

In this paper, the multidimensional data model, MD (Torlone, 2003) is used. In figures, dimensions are presented as round-cornered boxes, facts (measures) are presented as boxes, descriptions of dimensions are presented as small diamonds, and measures as circles. The circles drawn with a dashed line present calculated measures. The example traceability cube contains information about the environmental performance of objects moving through a supply chain enabling carbon disclosure from the whole supply chain-level down to the level of an individual product.

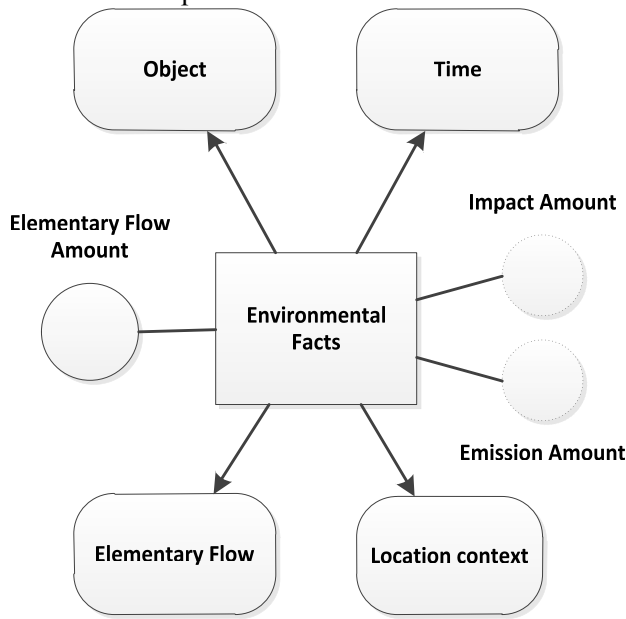


Figure 11. Traceability Cube Schema

Figure 11 presents the traceability cube with a few example dimensions. The cube consists of facts, the amount of elementary flows and volume of objects, and dimensions that categorize the facts. Dimensions can be seen as meta-data for the measures to which they provide information. The example cube also includes the calculated measures *Impact Amount* and *Emission Amount*. As an example, the calculated measure 'emission amount (carbon dioxide)' is calculated using environmental databases that contain life cycle inventory (LCI) data. One of these is the ELCD core database by the European Commission – DG Joint Research Centre – Institute for Environment and Sustainability which includes more than 300 LCI data sets. For example, according to this database, carbon dioxide emissions when using *articulated lorry transport; Euro 0, 1, 2, 3, 4 mix; 40 t total weight, 27 t max payload* total 4,442 g per kilogram of transported product per kilometer (PEInternational, 2002).

The environmental impact categories are calculated based on the emissions. For example, the climate change potential is calculated using carbon dioxide (CO₂), methane, nitrous oxide

and several other emissions - full list: PAS 2050 (CarbonTrust, 2008). The climate change potential is measured as a kg of CO₂ equivalent. The calculation is done by converting the other emissions to CO₂ by using predefined conversion factors. For example, the conversion factor for methane is 25 and for nitrous oxide the conversion factor is 298 (CarbonTrust, 2008). The calculated measures are calculated during the loading procedure and the calculation rules and conversion factors can be configured based on the different specifications. Figure 12 presents the calculation and sample instances over the traceability cube.

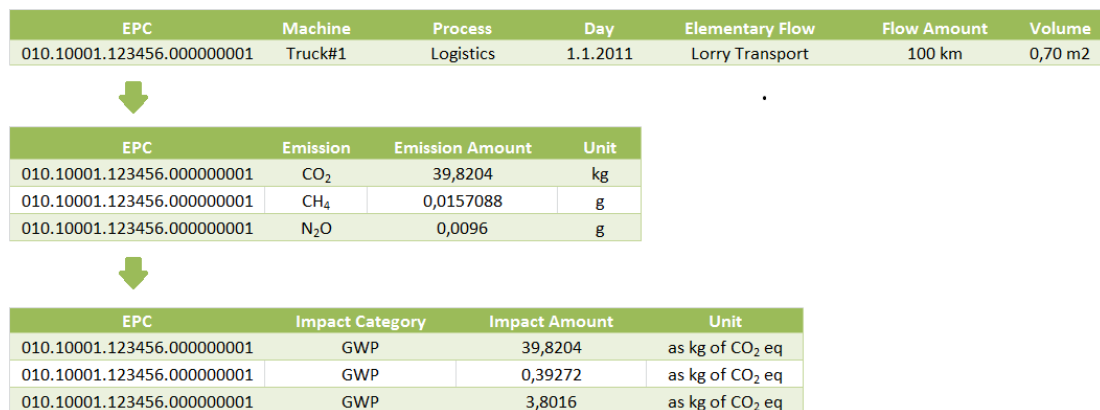


Figure 12. Emission and Impact Calculation.

The location context dimension contains the organizational hierarchy and other process-specific information, in this example Machine → Process → Site → Organization. The machine level includes the spatial description location that can be used for the spatial analysis of environmental burden. Moreover, the location context dimension can be used to benchmark between different processes or manufacturers, i.e. it can be used to compact the workflow. For example, logistics nodes could be merged, resulting in the total amount of transporting in the sample supply chain.

The object dimension is used to describe the properties of an object and it can be used to benchmark the environmental burden among different products and product types. By comparing the environmental performance of the same type of processes, companies can see whether the performance metrics of their process compare to the industry's best practices. This allows companies to notice where they could make the biggest improvements in their environmental performance. Similarly, consumers can select the products that feature the lowest environmental burden, thus forcing the manufacturers to adopt cleaner manufacturing procedures.

In data warehousing, the date dimension is the most frequently used dimension. The date dimension is used for analyzing the effect of time regarding the environmental impact. For example, with the date dimension we can perform a trend analysis for detecting a pattern of

behavior in a data series. Moreover, historical and periodical comparison analyzes can be made.

Common OLAP operations include *slicing* and *dicing* the data, providing a multidimensional view of data based on subsets corresponding to the selected dimensions. *Drilling down* and *up* rises or lowers the level of aggregation. For example, we can view the data at the daily, weekly, monthly or yearly level. *Pivot* and *rotate* operation provides an alternative view of the data. Pivot operations can be used to change the dimensional orientation of the data cube.

The *slicing* operation provides a multidimensional view to the data based on the selected dimension. For example, the data can be sliced according to the date dimension. Figure 13 illustrates how the environmental data of all locations and product types in January 2011 can be sliced from the OLAP cube.

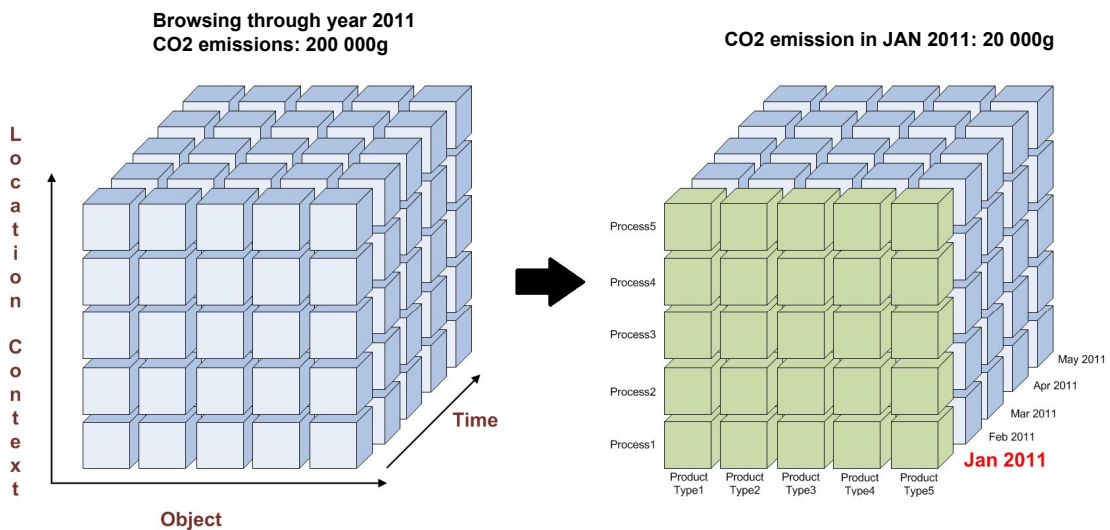


Figure 13. Slicing using time dimension.

Dicing means slicing the data on more than two dimensions. Figure 14 illustrates how environmental data is diced to include only the values of a certain process, product type and month. By comparing the environmental performance of the same type of processes, companies can see whether the performance metrics of their process compare to the industry's best practices. This allows companies to notice where they could make the biggest improvements in their environmental performance.

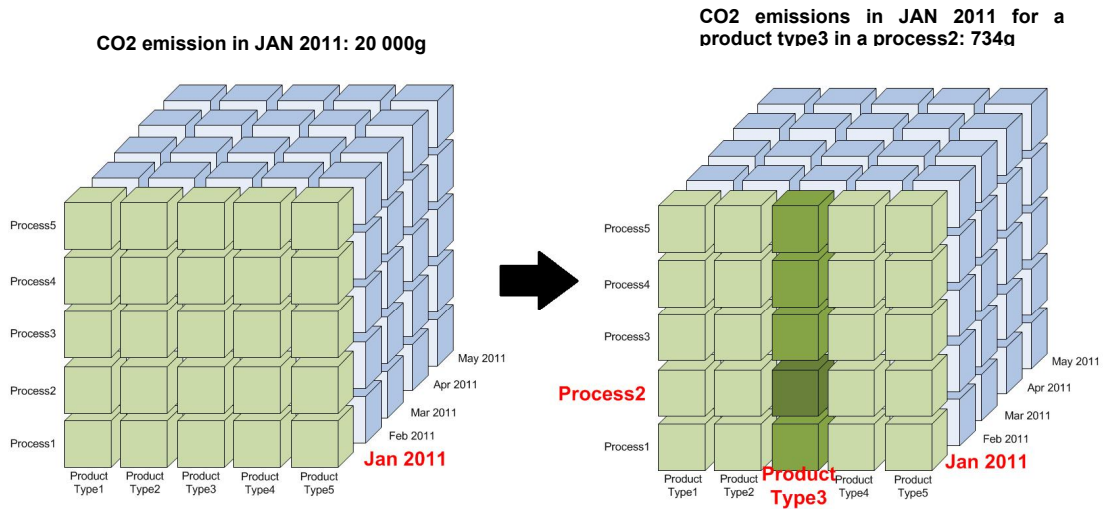


Figure 14. Dicing the Traceability Cube.

A cube can be drilled up or down, which means analyzing the data between the aggregation levels from the summarized data from the whole supply chains to the detailed data of an individual product. For example, the location context dimension can be drilled down from the logistics process to types of machines executing the process as illustrated in Figure 15. This information can be utilized for analyzing the environmental effect of different transporting methods.

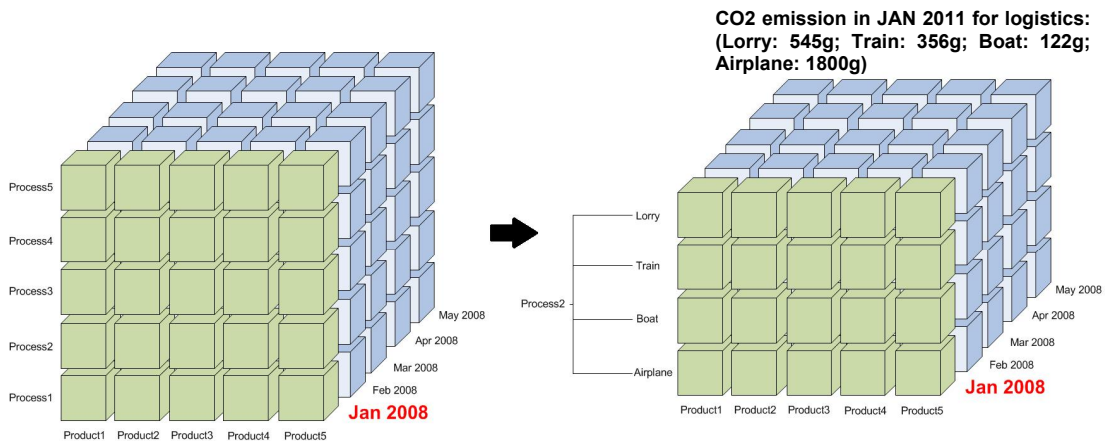


Figure 15. Drilling down in the Traceability Cube.

The analytics capabilities of the traceability cube can be used for analyzing the life cycle data. For example, environmental data can be summed up to create the total environmental impact of the whole life cycle of the product. The data can also be used for comparing the

performance between different manufacturers or manufacturing sites. The possibility to analyze the supply chain on the process and item levels allows end users to select the product with the lowest environmental burden. This creates pressure for manufacturers to improve the eco-efficiency of their supply chains.

7. Sharing the Life Cycle Information

In order to be able to generate life cycle information for the product throughout its whole life cycle, organizations share the environmental information of the products that were handled by them in their part of the supply chain. For example, some organizations are responsible for raw material extraction, others for manufacturing components and supplying these to the other manufacturers, some organizations are responsible for transporting the objects from one stakeholder to other, etc.

7.1 EPCglobal Architecture

The stakeholders of a supply chain own a part of the life cycle information of the final product. To be able to share product-related information in the complex supply chains, the organizations have to agree on a common standard. The EPCglobal Architecture - Figure 16 - framework standards (EPCglobal, 2009a) enable the supply chain stakeholders to capture, store and share product-related data. The EPC Information Services standard (EPCglobal 2007) defines the storing and sharing of the traceability data that is created when a product marked with an RFID tag passes an RFID reader in a process in a supply chain.

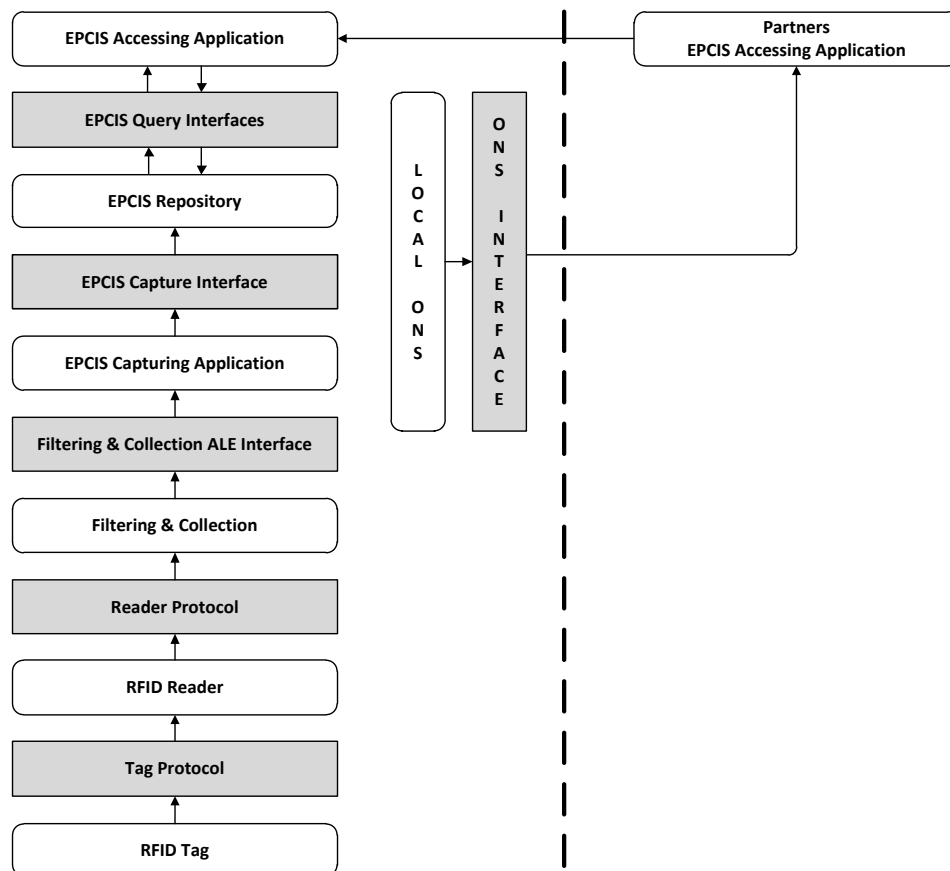


Figure 16. EPCglobal Architecture.

The main parts of the *RFID tag* are a chip, an antenna and a packaging. The chip holds the information attached to the physical object. The antenna transmits the information to the reader, and the packaging holds the antenna and the chip.

RFID readers read (or, depending on the reader, also write) information on the RFID tag without needing a clear line of sight. The RFID reader can be integrated into a machine or it can be a separate hand-held device. *Reader Protocol Interface* specifies the delivery of raw tag reads from RFID readers to the *Filtering & Collection* role. The raw event data contains a unique identification code, location and time ('Reader X in time Y the object Z was observed').

The *Filtering & Collection* role filters and aggregates the raw events using defined rules. The *Filtering & Collection Ale Interface* (EPCglobal, 2009b) specifies the delivery of cleansed event data to the *EPCIS Capturing Application*. This event data contains the following information: ('At Logical Reader X between time t1 and t2 a set of products with EPCs were observed'). Unlike raw event data, the set of EPCs does not contain duplicates.

The *EPCIS Capturing Application* manages the operations of readers and provides a context for the event data. In this work, environmental data can also be included in event data by using the extension mechanism provided by the EPC Information Service specification (EPCIS). For example, in a warehouse the electricity usage information is gathered from a smart metering sensor and the amount of objects in the warehouse from RFID readings.

EPCIS Interfaces are used to provide event data to the enterprise level roles like *EPCIS Accessing Applications*. In This level, the event data is meaningful for business applications. For example: 'At warehouse X between time t1 and t2 a set of products with EPCs were observed'. In our running example, the basic event data is extended with the life cycle information – data about environmental burden. 'At warehouse X between time t1 and t2 a set of products with EPCs were observed, with allocated set of elementary flows'. For example (Diesel, input, 100 liters; CO2, output, 300 kg).

The *EPCIS Repository* is used to store the RFID event data from the EPCIS Capturing Applications, and EPCIS Accessing Applications are different ERP and MES systems responsible for executing business processes like warehouse management and order handling aided with the acquired EPC data.

EPCIS Accessing Applications are ERP and MES systems than manage the business processes aided by traceability data. The traceability data is used, for example, in warehouse management and logistics.

The address of the EPC Information Services can be made available to other stakeholders by implementing the local EPCglobal *Object Name Service* (ONS) (EPCglobal, 2008a). The *Local ONS* is a service that answers the lookup requests for data about objects. The services are found using the root ONS that uses the domain name system DSN for looking up the local ONS services. Another option – not yet available – is to use the *Discovery Services*, a service standard under development by EPCglobal. The Discovery Service will provide a service to find all the EPC Information Services that have information about a specific product, also ensuring the authorization and authentication of requests, and to share the data in a secure way between stakeholders who do not necessarily know each other. This enables stakeholders to discover all services containing life cycle information about the products (who has information about this product & where the EPCIS is located, where to query life cycle information about the product).

7.2 Sharing Life Cycle Data

In order to handle the object transformation (division or composition) in the supply chain, the stakeholder responsible for the transformation part of the supply chain is also responsible for aggregating the life cycle information from up to that point. In other words, when a manufacturer is further processing products, this manufacturer is responsible for calling the EPC Information Services of a supplier and adding this information (derived attribute in the traceability graph) to the life cycle information of the further processed product. For example, a saw mill company would call the harvesting and logistics companies' EPCIS Query Interface to get the logs' life cycle information. Then the saw mill would cumulate this information to the life cycle information of manufactured boards. Then, the glued laminated timber manufacturer would call the saw mills' EPCIS Query interface to get the environmental information of the boards used for manufacturing the glued laminated timber. In this way the total environmental impact of a product can be made available to end customers.

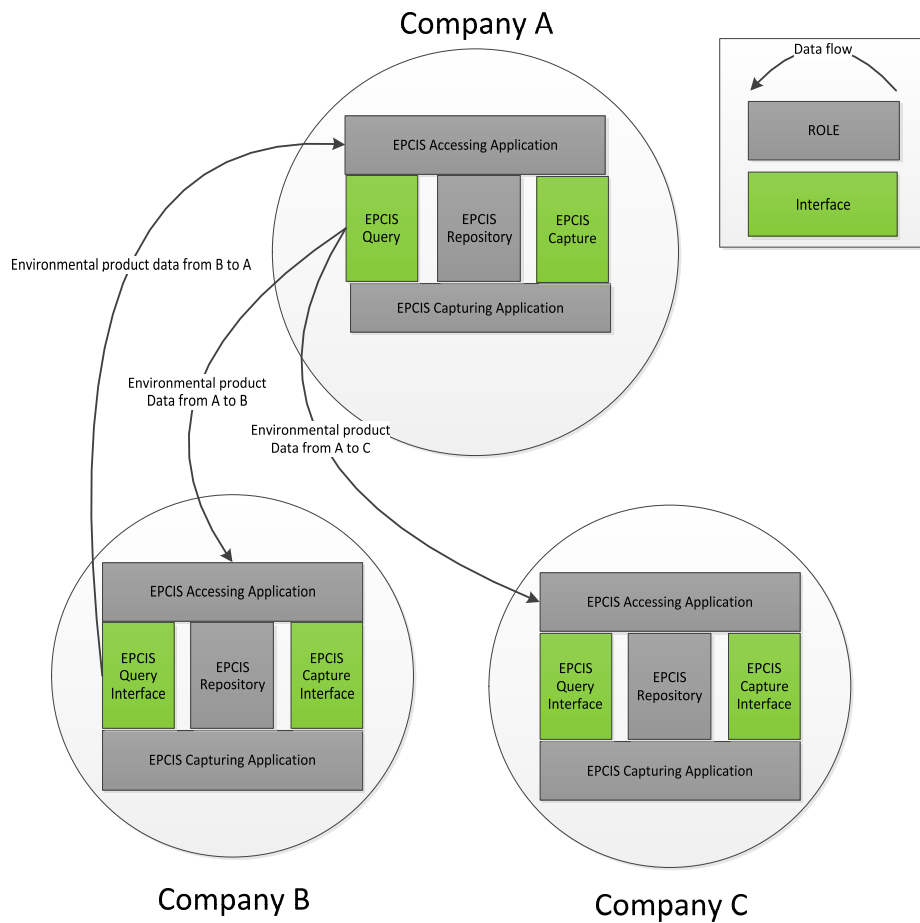


Figure 17. Exchanging life cycle information of a product within the EPCglobal architecture.

In Figure 17, the communication of product data is illustrated in a scenario where three companies exchange the life cycle information of products with each other by utilizing the EPCIS architecture. Each company has an EPCIS-compatible system realizing the roles defined in the EPCIS specification. Company A takes the EPCIS Accessing Application role when requesting environmental product data from Company B by calling the *EPCIS Query Interface*.

Figure 18 illustrates the life cycle data sharing between stakeholders in the supply chain from suppliers to consumers – for example, consumers could use a mobile phone as an EPC Accessing Application as suggested in (Dada et. al., 2010).

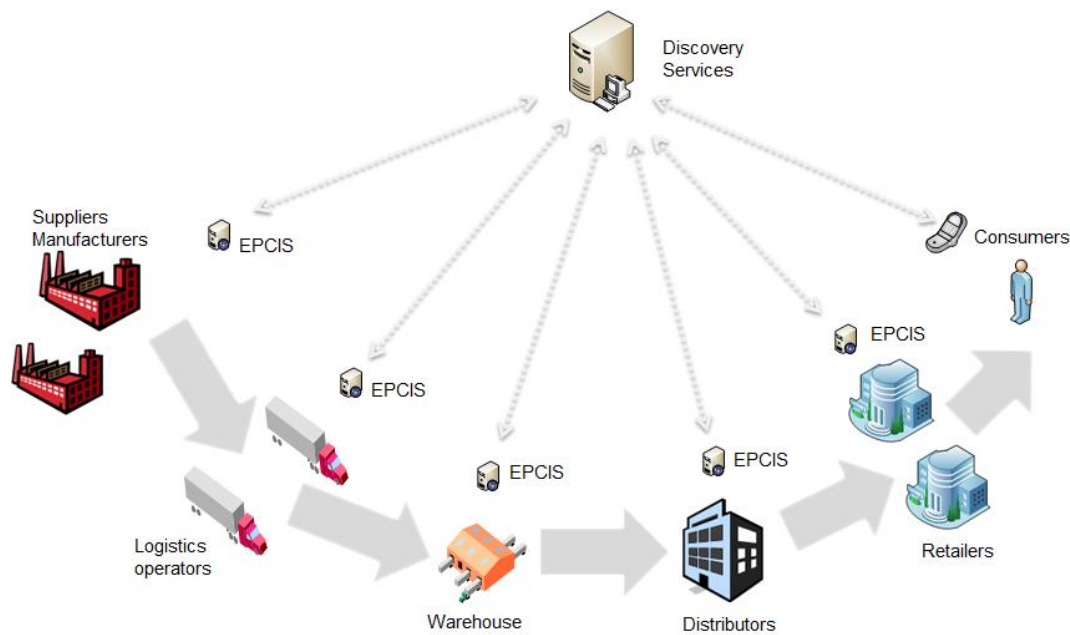


Figure 18. Sharing product information in EPCNetwork.

The main challenge of the system described above is the precision of underlying environmental data. In the optimal situation, all the products in the world would be traced through the full supply chain where all processes have smart metering systems for measuring the emissions generated when manufacturing products and this information would be shared through EPC information services. However, this is not realistic as a starting point.

As discussed in (Usva et. al., 2009), there is a need for a modular approach where the starting point is to use the environmental impact values produced by current approaches, varying from expert judgments and partial estimates to the usage of aggregated product group data generated from national input-output tables or life-cycle assessment data calculated using averages. These default values would be used when a part of the supply chain does not have environmental monitored traceability data available. These values should be defined so that the stakeholders would be encouraged to produce more accurate data and share it through the EPC network. This is achieved by using values slightly higher than the industry average. The same method should also be used inside the companies. The first process nodes to be monitored in real time using traceability and smart metering systems should be the ones that generate the biggest emissions and have greater variation between individual products.

For example, the transport distance of products varies considerably between product instances. Transport distance monitoring can be implemented by using the distance information acquired from the logistic planning system or, alternatively, by using the traceability information from the RFID system as described in (Dada et. al., 2010). In this

method, the transportation distance is acquired using the business locations of shipping and receiving events. The number of transported items is collected from EPC event data. The mass of items is acquired from an organization measurement system. The vehicle type used is acquired by mapping the mobile reader identification for a specific transport vehicle. For example, a reader is located in the articulated lorry (40 tons maximum capacity, EURO4 emission standard). In further steps, such as sawing, we can use the average values for all logs in a manner similar to the traditional life cycle assessment, because the emissions do not vary significantly between different objects.

7.3 Taxonomication

The traceability data management in EPCGlobal network architecture is a Grid computing system where each stakeholder has a service providing life cycle information about products.

The system can be characterized based on a Grid taxonomy (Yu and Buyya, 2005) as follows: The Grid taxonomy consists of five elements, workflow design, information retrieval, workflow scheduling, fault tolerance and data movement.

The temporal relation between the steps is defined by workflow, a structure that can be sequential, parallel, choice or iteration. The workflow model can be abstract or concrete depending on whether the specific resources of the Grid system are bound to a specific workflow task. The third aspect of workflow design is how the workflow model is built. In a user directed model, the user models the workflow using some modeling tools or languages. Automatic Grid systems provide the workflow model based on the data tokens.

The workflow structure of the infrastructure for sharing life cycle information is a *directed acyclic graph* where each task is performed when its conditions are true (*choice*). The model of the workflow is *concrete*; each product is handled by a specific stakeholder. The composition system for the traceability graph is *automatic*. The workflow – traceability graph - is composed based on the actual object movements through the supply chain.

The most important quality-of-service constraint for life cycle data sharing is *security*. The data acquired from the system must be trustworthy and the handling must be confidential if needed. These security constraints are specified at the *task level*.

The Grid taxonomy specifies three dimensions of information retrieval: *static information* related to architecture, *historical information* related to previous events occurred; and *dynamic information* about Grid resources are all managed by Discover Service in the EPCGlobal network that shares the location from EPCIS systems containing life cycle information about a specific object.

Scheduling is the most important concept for enabling scalability in Grid systems. The scheduling architecture can be divided into three types: centralized, hierarchical or decentralized. In centralized and hierarchical systems there is always one main scheduler that schedules the workflow steps. In decentralized systems each scheduler or service can communicate with each other and schedule sub work flows. The workflow scheduling architecture in EPCglobal network is *decentralized*, all the steps in the supply chain manage themselves as well as make *local* decisions about scheduling.

The planning scheme can be considered to be *Just in-time*; the data is queried when the EPCIS Accessing Applications queries the data from the supply chain through Discovery Service. A task scheduling strategy is not applicable to our example, because all the steps containing information about a product must be reached in order to be able to collect the total life cycle information about products. The same applies to performance estimation.

The Grid systems must be able to handle the concurrency and failures in workflows. There are many different methods of achieving the fault tolerance like retry, alternate resource, restart, replication, alternate task, redundancy, user defined exception handling and workflow recovery. Methods either mask the effect of the failure or alter the workflow steps to deal with the failure. Fault tolerance in EPCGlobal Network can be managed at the task level by *retrying* until the service providing the life cycle information becomes available.

Automatic data transfer in a Grid system can be centralized, mediated or peer-to-peer. The input and output data files must be staged to calculation steps in Grid. In the centralized approach, data is moved between calculation steps using a central location. In mediated approach, the data is managed by distributed data management systems and in peer-to-peer systems data is moved between calculation steps, saving time and resources. The intermediate data movement in EPCglobal network is *mediated*, with life cycle information managed by supply chain stakeholders, each collecting data about their part of the supply chain.

8. Papers

The dissertation consists of five articles. The first paper presents the data model for a traceability system in the forestry wood supply chain. The second paper presents a model that generalizes the model presented in Paper I and enables the allocation of life cycle information throughout the supply chain to the traced product. The third paper specifies the logical database schema for the model presented in Paper II. The fourth paper describes methods for querying and analyzing the product life cycle information. The methods are built on the database implementation presented in Paper III. The fifth paper presents a case study where the traceability graph model is used as a basis for creating an IT infrastructure for sharing the life cycle information of a product through supply chain stakeholders. The case study implements the model and methods presented in papers II, III and IV.

Paper I

Modeling traceability in the forestry wood supply chain describes how an automatic traceability system makes it possible to use raw material information throughout the forestry wood production system. RFID technology can achieve automatic traceability by enabling us to connect the physical world objects with their virtual counterparts. This traceability system needs to have a temporal data model to support the tracking of the raw material and monitoring of the processes. In this paper, such a data model for the traceability system of the forestry wood supply chain is presented. The model extends the usual RFID data model so that it supports tracing the evolution of the forest industry products and connects the process information to the product information, thus allowing the calculation of environmental and economical key performance indicators for the supply chain.

Paper II

Traceability Graph to Lifecycle Data Management focuses on how products travel through the supply chain and how the data about products is logically stored. The paper presents a model that generalizes the model presented in Paper I and enables the allocation of life cycle information throughout the supply chain to the traced product.

The model – traceability graph - enables tracing, monitoring, analyzing and querying the properties of processes and their mutual relationships. The formal specification allows services to handle the product's life cycle data formally. The model ensures correspondence between logical objects with real-life products of processes. i.e. it integrates the provenance of products and processes into traditional graph-based workflows. The approach supports attribute value propagation and aggregation in the supply chains. Input and output costs of

processes can be allocated to products, which enables tracing and analyzing these costs precisely.

The model can be applied to single products as well as larger batches. Unlike existing methods, the Traceability Graph enables the precise calculation of the input costs (e.g. recourses) and output costs (e.g. emissions and waste) of products and processes. So far, these calculations have been based on average values from a large set of processes. The model also has the ability to manipulate transformations of the products.

Paper III

Using RFID for Tracing Cumulated Resources and Emissions in a Supply Chain focuses on the logical storing structures of products' life cycle data into a workflow model, which allows storing data associated with processes and products and allocating the data to specific objects, i.e. information – such as the originality, resources and emissions - related to products and their supply chain. Paper III specifies the logical database schema for the model presented in Paper II.

This Paper pays attention to the correspondence between the data management model and real-world applications. That is, the objects are identified on both levels explicitly. In databases, objects are identified by a database solution, whereas in the real world physical RFID tagging is used. The present approach enables tracing product portions as well as single objects. This means that, for example, environmental burden such as greenhouse gases can be analyzed for different products and supply chains. The traceability graph is mapped to relational databases and sample analyzing possibilities are demonstrated.

Paper IV

Multidimensional Analysis of Supply Chain Environmental Performance describes methods for querying and analyzing product life cycle information. The methods are built on the database implementation presented in Paper III. Methods normally used today have problems in accurately quantifying the environmental burden of an individual product, because the supply chains are dynamic. In this paper we present a model that enables calculating and monitoring the environmental performance of products at the item level in a dynamic supply chain and performing a multidimensional analysis of environmental data using the *traceability cube*.

The model presents how emissions and resources can be monitored from the data management perspective. The model can be mapped to any precision level of physical tracing. At the most precise level, even a single physical object and its components can be

analyzed. This, of course, requires that the related objects and their components are identified and mapped to the database. From the opposite perspective, our model also supports the rough-level analysis of products and their histories. It was presented how a multidimensional analysis can be applied to life-cycle information based on the traceability graph.

Paper V

Monitoring Environmental Performance of the Forestry Supply Chain Using RFID presents a case study where the traceability graph model is used as a basis for creating an IT infrastructure for sharing the life cycle information of a product through supply chain stakeholders. The case study implements the model and methods presented in papers II, III and IV

The paper describes novel technology for traceability in the forest and wood industry; pulping-compatible UHF transponders for marking logs, robust RFID readers for harvesters and saw mills. RFID marking connects the physical objects with their database counterparts, thus allowing the automatic tracing of the objects. The software architecture for acquiring and sharing the traceability information was designed and implemented. The Traceability Services connect the steps of the supply chain together and allows new methods for analyzing the performance of the whole chain and any process within it. The developed architecture, which is based on EPCGlobal architecture, allows sharing traceability data within and across enterprises.

9. Discussion

This thesis gives a theoretical model for tracing the life cycle information of even a single product. The model also supports the tracing of patches of different granularities. Further, the thesis investigates how this model can be implemented in relational databases for storing the life cycle information. Then, a multidimensional model that can be used for analyzing this information is presented. The thesis also investigates the architecture for sharing life cycle information between all supply chain stakeholders. In tracing product instances, they must be identified by physical identifiers, which are then assigned to logical identities in databases.

The tracing of objects requires strict correspondence between objects at the logical and physical levels. In some applications, the marking can be made at the patch level if a set of physical products have similar features and supply histories. Now, also one logical object identity can be used to refer to this identifier. In other words, one logical identity and physical identifier can be used for referring to a set of physical products in some processes in a supply chain. In general, the traceability graph can be used for different granularities of product portions. The analysis precision depends on the granularity of the marking used at the physical level. It is also possible to do the marking with different frequencies. For example, in some cases it could be more reasonable to mark only the chosen objects and track the chain through estimates.

For physical marking, there are various methods in use: Imprinting, the finger print method, Laser marking, Label marking, Ink jet marking and transponder marking, but the thesis focus on RFID. However, the presented logical manipulation approach is independent of the physical marking method. In the transformation of objects (division or composition), old physical identifiers have to be replaced by new ones. This may be a critical point when aiming to utilize the presented approach in global supply chains. This is an issue for further studies. From the perspective of real enterprise systems, the balance of marking costs and the precision of tracing should be optimized. The extent of benefits achieved is dependent on the supply chain stakeholder engagement and interest in making use of the developed technology.

In order to present the data coherently and reliably, there has to be a standard set of rules which are enforced by an external auditing party. The ISO 14025 (International Organization for Standardization, 2006) specifies methods and programs for creating environmental declarations. One program operator is an international EPD system which maintains a specific set of product category rules that provide specifications for creating an environmental declaration for products. The rules could be extended so that the

environmental product declaration would be monitored in terms of real-time traceability-based environmental data. When the environmental impact of the whole supply chain is visible to the customers in a reliable and visible way, they can make educated choices to select the product with the lowest environmental impact, which would encourage companies to produce such data for their part of the supply chain and to optimize their production in a sustainable way.

10. Conclusions

The presented formal data-centric workflow model - called the traceability graph - integrates data-centric aspects of products and processes to traditional graph-based workflows. The approach supports the accumulation and aggregation of product life cycle data throughout the supply chain. This allows storing and analyzing life cycle information from the data management perspective. The model can be mapped to any precision level of physical tracing. At the most precise level, even a single physical object and its components can be analyzed. This, of course, requires that the related objects and their components are identified and mapped to the database. From the opposite perspective, the model also supports the rough-level analysis of products and their histories. The traceability graph is a general method for modeling supply processes, although our real-life example was from the forest industry.

The dissertation specifies the logical database schema for storing products' life cycle data and describes how a multidimensional analysis can be applied to the life cycle data based on the traceability graph using the traceability cube. The approach presented in the dissertation supports the different phases of the life cycle assessment as follows: The traceability graph is used to model the processes (LCA phase 1). The traceability data is collected to a database based on the graph (LCA phase 2). The traceability cube is used to represent the impact assessment and analyzing the results (LCA phases 3 and 4).

Based on the model, the presented infrastructure allows managing the life cycle data of products in a complex and dynamic supply chain in an item level. The system can be used to improve the life cycle data management of the products. The environmental impact calculation is used as an example, but the same methodology can be used for other life cycle management purposes, as well. For example, item-level traceability information would enable improved tracing of origin, more efficient call backs common in food and automotive industries and improved security against counterfeiting.

The dissertation presents an implementation of EPC Information Service which is used for storing and sharing information about the objects between the stakeholders of the supply chain. It is also proposed how this information can be shared between the supply chain stakeholders using EPC network.

The model presented can be used for creating a reliable system for measuring the environmental impact of a single product and to present this information to customers so that they are able to choose the products with the lowest environmental burden.

The dissertation consists of five papers. The first paper presents the data model for a traceability system in the forestry wood supply chain. The second paper presents a model that generalizes the model presented in Paper I and enables the allocation of life cycle information throughout the supply chain to the traced product. The third paper specifies the logical database schema for the model presented in Paper II. The fourth paper describes methods for querying and analyzing the product life cycle information. The methods are built on the database implementation presented in Paper III. The fifth paper presents a case study where the traceability graph model is used as a basis for creating an IT infrastructure for sharing the life cycle information of a product through supply chain stakeholders. The case study implements the model and methods presented in papers II, III and IV. All the papers together form a coherent story of the development of the system for product life cycle data management.

References

- Altintas, I., Berkley, C., Jaeger, E., Jones, M., Ludascher, B. and Mock, S. (2004). Kepler: an extensible system for design and execution of scientific workflows. *16th International Conference on Scientific and Statistical Database Management*, 423-424.
- Altintas, I., Barney, O., Jaeger-Frank, E. (2006). Provenance Collection Support in the Kepler Scientific Workflow System. *International Provenance and Annotation Workshop (IPAW 2006)*, 118-132.
- Ameri, F., Dutta, D. (2005), Product Lifecycle Management: Closing the Knowledge Loops. *Computer-Aided Design & Applications*, 2(5), 577-590.
- Akram, A., Allan, R. and Kewley, J. (2006). Data Centric approach for Workflows. *The Tenth IEEE International Enterprise Distributed Object Computing Conference Workshops (EDOCW 2006)*.
- Attie, P., Singh, M., Sheth, A. and Rusinkiewicz, M. (1993). Specifying and enforcing intertask dependencies. *19th International Conference on Very Large Data Bases (VLDB'93)*, 134-145.
- Barseghian, D., Altintas, I., Jones, M.B., Crawl, D., Potter, N., Gallagher, J., Cornillon, P., Schildhauer, M., Borer, E.T., Seabloom, E.W. and Hosseini, P.R. (2010). Workflows and extensions to the Kepler scientific workflow system to support environmental sensor data access and analysis. *Ecological Informatics* 5(1), 42-50.
- Barker, A. and van Hemert, J.I. (2007). Scientific Workflow: A Survey and Research Directions. *The Third Grid Applications and Middleware Workshop (GAMW'07)*, 746-753.
- Berthomieu, B. and Diaz, M. (1991). Modelling and verification of time dependent systems using Time Petri Nets. *IEEE Transactions on Software Engineering* 17(3), 259-273.
- Bonner, A. J. (1999). Workflow, transactions and datalog. *Proceedings of the Eighteenth ACM Symposium on Principles of Database System (PODS'99)*, 294-305.
- Booch, G., Rumbaugh, J. and Jacobson I. (1999). *The Unified Modeling Language User Guide*. Addison-Wesley, Reading.
- Carbon Trust (2008). *Specification for the assessment of the life cycle greenhouse gas emissions of goods and services*. PAS 2050:2008.

- Chaudhuri, S. and Dayal, U., (1997). An Overview of Data Warehousing and OLAP technology. *ACM SIGMOD Record* 26(1), 65-74.
- Chervenak, A., Foster, I., Kesselman, C., Salisbury, C. and Tuecke, S. (2001). The Data Grid: Towards an Architecture for the Distributed Management and Analysis of Large Scientific Datasets. *Journal of Network and Computer Applications* 23, 187-200.
- Clifford, B., Foster, I., Hategan, M., Stef-Praun, T., Wilde, M. and Zhao, Y. (2008). Tracking provenance in a virtual data grid. *Concurrency and Computation: Practice and Experience* 20(5), 565–575.
- Cohen, S., Boulakia, S. C. and Davidson, S. B. (2006). Towards a model of provenance and user views in scientific workflows. *Third International Workshop in Data Integration in the Life Sciences (DILS 2006)*, 264–279.
- Curcin, V. and Ghanem, M. (2008). Scientific workflow systems - can one size fit all? *5Th Biomedical Engineering Conference (CIBEC 2008)*, 18-20.
- Dada, A., Rau, A., Konkel, M., Staake, T. and Fleisch, E. (2010). The Potential of the EPC Network to Monitor and Manage the Carbon Footprint of Products: Dynamic Carbon Footprint Demonstrators. *Auto-ID Labs White Paper WP-BIZAPP-054*.
- Dada, A. and Staake, T. (2008). Carbon Footprints from Enterprises to Product Instances: The Potential of the EPC Network. *GI Jahrestagung, LNI 134*, 873-878.
- DiCesare, F., Harhalakis, G., Proth J.M., Silva, M. and Vernadat, F.B. (1993). *Practice of Petri Nets in Manufacturing*. Chapman and Hall.
- Ellis, C. A. and Nutt, G. J. (1980). Office Information Systems and Computer Science. *ACM Comp. Surv.* (12), 27-60.
- EPCglobal. (2007). EPC Information Services (EPCIS) Version 1.0.1 Specification, EPCglobal Ratified Standard. www.epcglobalinc.org/standards/epcis/epcis_1_0_1-standard-20070921.pdf. [Visited 2011-10-01].
- EPCglobal. (2008a). Object Naming Service (ONS) Standard Version 1.0.1, EPCglobal ratified standard, http://www.epcglobalinc.org/standards/ons/ons_1_0_1-standard-20080529.pdf. [Visited 2011-10-01]
- EPCglobal. (2008b). Tag Data Standard v. 1.4, EPCglobal Ratified Standard, June 2008. http://www.epcglobalinc.org/standards/tds/tds_1_4-standard-20080611.pdf. [Visited 2011-10-01]

- EPCglobal. (2009a). EPCglobal Architecture Framework Version 1.3, http://www.epcglobalinc.org/standards/architecture/architecture_1_3-framework-20090319.pdf. [Visited 2011-10-01]
- EPCglobal. (2009b). Application Level Events (ALE) Standard Version 1.1.1, EPCglobal ratified standard, http://www.epcglobalinc.org/standards/ale/ale_1_1_1-standard-core-20090313.pdf. [Visited 2011-10-01]
- Finkenzeller, K. (2004). *RFID Handbook Fundamentals and Applications in Contactless Smart Cards and Identification*, 2nd ed., Wiley.
- Floerkemeier, C. and Lampe, M. (2005). RFID middleware design - addressing both application needs and RFID constraints. *GI Jahrestagung* 1, 277-281.
- Foster, I., Kesselman, C. and Tuecke, S. (2001). The Anatomy of the Grid: Enabling Scalable Virtual Organizations. *International J. Supercomputer Applications* 15(3), 200-222.
- Freire, J., Koop, D., Santos, E. and Silva, C. (2008). Provenance for computational tasks: A survey. *Computing in Science & Engineering* 10(3), 11-21.
- Gonzalez, H., Han, J. and Li, X. (2006). FlowCube: Constructing RFID FlowCubes for Multi-Dimensional Analysis of Commodity Flows. *Proceedings of the 32nd international conference on very large data bases*, 834-845.
- Gonzalez, H., Han, J., Li, X. and Klabjan, D. (2006b), Warehousing and analyzing massive RFID data sets. *Proceedings of the 22nd International Conference on Data Engineering*, 83.
- Grieves, M. (2006). *Product Lifecycle Management – Driving the next generation of Lean thinking*. The McGraw-Hill Companies.
- Harrison, M. (2003). EPC information service - Data Model and Queries. *White paper, Auto-ID Centre for Manufacturing*, University of Cambridge.
- Häkli, J., Jaakkola, K., Nummila, K., Nilsson, Å., Puntanen, V. and Sirkka, A. (2010). Usage of RFID in the Forest & Wood Industry and Contribution to Environmental Protection. *Vision and Challenges for Realising the Internet of Things - IoT Clusterbook*. European Commission. 165-173.
- Ilic, A., Staake, T. and Fleisch, E. (2009). Using Sensor Information to Reduce the Carbon Footprint of Perishable Goods. *IEEE Pervasive Computing* 8(1). 22-29.

International Organization for Standardization. (1997). ISO 14040, Environmental Management - Life Cycle Assessment - Principles and Framework, ISO 14040:1997(e). International Organization for Standardization.

International Organization for Standardization (2006). ISO 14025 - Environmental labels and declarations -- Type III environmental declarations -- Principles and procedures”, ISO 14025:2006(e). International Organization for Standardization.

International Organization for Standardization. (2011). ISO 14067 - Carbon Footprint of Products, ISO 14067:2011(e). International Organization for Standardization.

Niemi, T., Junkkari, M., Järvelin, K. and Viita S. (2004). Advanced query language for manipulating complex entities. *Information Processing and Management*. 40(6), 869-889.

Junkkari, M. (2005) PSE: An Object-Oriented Representation for Modeling and Managing Part-of Relationships. *Journal of Intelligent Information Systems* 25(2), 131-157.

Kifer, M., Lausen, G. and Wu, J. (1995). Logical foundations of object-oriented and frame-based languages. *Journal of ACM* (42), 741-843.

Kosinski, P. R. (1973). A data flow language for operating systems programming. *SIGPLAN Notices* (8), 89-94.

Krompass, S., Aulbach, S. and Kemper, A (2007). Data Staging for OLAP- and OLTP-Applications on RFID Data. *Proceedings of Datenbanksysteme in Business, Technologie und Web (BTW 2007)*, 542-561.

PE International. (2002). Process data set: Process data set: Spruce wood; timber; production mix, at saw mill; 40% water content (en). PE International, <http://lca.jrc.ec.europa.eu/lcainfohub/dataset2.vm?id=49>. [Visited 2011-10-01]

Pennington, D., Higgins, D., Peterson, A.T., Jones, M.B., Ludaescher, B. and Bowers, S. (2007). Ecological niche modeling using the Kepler workflow system. *Workflows for eScience: Scientific Workflows for Grids, Chapter 8*. Springer.

Petri, C.A. (1966). Communication with automata. *DTIC Research Report AD0630125*.

Prodonoff Jr., V., Da Silva, S.B., Hwang, Y.M. and Mitchell, R. (2008). ROI from Auto-ID Deployment in Aerospace Logistics. *Aerospace ID publications*. www.aero-id.org/research_reports/AEROID-CAM-018-ROI.pdf [Visited 2011-10-15]

- Murata, T. (1989). Petri nets: properties, analysis and applications, *Proceedings of the IEEE* 77(4), 541-80.
- Simmhan, Y. L., Plale, B. and Gannon, D. (2005). A survey of data provenance in e-science. *SIGMOD Record* 34(3). 31–36.
- Sirkka, A. and Junkkari, M. (2010). Data management framework for monitoring and analyzing the environmental performance, *ICGREEN 2010, Proceedings of INNOV 2010*, 57-62.
- Stark, J. (2011), *Product Lifecycle Management – 21st Century Paradigm for Product Realisation*, Springer-Verlag, London.
- Stockman, H. (1948). Communications by means of reflected power, *Proceeding of the IRE* 36(10), 1196–1204.
- Sudarshan S. Chawathe, V. Krishnamurthy, S. Ramachandran, and S. Sarma, Managing RFID Data. *Proceedings of the Thirtieth international conference on Very large data bases (VLDB'04)*, 1189-1195.
- Torlone, R. (2003). Conceptual Multidimensional Models. *Multidimensional Databases: Problems and Solutions*. Idea Group Publishing, 69-90.
- Usva, K., Hongisto, M., Saarinen, M., Katajajuuri, J.M., Nissinen, A., Perrels, A., Nurmi, A., Kurppa, P. and Koskela, S. (2009). Towards certifies carbon footprints of products – a road map for data production. *VATT Research Reports 143:2*.
- van der Aalst, W. (1998). The application of Petri nets to workflow management. *J. of Circuits, Syst. and Computers* (8), 21-66.
- van der Aalst, W. and van Hee, K. (2002). *Workflow Management, Models, Methods, and Systems*. The MIT Press, Cambridge.
- van der Aalst, W., Ter Hofstede, A. H. M. (2005). YAWL: Yet another workflow language. *Information Systems* 30(4), 245-275.
- Wang, F., Liu, S. and Liu P. (2010). A Temporal RFID Data Model for Querying Physical Objects. *Journal of Pervasive and Mobile Computing* 6(3). 382-397.
- Yu J., Buyya, R. (2005). A Taxonomy of Workflow Management Systems for Grid Computing. *Technical Report, GRIDS-TR-2005-1*, Grid Computing and Distributed Systems Laboratory.

Modelling Traceability in the Forestry Wood Supply Chain

Antti Sirkka

University of Tampere
Kalevantie 4, 33014 Tampere, Finland
antti.sirkka@uta.fi

TietoEnator Forest & Energy Oy
Häkiläpolku 3 A, 33101 Tampere, Finland
antti.sirkka@tietoenerator.com

Abstract—Equivalent of € billion of wood raw material is going to waste in Europe. The reason is that information regarding the raw material is not available throughout the system. An automatic traceability system makes it possible to use the raw material information throughout the forestry wood production system. The RFID-technology can achieve automatic traceability by enabling us to connect the physical world objects with their virtual counterparts. This traceability system needs to have a temporal data model to support tracking of the raw material and monitoring of the processes. In this paper, that data model for the traceability system of the forestry wood supply chain is presented.

I. INTRODUCTION

The forestry wood supply chain can be defined as a network of autonomous or semiautonomous business entities collectively responsible for procurement, manufacturing and distribution activities of the product. In the supply chain different business entities are highly dependent on each other. The performance of the supply chain depends highly on the performance of each business entity. One pressing problem in several supply chains is the non-optimal use of resources. This can be solved by introducing the real-time traceability solution for the material flow.

The forestry wood supply chain has a concern with allocating the right raw material for the right final product. This is caused by the following facts: Firstly, the supply chain is not continuous and consists of many steps. Secondly, the biological raw material is very complex. Accurate data is collected during every step but most of this data is lost later in the supply chain. The core problem is to acquire product and production information for each item and enable utilisation of this information through the supply chain.

Nowadays wood is treated as bulk material in the forestry wood supply chain and it has to be graded in the final stages of the production process to find out if it meets the demands of the customer. However, we can define quality attributes needed for different end products. By using the RFID technology to identify the individual objects in the supply chain we can enable the pull model for the wood material.

Traceability improves control over the flow of wood and the follow-up of the processes of the supply chain. The more accurate and complete information about the wood origin and the processes allows big improvements in the supply chain. For example in the window frame manufacturing we can maximise the yield if we minimise the amount of knots in the boards. This can be achieved by selecting trees with high

internode length to be harvested and directed to the window frame manufacturer.

There are some previous researches made about RFID data modelling. Liu & Wang developed a model in [1] and it was developed further in [2-4]. Nguen et al. presented their model which includes business transactions in [5]. Harrison presented categories of RFID data and some query examples in [6]. However, these models do not take into account the specific nature of wood products production where the raw material evolves throughout the supply chain.

In this work the data model presented in these studies is extended to be able to handle the evolution of the wood objects. The presented model also includes the connection between the process parameters of the supply chain and the flow of the wood objects. This enables the traceability of the processes.

II. MODELLING FORESTRY WOOD SUPPLY CHAIN

The data concerning the raw material flow in the forestry wood supply chain is by nature temporal and dynamic. The individual associated data is collected throughout the supply chain using smart identification techniques. The individual associated data is a data concerning all the wood objects that have been used to create an upgraded product. For example if the upgraded product is a window frame the evolution path for the window frame is tree(s), log(s), board(s), pieces of board.

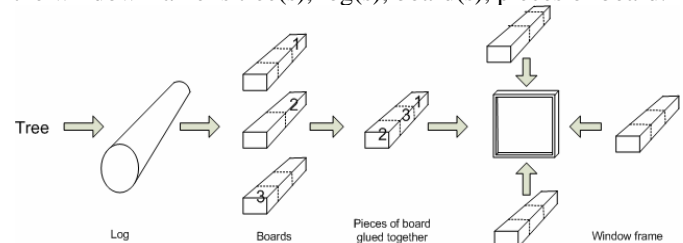


FIG. 1. TRACEABILITY CHALLENGE.

In the forestry wood supply chain information can be divided into dynamic and static information. Static information is information about processes, locations and entity types. Dynamic information is information about individual objects (tree-log-board-upgraded product) and process measurements (energy used, waste produced).

When researching different forestry wood supply chains we can detect the following entities that can be used to model all the supply chains which are examined: the object, the process, the observation and the measurement.

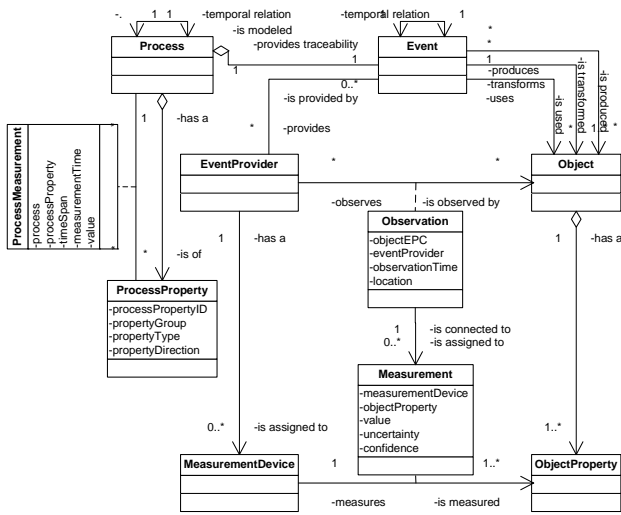


FIG. 2. MODEL OF FORESTRY WOOD TRACEABILITY SYSTEM

The object entity includes both separate objects such as tree, log, board and upgraded product and containment objects like package, kiln patch. The objects have also the parent-child relation for wood object evolution. For example a board is sawn from a log. All the objects have also some properties like moisture and a location. The purpose of the system is to keep track of the locations of the objects. The location can be either geographic (coordinates) or symbolic (the warehouse, the route between A and X). When keeping track of the changes in the system we need to take into account the ramification problem and the qualification problem. The ramification problem deals with the constraints of an action and the qualification problem is used to determine the constraint(s) in which the action is allowed to be executed.

Every distinguishable step in the forestry wood supply chain is called a process which is modelled with event(s). A process has a specific set of properties. There are three types of process properties: informational, input and output properties. The informational properties are not consumed in the process. The input and output properties can be either economical or environmental. Traceability allows connecting these process properties to individual objects and thus we can define all the costs and environmental effects concerning the life cycle of the individual objects. We distinguish three different types of events in the forestry wood supply chain – Usage, Produce and Transform. The usage type of the event uses a wood object. The produce type of the event produces a wood object and the transform type of the event changes the property values of the wood object. The relations between processes and events are modelled with the relations that were defined by Allen [7]. They define the relation between two time intervals and are: Equals, Precedes, Meets, Overlaps, Finishes Contains, Starts and their inverted relations. They cover all the temporal relations between two events.

The meaning of traceability is to track the dynamic interactions between processes and objects, e.g. to track the change of the location property of an object in the supply chain. This is done with observations. An observation is generated when an object is detected with an event provider.

An observation is connected to a measurement that is needed to be able to track the changes of other object properties like moisture. For example: In the log reception the log is observed with an RFID-reader and it is measured with a 3-d scanner, log reception event is then sent to the database.

III. USING THE TRACEABILITY INFORMATION

The traceability information improves control over the flow of wood and the follow-up of the processes of the supply chain. For example, the traceability system stores the information about wood object relations in the following tables.

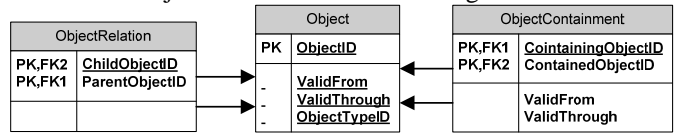


FIG. 3. OBJECT AND OBJECT RELATIONS

The traced wood objects and their life times are stored in the table Object. The table ObjectRelation is used to store object parent/child relations and the table object containment is used to store containment relations. By using these tables the user can fetch the evolution path of the wood object by the recursive query. For example in the Oracle database:

```

SELECT childObjectID
FROM ObjectRelation
START WITH childObjectID = 'EPC_CODE'
CONNECT BY PRIOR parentObjectID = childObjectID

```

By using the wood object traceability data together with the process property measurements the user can calculate the economic and environmental key performance indicators of the process. For example, the user can calculate the energy used in the drying process and correlate it to the amount of boards dried.

```

energy used = SELECT Value FROM ProcessMeasurement
WHERE process = 'KilnDrying' AND
processProperty = 'EnergyUsed' AND
trunc(measurementTime) = to_date('01-JAN-2008')

boards dried = SELECT count(objectEPC) FROM Observation
WHERE location = 'DryingKiln' AND
trunc(observationTime) = to_date('01-JAN-2008')

```

In this work we extend the RFID data model so that it supports tracing of the evolution of the wood objects and connects the process information to the product information, thus allowing the users to calculate environmental and economical key performance indicators for the supply chain.

REFERENCES

- [1] F.Wang, P.Liu. Temporal Management of RFID Data. In VLDB, 2005.
- [2] S. Liu, F. Wang and P. Liu. Integrated Data Modeling for Querying Physical Objects in RFID-Enabled Pervasive Computing. MDM 2007: 140-145
- [3] S. Liu, F. Wang and P. Liu, A Temporal RFID Data Model for Querying Physical Objects, Feb.2007, A Timecenter Technical Report.
- [4] Sudarshan S. Chawathe, V. Krishnamurthy, S. Ramachandran and S. Sarma, Managing RFID Data, in Proc. of VLDB, 2004.
- [5] Tuyen Nguyen, Young-Koo Lee, Rezwanul Huq, Byeong-Soo Jeong and Sungyoung Lee, A Data Model for EPC Information Services, DEWS2007, Japan, March, 2007.
- [6] M. Harrison, EPC information service - Data Model and Queries, White paper, Auto-ID Centre for Manufacturing, University of Cambridge, February 2003.
- [7] J. Allen, Maintaining knowledge about temporal intervals, Communications of the ACM, 26(11): 832-843, 1983.

Marko Junkkari and Antti Sirkka

Formal Definition of Traceability Graph



UNIVERSITY OF TAMPERE
SCHOOL OF INFORMATION SCIENCES
REPORTS IN INFORMATION SCIENCES 3

TAMPERE 2011

UNIVERSITY OF TAMPERE
SCHOOL OF INFORMATION SCIENCES
REPORTS IN INFORMATION SCIENCES 3
NOVEMBER 2011

Marko Junkkari and Antti Sirkka

Formal Definition of Traceability Graph

SCHOOL OF INFORMATION SCIENCES
FIN-33014 UNIVERSITY OF TAMPERE

ISBN 978-951-44-8652-4

ISSN-L 1799-8158

ISSN 1799-8158

Formal Definition of Traceability Graph

Marko Junkkari

School of Information Sciences
University of Tampere
Marko.Junkkari@uta.fi

Antti Sirkka

Tieto Finland
Antti.Sirkka@tieto.com

ABSTRACT

Data-centric workflows focus on how the data is transferred between processes and how it is logically stored. In addition to traditional workflow analysis, these can be applied to monitoring, tracing, and analyzing data in processes and their mutual relationships. In many applications, e.g. manufacturing, the tracing of products thorough entire lifecycle is becoming more and more important. In the present paper we define the *traceability graph* that involves a framework for data that adapts to different levels of precision of tracing. Advanced analyzing requires modeling of data in processes and methods for accumulating resources and emissions thorough the lifecycle of products. The traceability graph enables tracing and accumulation of resources, emissions and other information associated with products. The traceability graph is formally defined by set theory that is an established and exact specification method.

Keywords

Data-centric workflow, data model, lifecycle data management

1. INTRODUCTION

In many applications, e.g. manufacturing, workflows are used to model processes and the relationships among them. Processes are widely studied from the perspective of process modeling but seldom investigated from the perspective of the requirements of data models.

Data-centric workflow [4, 8] approach for designing workflows is based on defining how the data is transferred between processes and how it is logically stored. The approach examines how processes transform data and which entities send and receive the data. The main goal of data-centric workflow is to present the data sets in the workflow. In other words data-centric workflows must involve an integrated data model for storing and manipulating data.

In the present study, we develop a data-centric workflow approach involving such a data model. It supports dynamic data management techniques for data refinement such as aggregation, attribute value propagation, and embedded functions (derived attributes). Integration of object (entity) transformation (object division/composition) with other dynamic data management issues gives an advanced approach to analyze data associated with processes, products the processes yields, and their components. In our terminology product is a general term used to refer to objects resulting from a process. An identified (database) object is an entity represented with its properties.

The problem of focusing on the correspondence between database and real world objects (entities) is as old as the databases in general. Chen [9] defined that the entity is a “thing”, which can be distinctly identified. In object-orientation (see e.g. [7, 10, 24, 26]) an object is a thing that has existence *per se*. The problem of these approaches is how identifying of real world objects corresponds to database objects and vice versa. In practice a database object may represent e.g. 1) a single real word object, 2) a set of real word objects, or 3) a mass of material that can be identified e.g. based on its usage at a time. Furthermore, composed objects, having parts organized at several hierarchy levels, have their own manipulation needs (see e.g. [18, 23, 24]). Data modeling and manipulation in workflows share the complexity of composed objects because transitive relationships must be managed. Further a workflow may structurally correspond to a composed object because it can describe the related composing process.

In our approach a database object may correspond to a real life object, object set or mass of material that can be physically identified thorough the part of process chain in which it is participating. This means that physical objects of a patch are manipulated by a single database object if physical objects can not be individually identified. It is worth noting that in our terminology, the object set means the set of database objects, although a single database object would refer to a set of physical objects.

In data-centric workflows, the manipulation of objects requires specific features because objects may be changed into other objects in both the logical (in databases) and physical (real world) levels. Namely, an object may be divided into other objects or several objects can be composed into a single object. This means that object transformation must be modeled. In addition, objects are manipulated in patches that can be divided into subsets. In turn patches may be collected into larger patches. Table 1 summarizes the cases of object transformation.

Table 1. Identity manipulation

Transformation	Identity	Description
equivalence	maintain	a patch is transferred from a process to another as such
subsetting,	maintain	a patch is divided into subsets for refining
supersetting,	maintain	several patches are collected together for refining
division	change	objects in a patch are divided into several objects
composition	change	objects in patches are composed to single objects

We use the term supply chain, borrowed from manufacturing, to determine all the processes that are participating directly or indirectly in the production of a product. A supply chain is a directed subgraph of a workflow diagram. Structurally, a supply chain consists of supply processes following each other in a partial order, i.e. the result of a process is a raw material for another process. Processes possess properties, such as resources and emissions, associated with the result products of processes. From data-oriented perspective, these properties are accumulated in a supply chain, i.e. the information on the preceding processes of a process is also associated with the process and its products. For example, if we want to calculate used energy of a product, all history (preceding processes) must be taken into account. For modeling this accumulation of products, a process contains two values: ordinal and cumulated where the ordinal value is focused on an underlying

process whereas the cumulated value is aggregated from previous processes. The information from a process to another is sifted by derived attributes.

In data models the intensional (schema) and extensional (instance) levels are typically distinguished. In models of complex structures, such as composed objects, the strict correspondence between these levels is an essential challenge because instances (objects) of an object type may structurally vary from each other. The same concerns many applications of workflows, i.e. the supply chains of similar objects may vary from each other and different data are associated with objects and their components. For management of the structural diversity of composed objects, integrating the intensional and extensional levels with each other is proposed [18]. This allows advanced structural analysis and declarative query formulation thorough transitive relationships [21]. In the present paper, we adopt the idea of integrating the levels into each other, i.e. we do not have explicitly separated workflow schemata and instances. The same concerns the data associated with workflows. This gives a possibility for flexible forming of data-centric workflows.

In the present study we define a data-centric workflow model called the *traceability graph*. In general, the following goals are attached to the traceability graph.

1. **Embed logical storing structures of data into a workflow model.** Data associated with processes and products must be stored within data structures from which it can be search and redefined.
2. **Ability to manipulate objects, object sets and their transformation.** In processes, object sets can be divided into subset or be unionized into larger sets. Single objects can be divided into smaller objects or single objects can be composed from other objects.
3. **Support for querying and analysis of data.** The supply chain with cumulated resources and emissions of a patch or single objects can be derived from a traceability graph.
4. **Ability to manipulate the properties of processes and allocate them to different products.** The traceability graph involves the *cumulated values* of ordinary attributes and the concept of the *derived attribute*. The cumulated value is deduced from attributes based on previous processes. Derived attributes are associated with sifting of products from a process to another. The related calculation rules are based on rations of the amount of products in processes and their transforming among processes.
5. **Formal generality.** In formal specification we use set theory that gives freedom to implement the system in different database paradigms, e.g. relational, object-oriented or deductive data models.
6. **Application independency.** The model is not bound to any specific application area. Our sample application is from the forest industry, but the model can be applied to other domains.

The rest of the paper is organized as follows. In Section 2 we present a short survey on workflow models. In Section 3 we aim to motivate our work by introducing different use cases of data-centric workflows. Among them, we focus on the life cycle assessment because the requirements for strict tracing demands advanced analysis that is applicable to other use cases. Section 4 deals with the graphical notations of the traceability graph and an informal introduction for our sample application domain. The used mathematical notational conventions are given in Section 5 and the formal representation of the traceability graph is given in Section 6. The analytic capabilities of the traceability graph are described in Section 7 and in Section 8 we present problems and solutions to share the data of data-centric workflow with several stakeholders. Finally, the conclusions are given in Section 9.

2. RELATED WORKS

For describing functionality (dynamic aspects) of information systems, various graph based methods have been developed. Early data flow languages e.g. [20] were primarily intended for modeling the activity of computer programs, whereas workflow models [3] are focused on modeling physical processes and activities. Historically, the distinction between dataflow and workflow models is vague and some authors [13] use the term “information flow” for describing flowing of all information from a process to another. Nowadays, different types of activities are typically distinguished in workflow diagrams, like sending, transforming and packing of materials. Some modern modeling methods of information systems, e.g. UML [7], contain diagrams for dynamic aspects of programs (interaction diagrams) and modeling of workflows (activity diagram).

In workflows materials, documents and other information are transferred from a process to another [3, 6]. The workflow model of UML is informally defined and its purpose is to map real world activities to the underlying software solution (or visa versa). There are also a number of commercial applications that have a component for drawing workflow diagrams. Their common feature is that they support the illustration of different types of processes.

There are also formal methods for modeling functionality of information systems that are not primarily intended to any specific purpose. For example Petri nets are traditionally used for defining or describing functionality of computer programs, but they are also proposed for exact representation of workflow models [1], [3]. YAWL [2], Temporal logic [5], and Transaction Logic [6] are other good representatives for formalizing workflows. These methods emphasize timing within processes and supply chains. In general, these are very expressive languages that enable not only modeling of the functionality of systems but also other aspects of information systems. For example Transaction Logic is based on F-logic [20] that is a general framework for specifying object-oriented and deductive aspects. Workflow models have been investigated from the perspective of how they support different data-centric aspects [11]. The main aspects include concentrating on process functionality based on the data, i.e. each node has behavior instructions with regard to the data.

From our perspective, timing representation in YAWL, Temporal logic, and Transaction Logic is a secondary feature. Instead we emphasize handling the data aggregation and movement between processes.

Deutsch and others [12] present an advanced data-centric business processes model and its verification. They define several essential primitives such the artifacts schema, artifact instance, and service logically integrated with each other. Our study differs from their approach. First, instead of separated intensional and extensional levels we integrate these composed data structures. Of course, behind our approach may be a predefined data schema, but the data schema can also be derived from the traceability graph. In other words, we do not specify whether a data schema is predefined or derived. Second, we address the explicit object sifting, transformation and derivation of the information thorough of processes.

We aim to find the principal primitives needed to model and manipulate data-centric aspects in workflows in a way that enables tracing of products at different granularity levels. We do not bind our model to any existing data or workflow methods, i.e. we give freedom to apply our model into existing formalisms and systems.

In the related previous studies we have presented only graphical notations of the traceability graph and its implementation in relational data bases [19]. Based on the relational representation we

have introduced possibilities for OLAP analysis [26]. However, these studies were ad-hoc. Now we give a general formal definition for data-centric workflow model.

3. MOTIVATION

The traceability graph is a data-centric workflow model for tracing, analyzing and querying data. Next we introduce some practical use cases and their background for the Traceability Graph.

3.1 Life Cycle Assessment

The product level environmental impact assessment has become more and more important. The main approach for assessing this are international standards of the life cycle assessment (LCA) (ISO 14040 series [16]) and eco-labels (ISO 14020 [15]) and verification (ISO 14064 [17]). ISO has also started to develop standards for Quantification and Communication of the Carbon footprint of a product (ISO 14067).

The life cycle assessment (LCA) is a standardized method for calculating the environmental burden of a product throughout its lifespan, from raw materials to the disposal/recycle phase. The goal of LCA is to compare the environmental impact caused by a product so that the customer can choose the product that causes least damage to the environment. LCA has four main phases:

In the first phase, the goal and boundaries of the life cycle assessment are defined. In other words processes of the supply chain of the product are defined. In this phase also the functional unit is specified. For example a cubic meter of timber, one mobile phone or one tomato.

In the second phase, illustrated in Figure 1, called life cycle inventory analysis, each process is analyzed and the input and output flows of the processes are defined. There are two basic flow types.

- Elementary flows describe inputs of raw materials and energy resources, and outputs of waste and emissions.
- Product flows describe inputs of products which are an output of other processes, and outputs of products, by-products.

The third phase of LCA is impact assessment where the results of the life cycle inventory analysis are assigned to the environmental impact categories (e.g. Climate change, Ozone depletion and Acidification). For example, the carbon dioxide and methane are greenhouse gases and are assigned to the Climate Change category. The fourth phase is the interpretation where conclusions of the analysis are made.

The traceability graph, defined in the present paper, can be used for tracing and storing the inventory data, the impact assessment and interpretation phase is excluded. Unlike existing methods our model enables analyzing resources and emissions on the single product level – not only average values. Nowadays a common method for calculating the environmental impact is to measure the input and output flows of the whole supply chain during some time period and calculate the average environmental impact for the product (see e.g. [22]).

We have designed our sample system from the perspective of life cycle assessment in order to demonstrate traceability in a forest industry area. We especially aim to demonstrate possibilities that the traceability graph gives for tracing products that are composed from several components, and the components, in turn, have been divided from larger sub-products.

Although we emphasize environmental aspects in our example, the traceability graphs can be applied to other advanced tasks needed in different application domains.

3.2 Origin of Raw Material

The organizations are facing more pressure from consumers and legislators to accurately report the origin of their raw material. However, the complexities of today's supply chains pose a challenge for gathering this data accurately. For example in the forest industry the companies are certifying their product using PEFC¹ chain-of-custody certification. The PEFC chain-of-custody certification is a method for tracing wood from forest to the final product to ensure the wood or wood-fibre can be traced back to certified forest. The certification has two methods of realizing this:

- Percentage based method – the method allows mixing certified and non-certified raw materials taking into account that the percentage of certified raw material must be known. Company can sell as certified the proportion of its production, which equals the proportion of the certified raw material
- Physical separation method – the method requires certified and non-certified raw material to be physically separated throughout the supply chain.

Naturally, the methods can be integrated and improved using the Traceability Graph. The accurate item level information of the origin can be used to create a transparent and trustworthy certification system for origin.

3.3 Recalling Products

The traceability graph enables better and more accurate service for recalling unsafe, defective or hazardous products.

The item level traceability allows manufacturers to recall just the particular products that contain the unsafe elements. For example Toyota recalled approximately 9 million cars in 2009-2010 for pedal entrapment/floor mat problems and accelerator pedal problems. If the reason of recall is for example a faulty set of components, a manufacturer could have only recalled those cars that include the faulty components, not the whole set of manufactured cars in some time period.

The item or patch level traceability can also be used in food industry to avoid total recalls. The traceability graph enables manufacturers, suppliers and resellers to identify the products where some suspected raw material was used and all customers whom the products were delivered. Using the system companies are better equipped to retrieve the affected products and to protect their reputation and brand value.

3.4 Benchmarking

The traceability graph can also be used for benchmarking the processes. By comparing the environmental performance of the same type of processes companies can detect whether the performance metrics of their process are up to the best practices in the industry. This allows companies to notice where they could achieve the biggest improvements in their environmental performance.

¹ <http://www.pefc.org/>

4. INFORMAL DESCRIPTION OF TRACEABILITY GRAPH

The traceability graph can be used to model the supply processes of products and the data associated with the products. The traceability graph has an ability to manage the transformations of the products. For example a product may be manufactured using masses of raw materials or it can be composed from many parts. The traceability graph also has the ability to allocate the properties of processes to the products handled in processes.

Next we introduce the primitives of the traceability graph, a rough graphical representation, and the sample system used from now on.

4.1 Primitives of Traceability Graph

The traceability graph consists of nodes, edges and their properties. A node describes a process whereas an edge describes a flow between processes. Next we introduce data-centric primitives associated with nodes and edges.

- An *object* is a product unit uniquely identified in a supply chain. It may correspond to a single product or a patch or a mass of material, depending on the precision of the actual traceability system.
- A *product portion* determines a patch of products of a node. It can be associated with a set of objects or a mass of non-identified material. The product portion involves a ratio which is used to allocate the costs (e.g. emissions and resources) of the node to which the product portion belongs.
- An *attribute* of a node describes process information – elementary flows. Input attributes represents input costs, e.g. raw materials and resources, used in the process. Output attributes represents output costs, e.g. emissions and waste, generated in the process. For other information of processes the info attributes can be used. Each attribute has two values: an (ordinary) value for the underlying process and a cumulated value which is calculated from the preceding nodes via edges.

As a whole, a node involves an identity, a set of product portions and a set of attributes. An edge is identified by participating nodes, called start and end nodes. Furthermore it contains the following primitives:

- A *sifted product portion* contains products that are sifted from a process to another. A sifted product portion may involve only some objects from the original product portion.
- *Object mapping* belongs to a sifted product portion. It can be equivalence, subsetting, supersetting, division or composition. For example subsetting means that only a part of the product portion is selected for refining. Division means that products of the start node are divided in the end node. Composition means that products of the start node are components for the products of the end node. It is worth noting that unlike in set mappings the identities of objects are changed in division and composition.
- A *derived attribute* determines the quantity of emissions and raw materials associated with the sifted product portions. It is calculated using process specific rules. For example, the volume or the cost of the product can be used as a factor.

4.2 Graphical Notations

The graphical notation for the traceability graph is a rough level description about processes where data-centric primitives are mainly hidden. However, it gives a framework for information transferring among processes and illustrates object transformation (division and composition) between processes.

In the graphical representation, a node is illustrated by a circle, and an edge by an arrow – i.e. we follow the traditional illustration for graphs. Three kinds of edges are notationally distinguished:

- *Plain edge* is presented by a plain arrow and it determines that objects are transformed from the start node to the end node as such. This covers equivalence, subsetting and supersetting mappings of objects.
- *Division edge* represents that the objects of the start node are divided in the end node. In other words, several objects in the end node are mapped to one object in the start node. This is illustrated by a double head arrow.
- *Composition edge* means that objects in the start node are components for an object of the end node. In other words, several objects in the start node are mapped to one object in the end node. This is illustrated by an arrow having divided start.

4.3. Sample System

In terms of our example from the manufacturing and forest industry we demonstrate the use cases presented in Section 3. In Figure 1 a simplified production of glued laminated timber is illustrated. The processes included in the example are felling the trees (harvesting), sawing logs to boards, drying the boards and jointing the glued laminated timbers from boards.

The first phase of production has three instances (nodes), describing the amount of daily harvesting. The movement of the products is illustrated from left to right in the graph. For example, the products resulted from the harvesting nodes are transferred to the sawing node, a double headed arrow illustrates that in the sawing nodes, the objects (logs) from harvesting nodes are divided into several objects (boards).

Between Sawing and Drying nodes the objects are not changed. This is illustrated with a plain arrow. The mapping between the object sets in these cases can be equivalence, subsetting or supersetting. In other words the objects from the preceding node can be moved as such, or only part of the objects can be moved, or all the objects can be moved from the preceding node together with some objects from other nodes.

In the gluing nodes the objects from the drying nodes are composed as an object of the gluing node. In other words objects from the drying nodes are components of objects in the gluing nodes. An arrow with a divided start illustrates this. The emissions and resources associated with the products of the drying nodes are accumulated to a new set of objects in the gluing node.

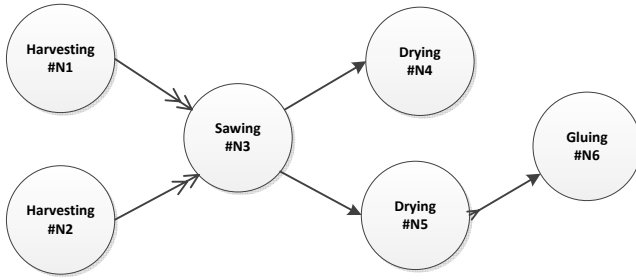


Figure 1. Sample Traceability Graph.

A supply chain of an object can be viewed as a network of processes that are associated with the product during its manufacturing. Using the information of the traceability graph it is possible to track the supply chain of the object throughout its entire supply chain and allocate all the information related to those processes to the object.

Given the running example, we are tracing the environmental burden of the glued laminated timber that belongs to the Gluing #N6 node. Then the supply chain of this glued laminated timber is the processing history of the objects. In Figure 2 the colored nodes are the processes that constitute the supply chain of the paper roll. For example, the glued laminated timber has participated in the nodes Drying#N5, Sawing#N3, Harvesting#N1 and Harvesting#N2. This sub graph of the total traceability graph is the supply chain of the glued laminated timber.

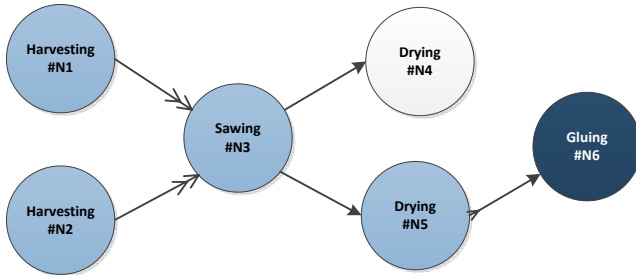


Figure 2. Sample supply chain for the object in node Gluing #N6

In the related formal example we demonstrate the precision of tracing that the tractability graph serves. Before that we introduce the used notational conventions for the definitions.

5. NOTATIONAL CONVENTIONS

Standard set theory is used for representing the traceability graph. Next we introduce only those notational conventions which have widely used alternative representations.

- *Tuple* is an ordered sequence of elements represented between angle brackets. For example $\langle a,b,c \rangle$ is a tuple.
- If t is a tuple and x its uniquely labeled member then $t.x$ refers to x in t . For example if $t = \langle a,b,c \rangle$ then $t.b$ refers to the second member of t .
- If it is not necessary to refer to a member of a tuple the underline space can be used. For example in 3-tuple $\langle _,x,_ \rangle$ the first and last members are not referred.
- The power set of the set S is denoted by $P(S)$
- Cartesian product between sets A and B is denoted by $A \times B$.
- Mapping f from a set X to another set Y is a 2-place relation ($\subseteq X \times Y$) denoted by $f: X \rightarrow Y$. X or Y may be a set consisting of sets, e.g. a power set.

- If R is a 2-place relation $R \subseteq X \times Y$ then the *domain* of R ($\{x \in X \mid \langle x,y \rangle \in R\}$) is denoted by $dom(R)$, whereas the *range* ($\{y \in Y \mid \langle x,y \rangle \in R\}$) is denoted by $rng(R)$.

A (directed) graph is a pair (N,E) where N is a set of *nodes* (vertexes) and E is a set of *edges*. Nodes and edges are represented by set theory as follows:

- A node is represented as a tuple $\langle \text{Node-id}, P_1, \dots, P_n \rangle$, where Node-id is the identity of the node and P_1, \dots, P_n are the properties associated with the node. For brevity, Node-id can be used to refer to the node. Thus, the notation Node-id. P_i refers to the property P_i in the node having the underlying identity.
- A directed edge is represented as a tuple $\langle \text{Node-id}_S, \text{Node-id}_E, P_1, \dots, P_n \rangle$, where Node-id_S and Node-id_E are the identities of the start and end nodes, respectively. P_1, \dots, P_n are the properties associated with the edge.

6. FORMAL REPRESENTATION OF TRACEABILITY GRAPH

In the traceability graph each node describes a process where resources are needed or new costs (e.g. environmental impacts) emerge. A node involves a set of attributes and a set of product portions. These are the properties of the node. An edge describes division, composition or transferring of products portions. Each edge possesses a set of product portions which are shifted to the following process. In an edge neither new resources are needed nor new costs are emerged, i.e. ordinary attributes are not associated with an edge. Instead, an edge may involve *derived attributes* that describe portions of previous product portions and attributes. In other words, *sifted product portions* and derived attributes are properties of an edge. Next we define our model in detail.

Products which are identified physically and logically in the application domain are called *objects*. For logical identifying each object of interest possesses an identity that is either an integer or a string (code used in the application domain at hand). The set of possible identities of an application domain is denoted by ID. In our sample domain $ID = \{id1, id2, \dots, id6000\}$.

In processes different products are manufactured or manipulated. Products can be divided in different portions based on their types or manipulation needs. The product portion is defined as follows:

Definition 1: *Product portion* is a tuple $\langle \text{P-Name}, C, \text{ID-set}, R \rangle$ where P-Name is the name of product, C is the amount of the portion, ID-set is the set of object identities in the portion, and R is the ratio of the portion related the underlying total amount of the products.

In Definition 1 the ratio R is calculated by some application specific method based on e.g. the weight of the product portion related to the total weight of products in the underlying process, or used time related to total time needed in the process. For a product portion associated with other than objects, ID-set is empty and the portion is manipulated as a mass without interest on individual products.

In following list we have some example product portions related to harvesting:

- $\langle \text{PineSawLog}, 350 \text{ m}^3, \{id1, \dots, id1000\}, 0.60 \rangle$
- $\langle \text{PinePulpWood}, 200 \text{ m}^3, \emptyset, 0.30 \rangle$
- $\langle \text{HarvestingWaste}, 20 \text{ ton}, \emptyset, 0.10 \rangle$

The products PinePulpWood, and HarvestingWaste are manipulated as a mass, i.e. they do not contain object identities. Harvesting waste is manipulated as a product because it can be used to bioenergy. PineSawLog has a set of identities for logs ($\{id1, \dots, id1000\}$), or there are one thousand

logs. This portion has also the biggest ratio value (0.60) which means that it involves 60% of costs of the underlying process.

An attribute describes some information bound to a process. The attributes are divided into the three category based on their nature as follows:

1. *Input attribute* describes costs, used materials and other resources needed in a process. For example the used fuel is an input attribute. In the present approach the input attribute has a numeric value.
2. *Output attribute* describes other matters than products that a process produces. For example, a process may produce some tons of CO-gas. The output attribute has a numeric value.
3. *Info attribute* contains other data or documents associated with a process. The info attribute has a set value.

An attribute involves two values: one for the underlying process (ordinal value) and another for the previous production chain (cumulated value). The cumulated value is derived from previous processes based on the given rules. These are defined after the definitions of primitives needed for them. Until that (Definition 6) we use examples where ordinal and cumulated values are the same.

Next, the attribute is formally defined.

Definition 2: *Attribute* is a tuple $\langle A\text{-Name}, T, V, W \rangle$ where *A-Name* is the name of attribute, *T* is the type of the attribute ($\in \{\text{input}, \text{output}, \text{info}\}$), *V* is the ordinal value of the attribute, and *W* the cumulated value of the attribute.

In the following list we have some example attributes where the ordinal and cumulated values are same.

- $\langle \text{Diesel}, \text{input}, 100 \text{ liters}, 100 \text{ liters} \rangle$
- $\langle \text{CO}_2, \text{output}, 300 \text{ kg}, 300 \text{ kg} \rangle$
- $\langle \text{CompanyCode}, \text{info}, \{111\}, \{111\} \rangle$

In the example “Diesel” is an input attribute describing that one hundred liters diesel is used (*V* value), “CO₂” is an output attribute describing that the process caused 300 kilograms of carbon dioxide emission, and the company code 111 indicates a manufacturer associated with the process.

Now we are able to define the process node involving product portions and attributes in Definition 3.

Definition 3. *Process node* (simply node) is a tuple $\langle \text{Nid}, \text{N-type}, \text{P-set}, \text{A-set} \rangle$ where *Nid* is the identity of the node, *N-type* is the type of the process, *P-set* is the set of product portions and *A-set* is the set of attributes associated with the node.

A harvesting node (*Nid* = 1) involving product portions and attributes is given below. From now on we mark by boldface numeric attribute values and ratios needed for cumulating forthcoming values in the supply chain.

$\langle 1, \text{Harvesting},$
 $\{\langle \text{PineSawLog}, 350 \text{ m}^3, \{\text{id1}, \dots, \text{id1000}\}, \mathbf{0.60}\},$
 $\langle \text{PinePulpWood}, 200 \text{ m}^3, \emptyset, 0.30\rangle,$
 $\langle \text{HarvestingWaste}, 20 \text{ ton}, \emptyset, 0.10\rangle\},$
 $\{\langle \text{CarbonDioxide}, \text{output}, 300 \text{ kg}, \mathbf{300 \text{ kg}}\},$
 $\langle \text{Diesel}, \text{input}, 100 \text{ liters}, \mathbf{100 \text{ liters}}\},$
 $\langle \text{CompanyCode}, \text{info}, \{111\}, \{111\}\},$
 $\langle \text{Location}, \text{info}, \{\text{lat } 62.87 - \text{lon } 22.86\}, \{\text{lat } 62.87 - \text{lon } 22.86\}\rangle\}$

In a supply chain, information on previous processes (nodes) is propagated to forthcoming processes. This information consists of object identities and the values of attributes. An *identity shift* determines those objects that are transferred from a process to another or a mapping among objects. There are five types of the identity shift. 1. *Equivalence* means that objects of the start and end nodes are the same. 2. *Subsetting* means that some objects (but not all) are transferred. 3. *Superseting* means that all the objects are transferred but there also are objects from another process. 4. *Composition* means that several objects are composed to single objects. 5. *Division* means that single objects are auto-identification into several objects. Formally the identity shift is defined as follows:

Definition 4: *Identity shift* is a mapping M among identities or the sets consisting of them. The mapping M may be:

1. *equivalence*, where $M:ID \rightarrow ID$ and $dom(M) = rng(M)$
2. *subsetting*, where $M:ID \rightarrow ID$ and $dom(M) \subset rng(M)$
3. *superseting*, where $M:ID \rightarrow ID$ and $dom(M) \supset rng(M)$
4. *division*, where $M:ID \rightarrow P(ID)$ and $\forall Y \in rng(M) \exists x \in dom(M): |Y| > 1$
5. *composition*, where $M:P(ID) \rightarrow ID$ and $\forall y \in rng(M) \exists X \in dom(M): |X| > 1$.

In Definition 4, cases from 1 to 3 maintain the object identities, whereas in cases 4 and 5 object identities are typically changed. In case 4, a single object is mapped to a set of object identities, whereas in 5 a set of object identities is mapped to a single object identity.

For propagating information represented as attributes among nodes, the notation of the derived attribute is used. Attribute value propagation rules are based on the types of attributes, i.e. propagation for input and output attributes requires calculation whereas info attributes are propagated by collecting all the data and documents for forthcoming processes. These rules are involved in the definition of the edge below.

A (shift) edge describes the connection between two nodes. An edge involves a set of derived attributes and a set of *sifted product portions*. A sifted product portion describes products that are shifted from a process to another. In order to maintain a product portion and the related objects, an identity shift is associated with sifted product portions. The edge is formally defined as follows:

Definition 5. *Shift edge* (simply edge) is a tuple $\langle N_S, N_E, SP, D \rangle$, where N_S and N_E are identities of the start and end nodes, SP is a *sifted product portion*, and D is a set of *derived attributes*.

- SP is a tuple $\langle P\text{-Name}, C, M, Rp \rangle$, such that there exists $\langle P\text{-Name}, C', ID\text{-set}_S, _ \rangle \in N_S.P\text{-set}$ and $\langle _, _, ID\text{-set}_E, _ \rangle \in N_E.P\text{-set}$. The mapping M is an identity shift where objects in $dom(M)$ belong to $ID\text{-set}_S$ and objects in $rng(M)$ belong to $ID\text{-set}_E$. C is the amount of the sifted product portion and $Rp = C/C'$ is the ratio of the sifted product portion.
- A derived attribute ($\in D$) is a tuple $\langle A.A\text{-Name}, A.T, DV \rangle$ where $A \in N_S.A\text{-set}$, i.e. $A.A\text{-Name}$ is the name of an attribute in N_S and $A.T$ its type. DV is the value of the derived attribute. It is

$$\begin{cases} A.W, & \text{if } A.T = \text{info} \\ A.W \cdot P.R \cdot SP.Rp & \text{where } P \in N_S.P\text{-set: } P.P\text{-Name} = SP.P\text{-Name, if } A.T \in \{\text{input, output}\} \end{cases}$$

In Definition 5 the sifted product portion is $\langle P\text{-Name}, C, M, Rp \rangle$ where Rp is the ratio of the portion. This is calculated such that the sifted amount C is divided by the original amount C' of the

start node. This ratio is used for calculating the value of derived attributes. A derived attribute is a 3-tuple where the first and second members possess the name and the type of the attribute inferred from the start node. The value of a derived attribute is cumulated as such from the start node if the type of the attribute is info. Otherwise, the original value is multiplied by the ratio of the original product portion (P.R), and by the ratio of the sifted product portion (SP.Rp). Below we introduce these aspects by examples.

In the running example (see Figure 2) there is a division edge between Nodes 1 (a harvesting node) and 3 (a sawing node). This edge describes that 10% of pine saw logs are selected to the sawing node. In the harvesting node (see above) the product portion of the logs possesses the ratio 0.6. This means that the values of input and output attributes must be multiplied by these ratios. Thus, the derived values of carbon dioxide and diesel are $300 \cdot 0.6 \cdot 0.1 = 18$ and $100 \cdot 0.6 \cdot 0.1 = 6$, respectively. The value of the info attribute is propagated as such. The type (division) of the edge means that single objects are divided into the sets of new objects. In the example objects with identities between id3000 and id3200 are balks whereas boards involve the identities between id4000 and id4800. A log is divided into one balk and four boards. In the object level this is described by an identity shift for one log object to a set of objects consisting of one balk and four boards. For example, the identity shift instance $\langle id1, \{id3000, id4000, id4001, id4002, id4003\} \rangle$ means that a log object (id1) is mapped to a balk (id3000) and boards (id4000, id4001, id4002, id4003). The related sample edge is represented as follows:

$\langle 1, 3, \langle \text{PineSawLog}, 35 \text{ m}^3, \{ \langle id1, \{id3000, id4000, id4001, id4002, id4003\} \rangle, \langle id2, \{id3001, id4004, id4005, id4006, id4007\} \rangle, \langle id3, \{id3002, id4008, id4009, id4010, id4011\} \rangle, \dots, \langle id100, \{ \dots \} \rangle \rangle, \mathbf{0.1}, \{ \langle \text{CarbonDioxide}, \text{output}, \mathbf{18} \text{ kg} \rangle, \langle \text{Diesel}, \text{input}, \mathbf{6} \text{ liters} \rangle, \langle \text{CompanyCode}, \text{info}, \{111\} \rangle \} \rangle$

In a traceability graph there are two kinds of nodes based on their roles in the graph. Initial nodes have no predecessors, i.e. there is no edge to them. Other nodes possess at least one predecessor. This distinction is essential because attribute values are cumulated in other nodes than initial ones. In initial nodes an attribute has the same cumulated value as the ordinal value. For attributes in the other nodes, the cumulated value is derived from previous nodes via edges. If the underlying attribute is an info attribute, then the cumulated value is the set consisting of the ordinal value and values of derived attributes in incoming edges. Otherwise it is the sum of the value of the ordinal attribute and the corresponding values of derived attributes in incoming edges. Formally, the cumulated value is defined as follows:

Definition 6: Let A be an attribute in node N , V its ordinal value and $S = \{E | E.N_E = N\}$ the set of immediate incoming edges E to N , then the *cumulated value* W of A is

$$\begin{cases} V \cup \bigcup_{\langle \dots, D \rangle \in S} DV: \langle A, _, DV \rangle \in D, \text{ if } A.T = \text{info} \\ V + \sum_{\langle \dots, D \rangle \in S} DV: \langle A, _, DV \rangle \in D, \text{ if } A.T \in \{\text{input}, \text{output}\} \end{cases}$$

We demonstrate calculation of cumulated values below. Before that we define the traceability graph as follows:

Definition 7: Let N-Set be a set of process nodes and E-Set a set of shift edges, then $\langle N\text{-Set}, E\text{-Set} \rangle$ is a *traceability graph*.

Next we present the rest of our sample traceability graph. Node 1 and the edge between Nodes 1 and 3 are given above. In the example there are two harvesting nodes. In Node 1, logs have identities id_1, \dots, id_{1000} , whereas in Node 2 they have identities $id_{1001}, \dots, id_{1800}$. Ratios and the values of attributes differ in some extend from Node 1.

$\langle 2, \text{Harvesting},$

$\{ \langle \text{PinePulpWood}, 300 \text{ m}^3, \emptyset, 0.45 \rangle,$
 $\langle \text{PineSawLog}, 300 \text{ m}^3, \{id_{1001}, \dots, id_{1800}\}, \mathbf{0.45} \rangle,$
 $\langle \text{HarvestingWaste}, 20 \text{ ton}, \emptyset, 0.10 \rangle \},$
 $\{ \langle \text{CO}_2, \text{output}, 270 \text{ kg}, \mathbf{270 \text{ kg}} \rangle,$
 $\langle \text{Diesel}, \text{input}, 90 \text{ liters}, \mathbf{90 \text{ liters}} \rangle,$
 $\langle \text{CompanyCode}, \text{info}, \{211\}, \{211\} \rangle,$
 $\langle \text{Location}, \text{info}, \{ \text{lat } 65.21 - \text{lon } 21.36 \}, \{ \text{lat } 65.21 - \text{lon } 21.36 \} \rangle \}$

From Nodes 1 and 2 one hundred logs (30 cubic meters) are selected to the underlying sawing process. In the example these logs have identities $id_1 - id_{100}$ (edge from Node 1 to Node 3) and $id_{1001} - id_{1100}$ (edge from Node 2 to Node 3).

$\langle 2, 3, \langle \text{PineSawLog}, 30 \text{ m}^3, \{ \langle id_{1001}, \{id_{3101}, id_{4400}, id_{4401}, id_{4402}, id_{4403} \rangle \}, \dots, \langle id_{1100}, \{ \dots \} \rangle \rangle, \mathbf{0.1} \rangle,$
 $\{ \langle \text{CO}_2, \text{output}, \mathbf{12.15 \text{ kg}} \rangle,$
 $\langle \text{Diesel}, \text{input}, \mathbf{4.05 \text{ liters}} \rangle,$
 $\langle \text{CompanyCode}, \text{info}, \{211\} \rangle \}$

In the edge from Node 2 to Node 3 the values of input and output attributes are calculated such that original attribute is multiplied with the ratio of the corresponding product portion (0.45) and the ratio of shifted product portion (0.1). For example the value of carbon dioxide (12.15) is achieved by the product $270 \cdot 0.45 \cdot 0.1$.

In the sawing node (3) there are three product portions: balk, board and sawing waste. The node has the attributes cumulated from the previous nodes and an additional attribute electric energy. The cumulated value of former input and output attributes is the sum of the values of derived attributes in incoming edges and the ordinal value of the attribute. For example the cumulated value of CO₂ is $10 + 18 + 12.15 = 40.15$. The ordinal value of the attribute is based on the used electric energy (0.5 kg per one kWh).

$\langle 3, \text{Sawing},$

$\{ \langle \text{balk}, 1000 \text{ m}, \{id_{3000}, \dots, id_{3200}\}, \mathbf{0.50} \rangle,$
 $\langle \text{board}, 4000 \text{ m}, \{id_{4000}, \dots, id_{4800}\}, \mathbf{0.40} \rangle,$
 $\langle \text{SawingWaste}, 5 \text{ ton}, \emptyset, 0.1 \rangle \},$
 $\{ \langle \text{CO}_2, \text{output}, \mathbf{10 \text{ kg}}, \mathbf{40.15 \text{ kg}} \rangle,$
 $\langle \text{Diesel}, \text{input}, \mathbf{0}, \mathbf{10.05 \text{ liters}} \rangle,$
 $\langle \text{CompanyCode}, \text{info}, \{311\}, \{111, 211, 311\} \rangle,$
 $\langle \text{ElectricEnergy}, \text{input}, \mathbf{20 \text{ kWh}}, \mathbf{20 \text{ kWh}} \rangle \}$

Next the balks are transferred to a drying process. In the edge from Node 3 to Node 4, the objects maintain their identities and all the balks are transferred (shifted ratio value 1). The balks represent

50% of costs of the previous node, i.e. the values of input and output attributes are multiplied by 0.5.

⟨3, 4, ⟨balk, 1000 m, {⟨id, id⟩ | id ∈ {id3000,...,id3200}}, 1⟩,
 {⟨CO2, output, **20.075** kg⟩,
 ⟨Diesel, input, **5.025** liters⟩,
 ⟨CompanyCode, info, {111, 211, 311}⟩,
 ⟨ElectricEnergy, input, **10** kWh⟩}

The balks are dried in Node 4 which produces 1000 kg carbon dioxide and takes 2000 kWh of electric energy.

⟨4, Drying,
 {⟨balk, 1000 m, {id4000,...,id4800}, 1⟩},
 {⟨CO2, output, **1000** kg, **1020.075** kg⟩,
 ⟨Diesel, input, **0**, **5.025** liters⟩,
 ⟨CompanyCode, info, {311}, {111, 211, 311}⟩,
 ⟨ElectricEnergy, input, **2000** kWh, **2010** kWh⟩}

Drying of boards is similar to the drying of balks. The following edge and node represent drying of boards.

⟨3, 5, ⟨board, 4000 m, {⟨id, id⟩ | id ∈ {id4000,...,id4800}}, 1⟩,
 {⟨CO2, output, **16.06** kg⟩,
 ⟨Diesel, input, **4.02** liters⟩,
 ⟨CompanyCode, info, {111, 211, 311}⟩,
 ⟨ElectricEnergy, input, **8** kWh⟩}

⟨5, Drying,
 {⟨Board, 4000 m, {id4000,...,id4800}, 1⟩},
 {⟨CO2, output, **800** kg, **816.06** kg⟩,
 ⟨Diesel, input, **0**, **4.02** liters⟩,
 ⟨CompanyCode, info, {311}, {111, 211, 311}⟩,
 ⟨ElectricEnergy, input, **1600** kWh, **1608** kWh⟩}

Next dried boards are transferred to a gluing process. In this phase we assume that 10% of boards are disallowed, because for some reason they are flawed. This means that the ratio of the shifted product is 0.9. In gluing several boards (say ten) are composed to one glued beam. In the corresponding identity shift ten board objects are mapped to one glued beam object. For example in ⟨{id4000, ..., id4010}, id5000⟩ the identities id4000, ..., id4010 are board objects and id5000 is a glued beam object.

⟨5, 6, ⟨board, 2000 m, {{⟨id4000, ..., id4010⟩, id5000},
 {⟨id4010, ..., id4020⟩, id5001}, ... }, **0.9**⟩,
 {⟨CO2, output, **734.454** kg⟩,
 ⟨Diesel, input, **3.618** liters⟩,
 ⟨CompanyCode, info, {411}, {111, 211, 311, 411}⟩,
 ⟨ElectricEnergy, input, **1474.2** kWh⟩}

Finally, in the gluing process glued beams are composed. There is also an additional attribute ‘glue’ that describes the amount of the used glue.

⟨6, Gluing,

{⟨GluedBeam, 2000 m, {id5000,..., id5500}, **0.9**⟩,
 ⟨WoodWaste, 0.1 ton, ∅, 0.10⟩},
 {⟨CO2, output, **100 kg**, **834.454 kg**⟩,
 ⟨Diesel, input, 0, **3.618** liters⟩,
 ⟨CompanyCode, info, {311}, {111, 211, 311}⟩,
 ⟨ElectricEnergy, input, **200 kWh**, **1647.2 kWh**⟩,
 ⟨Glue, input, 100 kg, 100 kg⟩}

Now we are able to calculate cumulated resources and emission for single glulam beams. For example emissions of single products or a set of products can be calculated.

7. ANALYZING TRACEABILITY GRAPH

In this section we introduce different analyzing possibilities based on our data-centric approach. We assume a traceability graph ⟨N-Set, E-Set⟩ notation behind the formalization.

7.1 Basic Functions for Analyzing Object Structure

Objects, their mutual structures and properties are embedded in the traceability graph. Next, we introduce how they can be derived from a traceability graph.

The predecessors of an object are in special interest because they determine materials and components needed for the object. Immediate predecessors are the nearest predecessors of the object (id) and they can be achieved by the function $i_predecessors$ that is defined as follows:

$$i_predecessors(id) = \{id' \mid E \in E\text{-Set} \wedge id \in rng(E.SP.M) \wedge id' \in dom(E.SP.M): \langle id', id \rangle \in E.SP.M\}$$

where E is an edge, SP its sifted product portion (presented as a tuple) an M the mapping among objects.

For example, $i_predecessors(id5000) = \{id4000, \dots, id4010\}$, i.e. glulam beam id5000 is glued from boards id4000-id4010.

All predecessors can be achieved recursively as follows:

$$predecessors(id) = \begin{cases} S \cup \bigcup_{i \in S} predecessors(i), & \text{where } S = i_predecessors(id), \text{ if } S \neq \emptyset \\ \emptyset, & \text{otherwise} \end{cases}$$

For example, $predecessors(id5000) = \{id4000, \dots, id4010, id1, id2, id3\}$, i.e. glulam beam id5000 is glued from boards id4000-id4010 which are sawed from logs id1, id2 and id3.

Successors of an object mean, for example, those objects for that the object has been raw material or component. The functions $i_successors$ and $successors$ yield the immediate and all successors, respectively.

$$i_successors(id) = \{id' \mid E \in E\text{-Set} \wedge id \in rng(E.SP.M) \wedge id' \in dom(E.SP.M): \langle id, id' \rangle \in E.SP.M\}$$

$$successors(id) = \begin{cases} S \cup \bigcup_{i \in S} successors(i), & \text{where } S = i_successors(id), \text{ if } S \neq \emptyset \\ \emptyset, & \text{otherwise} \end{cases}$$

For example, $successor(id1) = \{id3000, id4000, \dots, id4003, id5000\}$, i.e. log with id1 is sawn to boards id4000-id4003 and beam id3000. The boards are used to manufacture the glulam beam with id5000.

An object may belong to several nodes in process chain. The function $node(id)$ yields all the nodes that the object with id has associated with:

$$node(id) = \{N.Nid \mid N \in N\text{-set}: t \in N.P\text{-set} \wedge id \in t.ID\text{-set}\}$$

Among these nodes the first and last ones are of special interest because the former determines the node where the object is created and the latter refers to the final state of the object. The functions $first_node(id)$ and $last_node(id)$ return them as follows.

$$first_node(id) = N \in node(id) \wedge id \notin predecessors(id)$$

$$last_node(id) = N \in node(id) \wedge id \notin successors(id)$$

Below we demonstrate the use of these functions.

7.2 Horizontal and Vertical Views

Traditionally views are predefined queries containing a derivation rule. In the context of the traceability graph the view means a sub graph determined by a rule for some purpose. In this paper we consider horizontal and vertical views. A horizontal view means an extensionally connected subgraph fired by object/objects. The connection at the extensional level means that objects of nodes in a view are connected through object mapping. Vertical view means that nodes of the same type are merged within a meta-node.

Among horizontal views we consider two basic cases: supply chain of an object and range distribution of an object. The supply chain of an object contains all preceding nodes that are extensionally connected (via object mapping) with the object. In terms of the function $predecessors$ the supply chain of the object id is defined by the function $SC(id)$ as follows:

$SC(id) = \langle N\text{-Set}', E\text{-Set}' \rangle$ such that

$$\begin{cases} N\text{-Set}' = \{N \in N\text{-set} \mid t \in N.P\text{-set} \wedge id \in t.ID\text{-set} \cup predecessors(id)\} \\ E\text{-Set}' = \{E \in E\text{-set} \mid N1, N2 \in N\text{-set}' \wedge E.N_S = N1 \wedge E.N_E = N2\} \end{cases}$$

For example considering the glulam beam with id5000. The horizontal view contains the nodes and edges the glulam beam has participated in. For the sake of brevity we refer to nodes by their identities and edges to the participating node identities.

$$SC(id5000) = \langle \{N1, N3, N5, N6\}, \{\langle N1, N3 \rangle, \langle N3, N5 \rangle, \langle N5, N6 \rangle\} \rangle$$

The distribution of an object means those nodes where the object or its part has been a participating in the TG. The distribution can be defined as follows:

$dist(id) = \langle N\text{-Set}', E\text{-Set}' \rangle$ such that

$$\begin{cases} N\text{-Set}' = \{N \in N\text{-set} \mid t \in N.P\text{-set} \wedge id \in t.ID\text{-set} \cup successors(id)\} \\ E\text{-Set}' = \{E \in E\text{-set} \mid N1, N2 \in N\text{-set}' \wedge E.N_S = N1 \wedge E.N_E = N2\} \end{cases}$$

For example considering the log with id1. The horizontal view shows all the nodes and edges the log has participated in.

$$dist(id1) = \langle \{N1, N3, N4, N5, N6\}, \{\langle N1, N3 \rangle, \langle N3, N4 \rangle, \langle N3, N5 \rangle, \langle N5, N6 \rangle\} \rangle$$

Horizontal views can also be applied to tracing a set of objects having a specific property.

A vertical view is conceptually different of the horizontal ones. Namely it is basically defined on the intensional level - for a set of nodes of the same type merged onto one pseudo node. However, the selection criteria for nodes in the view can be extensional as well. For example nodes of objects having a property may be a selection criterion for a vertical view. The following definition for the vertical view may at the first view look complex but basically the its parts follow the similar way to deriving. Next we give the definition of vertical view as a whole and then introduce in detail.

Let $N\text{-Set}' (\subseteq N\text{-Set})$ be a set of nodes of the same type, i.e. for each nodes $N_1, N_2 \in N\text{-Set}'$: $N_1.N\text{-type} = N_2.N\text{-type}$. The horizontal view node is represented as a tuple $\langle Nid_h, N\text{-type}, P\text{-set}_h, A\text{-set}_h \rangle$ where

- $Nid_h = \bigcup_{N \in N\text{-Set}'} \{N.Nid\}$
- $N\text{-type} = N.N\text{-type}: N \in N\text{-Set}'$
- $P\text{-set}_h$ is a set of tuples each of the form $tp = \langle P\text{-name}_h, C_h, ID\text{-set}_h, R_h \rangle$ such that

$$\left\{ \begin{array}{l} P\text{-name}_h \in \{t.P\text{-Name} \mid N \in N\text{-Set}': t \in N.P\text{-set}\} \\ C_h = \sum_{N \in N\text{-Set}'} t.C: t \in N.P\text{-set} \wedge t.P\text{-name} = P\text{-name}_h \\ ID\text{-set}_h = \bigcup_{N \in N\text{-Set}'} t.ID\text{-set}: t \in N.P\text{-set} \wedge t.P\text{-name} = P\text{-name}_h \\ R_h = \frac{\sum_{N \in N\text{-Set}'} t.R: t \in N.P\text{-set} \wedge t.P\text{-name} = P\text{-name}_h}{|N\text{-Set}'|} \end{array} \right.$$

- $A\text{-set}_h$ is a set of tuples each of the form $ta = \langle A\text{-Name}_h, T_h, V_h, W_h \rangle$ such that

$$\left\{ \begin{array}{l} A\text{-name}_h \in \{t.A\text{-Name} \mid N \in N\text{-set}': t \in N.A\text{-set}\} \\ T_h = t.T \mid N \in N\text{-Set}' \wedge t \in N.A\text{-set} \wedge t.A\text{-name} = A\text{-name}_h \\ V_h = \begin{cases} \bigcup_{N \in N\text{-Set}'} t.V: t \in N.A\text{-set} \wedge t.A\text{-name} = A\text{-name}_h, & \text{if } T_h = \text{info} \\ \sum_{N \in N\text{-Set}'} t.V: t \in N.A\text{-set} \wedge t.A\text{-name} = A\text{-name}_h, & \text{otherwise} \end{cases} \\ W_h = \begin{cases} \bigcup_{N \in N\text{-Set}'} t.W: t \in N.A\text{-set} \wedge t.A\text{-name} = A\text{-name}_h, & \text{if } T_h = \text{info} \\ \sum_{N \in N\text{-Set}'} t.W: t \in N.A\text{-set} \wedge t.A\text{-name} = A\text{-name}_h, & \text{otherwise} \end{cases} \end{array} \right.$$

In the formula the set of node identities is Nid_h , $N\text{-type}$ is the common type of the nodes, $P\text{-set}_h$ is the set of unionized product portions and $A\text{-set}_h$ is the set of merged attributes. In a product portion, $P\text{-name}_h$ is the name of a product, C_h is the total amount of the products and $ID\text{-set}$ is the set of all object identities in the unionized product portion. The ratio R_h is calculated by dividing the corresponding amount of products by the number of the nodes participating in the view. A merge attribute consists of its name $A\text{-Name}_h$, the type T_h of the attribute, merged ordinal value V_h , and cumulated value W_h . The values are calculated by summing or unionizing the original values depending on the type of the attribute.

A horizontal view can be used to compact a workflow diagram. For example we could merge the harvesting nodes of our running example and get the total amount of harvesting in our sample supply chain.

<{1,2}, harvesting,
 <{PineSawLog, 650 m³, {id1, ..., id1800}, 0.525},
 <{PinePulpWood, 500 m³, ∅, 0.375},
 <{HarvestingWaste, 40 ton, ∅, 0.10}>,
 <{CarbonDioxide, output, 570 kg, 570 kg},
 <{Diesel, input, 190 liters, 190 liters},
 <{CompanyCode, info, {111, 211}, {111, 211}}>,
 <{Location, info, {lat 62.87 - lon 22.86, lat 65.21 - lon 21.36}, {lat 62.87 - lon 22.86, lat 65.21 - lon 21.36}}>

In an enlarged example we could analyze the effect of different drying programs by merging the drying nodes based on the info-attribute that indicates the drying program.

In Figure 3 the enlarged example is represented. In the example the drying nodes #D12, #D13 and #D14 are merged to node #D12-14 and nodes #D15, #D16 and #D17 are merged to node #D15-17. The merging of traceability graph can be used to benchmark group of processes as described in the next section.

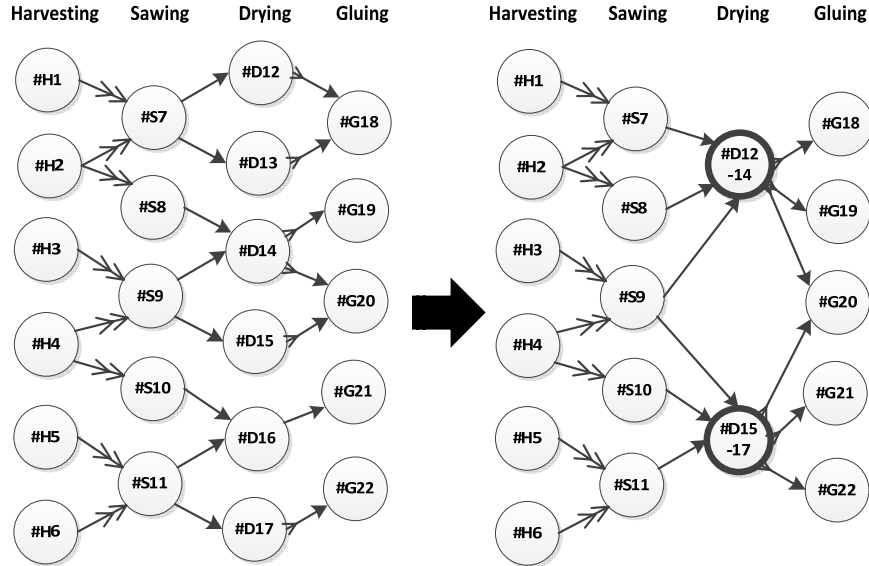


Figure 3. Merging Traceability Graph.

7.3 Examples

We demonstrate querying possibilities by sample queries that correspond to the uses cases described in Section 2.

Sample query 1: Calculating the item level carbon footprint

The carbon footprint of the object can be calculated by using the cumulated value of the CO2 attribute of the final node that the object has participated in. For example the carbon footprint of the glulam beam with id5000 is:

$$\frac{| \{id5000\} |}{| \{t.ID - set\} |} \cdot t.R \cdot t'.W$$

where $t \in N.P\text{-set}$: $id5000 \in t.ID\text{-set} \wedge t' \in N.A\text{-set}$: $t.A\text{-name} = CO2$ such that $N = last_node(id)$

Carbon footprint = $(1 / 500) \cdot 0.9 \cdot 834.454 \text{ kg} = \mathbf{1,5020172 \text{ kg}}$

The formula can easily be extended to concern several objects when the dispensation would be multi-valued set.

Sample query 2: Origin of raw material

As described in Section 3 the information about the origin of raw material is becoming more and more important. Using the traceability graph we can trace the origin of the product. The origin of the gulam beam with id5000 is achieved by the location attribute in the harvesting nodes as follows:

$$t.V \mid t \in N.A\text{-set} \wedge t.A\text{-name} = \text{location}: N \in N\text{-set}' \wedge N \notin \text{rng}(E\text{-set}') \text{ where } \langle N\text{-set}', E\text{-set}' \rangle = \text{SC}(\text{id5000})$$

In other words, CS determines the supply chain of the gulam beam and the V value is returned from the location attribute. The condition $N \notin \text{rng}(E\text{-set}')$ ensures that the node is an initial process. The result is $\{\text{lat } 62.87 - \text{lon } 22.86\}$.

It is worth noting that above definition is based on a non-cumulated attribute. In cumulated attributes a corresponding query is simpler. CompanyCode is such an attribute. In our example customer is a corresponding cumulated attribute. For example the custody of the gulam beam with id5000 can easily achieved as follows:

$$t.W \mid t \in N.A\text{-set} \wedge t.A\text{-name} = \text{CompanyCode}: N = \text{last_node}(\text{id5000})$$

The result is $\{111, 211, 311\}$

Sample query 3: Recalling products

The traceability information can be used to recall products accurately and rapidly without needing to do the total recall throughout the supply chain. With information of the traceability graph we can find out all the products that some object is used as a component or raw material. The function *final_p* gives the final products in which an object (id) is used.

$$\text{final_p}(\text{id}) = \{\text{id}' \in \text{successors}(\text{id}) \mid \text{last_node}(\text{id}') \notin \text{rng}(E\text{-Set})\}$$

In our example, $\text{final_prod}(\text{id1}) = \{\text{id3000}, \text{id5000}\}$, i.e. beam id3000 and gulam beam id5000 are the recalled final products.

Sample query 4: Benchmarking

Benchmarking the processes between companies and manufacturing facilities enables to identify the processes with biggest environmental impact so that we improve the environmental performance of the supply chain. By using the vertical view we can calculate a key performance indicator for the nodes using the *bench()* functions.

- nodes define the set of nodes used in benchmarking
- prod_name defines the product portion used in benchmarking
- prop_name defines the attribute used to calculate the key performance indicator.
- group defines the attribute used to analyze the traceability graph.

$\text{bench}(\text{nodes}, \text{prod_name}, \text{prop_name}, \text{group}) =$

$$\{\langle x, y \rangle \mid x \in t_1.V: t_1 \in N.A\text{-set} \wedge t_1.A\text{-name} = \text{group} \wedge t_1.T = \text{info} \wedge y = \frac{\sum_{i \in S} i}{\sum_{j \in T} j} \text{ where}$$

$$\begin{cases} S = \{t_2.V \mid t_2 \in N.A\text{-set} \wedge t_2.A\text{-name} = \text{prop-name}\} \\ T = \{t_3.V \mid t_3 \in N.P\text{-set} \wedge t_3.P\text{-name} = \text{prod-name}\} \end{cases}$$

where $N \in \text{nodes}$

In our running example we can calculate the harvesting efficiency as follows:

bench({1,2}, PineSawLog, Diesel, CompanyCode)

The result $\{\langle 111, 0.286 \rangle, \langle 211, 0.3 \rangle\}$ presents how much diesel companies have used per cubic meter of saw logs. The comparison value for the company 111 is 0.286 and for the company 211 is 0.3.

8. DISCUSSION

The presented data-centric workflow model enables tracing, monitoring, analyzing and querying the properties of processes and their mutual relationships. The formal specification allows services to handle the products lifecycle data formally. To be able to share the life cycle data in real a world supply chain, we must:

1. ensure correspondence between logical objects with real life products of processes
2. have an infrastructure that enables multiple companies in a supply chain to share and use the information regarding products.

In tracing products, in addition to logical identities, they must be identified by physical identifiers. For physical products, various marking methods are in use: Imprinting, the finger print method, Laser marking, Label marking, Ink jet marking and transponder marking. In practice a physical identifier corresponds to object identity in database. This also gives natural interpretation for an object in the traceability graph. Below we consider an RFID (Radio Frequency IDentification) marking case related to our running example.

The modularity of a supply chain means that each actor is responsible for generating data from a part of the supply chain of the product and to share it with other stakeholders. To be able to share product related information in the complex supply chains the organizations have to agree on a common standard. One of the most promising is EPCglobal Architecture Framework² standards which are generally accepted methods for sharing product data in supply chains. They enable supply chain stakeholders to capture, store and share product related data. The EPCGlobal architecture includes EPC Information Services specification [17] that defines storing and sharing the traceability data that is created when a product marked with an RFID-tag passes an RFID-reader in a process in a supply chain. This event data normally contains unique identification code, location and time. By extending the EPC Information Service specification also environmental data can be included in event data. For example: ‘At location X in time Y the object Z was observed with the environmental data [elementary flow #1, elementary flow #2]’

To be able to generate a total carbon footprint for a product the organizations must share environmental information of the products that were handled by them in their part of the supply chain. For example, in our running example, some organizations are responsible for harvesting the timber; sawmill companies handle the sawing and glued laminated timber manufactures are using the boards sawn in sawmills. All these stakeholders own a part of the final product’s life cycle information. To be able to share environmental information each stakeholder must implement an

² <http://www.epcglobalinc.org/standards/>

EPC Information Service that implements the extension for environmental data. To be able to handle the object transformation (division or composition) in the supply chain, the stakeholder responsible for the transformation part of the supply chain is also responsible for aggregating the environmental information from up to that point. In other words, when a manufacturer is further processing products, the manufacturer is responsible for calling the EPC Information Services of a supplier and to add this (derived attribute in the traceability graph) information to the environmental information of the further processed product.

9. CONCLUSIONS

We have presented a data-centric workflow model, called the traceability graph. It integrates data-centric aspects of products and processes to traditional graph-based workflows. The approach supports attribute value propagation and aggregation in the supply chains. Input and output costs of processes can be allocated into products, which enable tracing and analyzing these costs precisely. The model can be applied to single products as well as larger patches. Unlike existing methods the traceability graph enables precisely calculated input costs (e.g. resources) and output costs (e.g. emissions and waste) of products and processes. So far these have been based on average values from a large set of processes.

Through the presented object transformation it is possible to model and manipulate the composition and division of objects in processes. We defined horizontal and vertical views. A horizontal view can represent a supply chain or the distribution of resources or components. In terms of the vertical view a traceability graph can be compacted by collecting similar processes together. We demonstrated analyzing possibilities of the traceability graph by several sample functions and use cases.

10. REFERENCES

- [1] van der Aalst, W. M. P. The application of Petri nets to workflow management. *The J. of Circuits, Syst. and Computers*, 8(1), 1998, 21-66.
- [2] van der Aalst, W. M. P. and Ter Hofstede, A. H. M. YAWL: Yet another workflow language. *Inf. Syst.*, 30(4), 2005, 245-275.
- [3] van der Aalst, W. and van Hee, K. *Workflow Management, Models, Methods, and Systems*, The MIT Press, Cambridge, MA, 2002.
- [4] Akram, A., Kewley, J. and Allan, R.: A Data Centric approach for Workflows. EDOC Workshops 2006, 10
- [5] Attie, P., Singh, M., Sheth, A., and Rusinkiewicz, M. Specifying and enforcing intertask dependencies. In *Proceedings of 19th International Conference on Very Large Data Bases (VLDB'93)* (Dublin, Ireland, August 24-27, 1993), 134-145.
- [6] Bonner, A. J. Workflow, transactions and datalog. In *Proceedings of the Eighteenth ACM Symposium on Principles of Database System (PODS'99)* (Philadelphia, Pennsylvania, May 31-June 2, 1999), 294-305.
- [7] Booch, G., Rumbaugh, J and Jacobson I. *The Unified Modeling Language User Guide*, Addison-Wesley, Reading, MA, 1999.
- [8] Caswell, N. S. and Nigam, A. Business artifacts: An approach to operational specification. *IBM Systems Journal*, 42(3), 2003, 428-445.

- [9] Chen, P. The entity-relationship model—toward a unified view of data , *ACM Transaction on Database Systems*, 1(1), 1976, 9-36.
- [10] Coad, P. and Yourdon, E. *Object-Oriented Analysis*. Prentice-Hall: 1990.
- [11] Curcin, V. and Ghanem, M. Scientific workflow systems - can one size fit all? In *Proceedings Biomedical Engineering Conference (CIBEC 2008)*. (Cairo, 18-20 Dec. 2008), 1-9.
- [12] Deutch, A., Hull, R., Patrizi, F. and Vianu, V. Automatic verification of data-centric business processes. In *Proceedings of the 12th International Conference on Database Theory*, (St. Petersburg, March 23-26, 2009), 252-267.
- [13] Ellis, C. A. and Nutt, G. J. Office Information Systems and Computer Science, *ACM Comp. Surv.*, 12(1), 1980, 27-60.
- [14] EPCGlobal, EPC Information Services (EPCIS) Version 1.0.1 Specification, EPCglobal Ratified Standard, 2007.
- [15] ISO 14020, Environmental labels and declarations -- General principles, ISO 14020:2000, International Organization for Standardization, Geneva, Switzerland, 2000.
- [16] ISO 14040, *Environmental Management - Life Cycle Assessment - Principles and Framework*, ISO 14040:1997(e), International Organization for Standardization, Geneva, Switzerland, 1997.
- [17] ISO 14064, Greenhouse gases -- Part 1: Specification with guidance at the organization level for quantification and reporting of greenhouse gas emissions and removals, ISO 14064-1:2006, International Organization for Standardization, Geneva, Switzerland, 2006.
- [18] Junkkari, M. PSE: An Object-Oriented Representation for Modeling and Managing Part-of Relationships, *J. of Intelligent Inf. Syst.*, 25(2), 2005, 131-157.
- [19] Junkkari, M. and Sirkka A. Using RFID for tracing cumulated resources and emissions in supply chain, *Int. J. Ad Hoc and Ubiquitous Computing*, 8(4), 2011, 220-229.
- [20] Kifer, M., Lausen, G. and Wu, J. Logical foundations of object-oriented and frame-based languages, *J. of ACM*, 42(4), 1995, 741-843.
- [21] Niemi, T., Junkkari, M., Järvelin, K. and Viita, S. Advanced query language for manipulating complex entities, *Inf. Processing and Management*, 40(6), 2004, 869-889.
- [22] Puettmann, M. and Wilson, J. 2005. Gate-to-gate Life-Cycle Inventory of Glued Laminated Timbers Production. *Wood Fiber Sci.*, 37, 2005, 99-113.
- [23] Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F., and Lorenzen, W. *Object-Oriented Modeling and Design*. Englewood Cliffs, NJ: Prentice Hall. 1991.
- [24] Rumbaugh, J., Jacobson, I., & Booch, G. *The Unified Modeling Language Reference Manual*. Reading, MA: Addison-Wesley. 1999.
- [25] Savnik, I., Tari, Z., And Mohoric, T. QAL: A query algebra of complex objects. *Data & Knowledge Engineering*, 30(1), 1999, 57-94.
- [26] Sirkka, A. and Junkkari, M. Data management framework for monitoring and analyzing the environmental performance, In *Proceedings of INNOV 2010, ICGREEN 2010*, (Athens, July 29-31, 2010), 57-62.

Using RFID for tracing cumulated resources and emissions in supply chain

Marko Junkkari*

School of Information Sciences,
University of Tampere, FIN 33014, Finland
E-mail: Marko.Junkkari@cs.uta.fi
*Corresponding author

Antti Sirkka

Tieto Finland Oy,
Hatanpäänvaltatie 30,
Tampere 33101, Finland
E-mail: Antti.Sirkka@tieto.com

Abstract: The tracing of resources and emissions has been recognised increasingly important in supply chains. The developed traceability graph enables tracing of information associated with products and their parts. Tracing in a supply chain requires the three integrated levels of actions:

- 1 at the physical level the Radio-Frequency Identification (RFID) is used for physical objects
- 2 at the data-storing level logical counterparts of RFID tags and database objects are mapped
- 3 operational actions determine how resources, emissions etc. are portioned in different parts of the traceability graph.

Briefly, we integrate RFID technology with databases and operational actions of the supply chain.

Keywords: RFID; radio-frequency identification; supply chain; tracing; traceability graph; cumulated resources; cumulated emissions; life-cycle assessment; environmental management; modelling; accumulation of attributes; database.

Reference to this paper should be made as follows: Junkkari, M. and Sirkka, A. (2011) 'Using RFID for tracing cumulated resources and emissions in supply chain', *Int. J. Ad Hoc and Ubiquitous Computing*, Vol. 8, No. 4, pp.220–229.

Biographical notes: Marko Junkkari got Master's Degree in Computer Science from the University of Tampere in 1997. He achieved Licentiate Degree in 2002 and PhD in 2007 in the same university. His research activities cover conceptual structures, graph-based methods, development and integration of data models, query languages, and XML information retrieval. He has been a reviewer in several international journals and a member of ACM SIGIR programme committee. Since 2006, he has a position as an Associate Professor of Data Management in the University of Tampere. His teaching activities include database programming, XML, special topics on data management and fuzzy logic.

Antti Sirkka got Master's Degree in Computer Science from the University of Tampere in 2001. He has then worked in Tieto building data management systems for the forest industry, and is currently working on the European Commission's Sixth Framework Programme-funded project, Indisputable Key as a responsible for IT Architecture. The project researches wood traceability via RFID. He is also working towards completing his thesis on supply chain performance management using traceability.

1 Introduction

Physical assemblies and other physical products have their own supply histories represented by supply chains. The supply chain can be understood as a network of autonomous or semiautonomous business entities collectively responsible for procurement, manufacturing and distribution actions of a product. A supply chain is a

directed subgraph of a graph used for modelling manufacturing processes, e.g., workflow diagrams (van der Aalst, 1998; van der Aalst and Ter Hofstede, 2005; van der Aalst and van Hee, 2002; Attie et al., 1993; Bonner, 1999; Booch et al., 1999).

Structurally, a supply chain consists of supply processes following each other in a partial order, i.e., the result of a

process is a raw material for another process. Each process involves resources and emissions associated with the products that are manufactured in the process. This means that the resources and emissions are cumulated in a supply chain. A single product may be an output of a supply chain that requires a great volume of energy and produces environmental toxins and greenhouse gases. In recent times, the importance of environmental aspects has been widely recognised, which means that the cumulated resources and emissions must be partitioned on different products and batches more and more precisely. This paper deals with modelling and manipulation of supply chains such that cumulated resources and emission involving environmental aspects (e.g., CO₂ emissions) can be analysed for products.

There are many programmes launched to create certificates for environmentally friendly products such as Carbon footprint (Carbon Trust, 2006, 2007), European Eco-label (EC, 2000) and EPD (ISO 14025, 2006). Environmental labels are granted based on a method for calculating the environmental effects of the supply chain. EU Eco-Management and Audit Scheme (EMAS) (EC, 2001) and Life-Cycle Assessment (LCA) (ISO 14040, 1997; ISO 14041, 1998; ISO 14044, 2006) are proposals for this. EMAS is a management tool for companies and other organisations to evaluate, report and improve their environmental performance. In this paper, we focus on LCA, which is a standardised method for calculating the environmental burden caused by the manufacturing of a product. The goal of the LCA is to compare the environmental impact of a product or service throughout its lifespan to be able to choose the least burdensome one. The existing approach for calculating the environmental impact in a supply chain is to analyse the yearly average information about emissions and production.

In this paper, we present the model for tracing and storing the inventory data, the impact assessment and interpretation phase is excluded. Unlike the existing methods, our model enables analysing resources and emissions on the single product level even in a real time – not only average yearly values. Nowadays, a common method for calculating the environmental impact is to measure the input and output flows of the whole supply chain during some time period and calculate the average environmental impact for the product (PE International, 2002; Puettmann and Wilson, 2005). Further, the dynamic nature of the supply chain is not taken into account. For example, in the *European Reference Life Cycle Data System (ELCD) core data set 1.0.1*. (PE International, 2002), an average value used for transportation distance for timber is 114 km. A product manufactured using raw material transported from another continent is bound to have different environmental effect than another product manufactured using raw material from a nearby source. By using the autoidentification, we can create a methodology that enables real-time monitoring of the environmental impact for the single products.

A supply process is focused on product portions consisting of single products (called objects) or a mass of

material. In other words, a product portion is a part of a patch selected, e.g., for refining. Product portions can be unionised or a product portion can be divided into smaller parts. Manipulation of objects requires specific features because objects may be changed into other objects in both the logical (in databases) and physical (real world) levels. Namely, a single object may be divided into other objects or several objects can be composed to a single object. This means that object transformation must be modelled.

In a process, objects may be divided or composed. This means new objects must be identified at the logical level (in databases) and physical level (in real world). In the physical level, there must be an explicit identifier that corresponds to the logical identities (or vice versa). The development of a smart identification enables us to identify an object moving in the supply chain. This means that we can connect the physical world objects with their virtual counterparts in databases. With the traceability, we can track the relationships among properties of processes, in this case the environmental burden caused by processes and actual product instances.

To enable strict and real-time traceability in a supply chain, the objects of interest have to be automatically identified. The development of the smart identification enables us to identify single objects moving in the supply chain and this enables us to track their movements and monitor their changes. There are numerous marking methods that can be used in a supply chain: Imprinting, Laser marking, Label marking, Ink jet marking and transponder marking. However, using of these conventional methods is often problematic because the methods require clear line of sight that is impossible to achieve in several manufacturing areas. For example, in forest industry, a log may be covered with mud or snow, which prevents of using these marking methods. Further, automatic reading of conventional marking is slow because a robotic system typically must search the identifier in different sizes of an object.

In this study, we consider the RFID technology for physical marking of the objects. The RFID technology can be compared with the bar code identification: an identification code is embedded to an object. The RFID technology and data management has been researched extensively during the last decade, e.g., in the software architecture and online data management (Chaudhuri and Dayal, 1997), and the RFID data warehousing (Gonzalez et al., 2006a, 2006b). RFID data staging to OLTP- and OLAP-applications (Krompass et al., 2007) provides a method that answers to different kinds of needs that transactional and analytical applications pose to RFID data. Dada and Staake (2008) describe a method of presenting the carbon footprint of the product instance to the customers. In this paper, we demonstrate how the traceability data – in this example RFID data – can be used to allocate the environmental burden to individual products.

Our aim is to develop a *traceability graph* for modelling supply chains of physical products. In this context, an *object* means a product that is identified on the logical and physical

levels. In general, the following features are attached to the traceability graph.

- 1 *Ability to manipulate objects, object sets and their transformation:* Object sets can be divided into subset or object sets can be unionised into larger sets. Single objects can be divided into smaller objects or single objects can be composed from other objects. Strict correspondence of the real-world actions and databases must be maintained.
- 2 *Ability to manipulate the properties of processes and to allocate them for underlying product portions:* In a process, different product portions are formed and transferred from a process to another. These portions can further be divided or composed into new proportions. For strict traceability, resources and emissions must be focused on different product portions, which are manipulated at three levels of actions. Among other things, this means that attributes representing tracing of recourses and emissions must be cumulated in the supply chain. The related calculation rules are based on rations of product portions in processes and their transforming among processes.
- 3 *Application independency:* The model is not bound to any specific application area. Our sample application is from the forest industry, but the model can be applied to any other domain of physical products.
- 4 *Support for advanced data analysis:* The supply chain with cumulated resources and emissions of a product portion or single objects can be derived from a traceability graph.

The rest of the paper is organised as follows. Section 2 deals with graphical notations of the traceability graph and an informal introduction for our sample application domain. Section 3 investigates physical, operational and data-storing actions related to the different parts of the traceability graph. An implementation for the traceability graph to relational databases is presented in Section 4. Sample analyses are also given in this section. In Section 5, we discuss problems of the present approach and their solutions. Finally, the conclusions are given in Section 6.

2 Traceability graph

A manufacturing process is an event that transforms the input elements into products. The inputs of the process are raw material and natural resources like energy. The outputs of the process are emissions, waste and generated products. Products that are identified in the underlying application domain are called *objects*.

A process involves *product portions*, which are output products of the process. A product portion contains a set of objects or a mass of non-identified products. Furthermore, a product portion involves the amount of products and a ratio for describing the piece of resources of the product portion in the process. *Attributes* of a process describe resources, emissions or other relevant properties associated with

the process. An *input attribute* describes a resource of a process, an *output attribute* describes other outputs than products, and an *info attribute* contains other information associated with the product.

The end products of a process may be raw material for another process. In other words, a product portion is shifted to another process. With product portions, the properties associated with previous process are transferred to the new process because they describe earlier resources related to the product portions. The relative part of attributes for a product portion can be calculated based on the ratio of the product portion.

Next, we introduce graphical notations of the traceability graph used to describe processes following each other in a partial order and give a sample system. Product portions and attributes are not represented graphically.

2.1 Introduction to traceability graph and graphical notations

In a traceability graph, a (process) node represents a process where manufactured products are categorised into product portions consisting of objects or other mass of products. Resources are divided and calculated for them based on their amount of used resources.

An edge is used for describing the elementary or product flows between two nodes, called start and end nodes. If the underlying flow contains a product portion consisting of objects, integrity constraints are associated with the edge. For this purpose, the traceability graph contains three kinds of edges:

- 1 *Plain edge* describes that an object does not transform in the node. Instead, the properties of objects and grouping of objects may be changed. If the underlying product portion does not consist of objects, then the plain edge is used.
- 2 *Division edge:* Single objects are divided into several objects. In other words, there is a mapping from a single object to a set of objects.
- 3 *Composition edge:* Objects are composed from several objects. In other words, there is a mapping from a set of objects to a single object.

The visual symbols for the node and edges are given in Figure 1.

2.2 Sample system

Next, we introduce our sample system used for illustrating forthcoming implementation. The sample traceability graph illustrates manufacturing of wood barks and glulam (glued laminated timber) beams. Graphically, the sample traceability graph is given in Figure 2. There are two harvesting nodes (N1 and N2). A harvesting node means the daily production of a single harvester.¹ The production of a harvester is transported to a manufacturer, for example saw logs to the saw mill and pulp wood to the pulp mill. In the example, only supply chains for wood barks and glulam

balks are represented. Further, when compared with a real-life system, several processes like forwarding, transporting, sorting and planning are left out.

Figure 1 Visual symbols for traceability graph



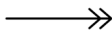
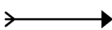
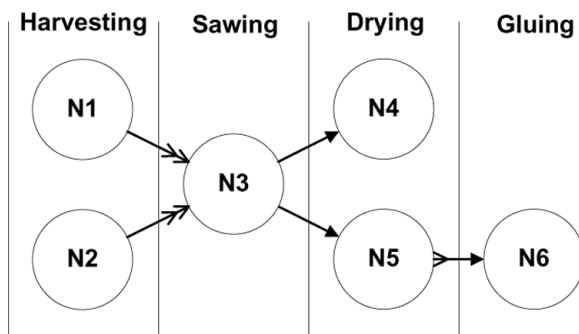
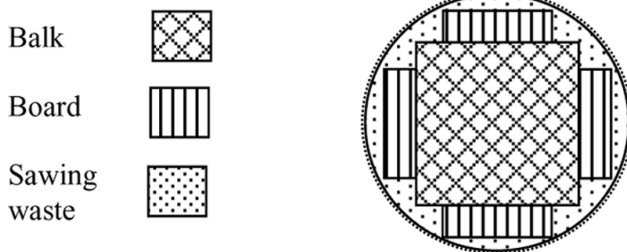
NODE	
Used to describe a process where resources are focused to specific product portions.	
EDGES	
Plain edge. Objects maintain their identities.	
Division edge. Double head arrow illustrates that in the flow, for each object related to the start node, there are several objects belonging to the end node.	
Composition edge. In the flow, for each object related to the end node, there are several objects belonging to the start node.	

Figure 2 An example on traceability in the forest industry



In the harvesting nodes, products are sorted such that sawing logs form one product portion in each node. In the example, parts of these portions are transferred to the sawing process (N3). There are various ways to saw a log to boards or balks. In this example, a sawing pattern (Figure 3) used in sawing produces one balk and four boards from each log. In other words, a log is a parent object for one balk object and four board objects.

Figure 3 A sawing pattern for logs



Next step in the sample supply chain is kiln drying where the boards and balks are dried to a predefined moisture percent. Since the dimensions of balk and boards differ, they are dried using different drying programmes (N4 and N5). Dried balks are end products, whereas the dried boards are used to manufacture glulam beams. Boards are glued (N6) together to create one beam.

3 Three-level framework for actions in the traceability graph

Graphical notations for the traceability graph are a rough-level description of supply chains in manufacturing and related areas. Next, we introduce actions of information system needed in strict tracing of resources and emissions. The underlying levels are physical (RFID), operational and data storing. All these levels interact with each other in the traceability graph.

The RFID system can be divided into three components:

- *RFID tag:* The main parts of the tag are a chip, an antenna and a packaging. The chip holds the individual code attached to the physical object. The antenna transmits the information to the reader and the packaging holds the antenna and the chip.
- *RFID reader:* RFID readers read the object identifier from the RFID tag without needing a clear line of sight. The RFID-reader can be integrated to a machine or it can be just a hand-held device.
- *Host:* The host computer hosts the application that maps the RFID data stream from RFID readers to more understandable form that is useful for application-level interactions.

In our example, logs, balks, boards and glulam beams are identified by RFID tags.

In databases, information on processes and their interactions, such as the logical counterparts of RFID tags, grouping of objects and mapping of objects among processes, are stored. Further, the database contains information on tracked resources and emissions presented by attributes.

Operational level consists of rules how resources and emissions are calculated in different parts of the traceability graph. For example, when sawing a log, the central balk, boards and sawing waste get their own parts of resources and emissions based on application-specific ratios.

In general, the following actions are associated with process nodes:

- *RFID actions:* Tags are attached to physical products or product portions.
- *Operational actions:* Products are divided into portions and the resources and emissions are partitioned to them. Possible earlier resources and emissions based on incoming edges are calculated and divided into product portions. The value of a cumulated attribute for a product portion is the sum of its proportions in the underlying and earlier processes.
- *Data-storing actions:* Identities of products and their groupings (portions) are stored following the granularity of RFID tagging. Relationships among identified products and portions are stored. The values of the resource and emission attributes of the process are stored. Cumulated values for product portions are stored.

Edges represent material and information flows from a node to another. An edge itself is not a process although it involves operational notations such as compositions or division of objects. Instead, an edge is an action model or pattern how objects are sifted to a process to another and what are the related actions in different levels. In Table 1, we present these actions.

Next, we give a database implementation for the traceability graph that contains schema for needed information of operational and data-storing levels.

4 Database implementation

The present database allows storing and analysing the information of the traceability graph. For illustrating the database implementation, we use the running example on the forest industry. In Figures 4, 5, 7 and 9, PK means a primary key whereas FK denotes a foreign key. For distinguishing attributes in relational database and traceability graph, we call attributes in the database as columns. The implementation of the graph is divided into three steps: First, the properties of the graph are implemented. Second, the identity shift is implemented. In the third phase, objects are associated with those

nodes they have travelled through. In the second and third phases, interactions of the database and RFID data are demonstrated.

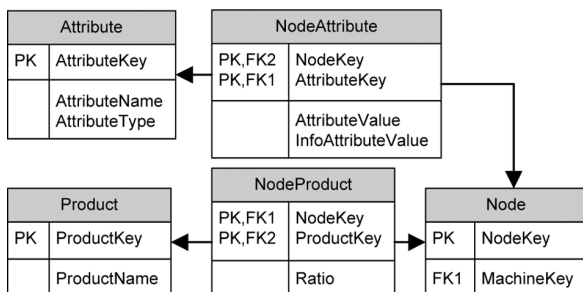
4.1 Product portions and attributes

The nodes of the application are stored into the *Node* relation. For the data associated with the properties of the traceability graph, we define four relations. The attributes and products are stored into the relations *Attribute* and *Product*, respectively. The data about product portions of the specific node is stored in the relation *NodeProduct*. The column *Ratio* is used in allocation of environmental burden between product portions. For example, the relation would include a row (Harvesting, PineSawLog, 0.6), which means that 60% of recourses and emissions are allocated to pine saw logs in the underlying harvesting node. For the node-specific information about attributes, the relation *NodeAttribute* is used. The column *AttributeValue* is used for numeric (input and output) attributes and the column *InfoAttributeValue* is used for textual (info) attributes. The cumulated value of an attribute is not stored as a column because it can be calculated based on the current values as presented later. These relations are illustrated in Figure 4.

Table 1 Edge patterns

	<i>RFID action pattern</i>	<i>Operational pattern</i>	<i>Data-storing pattern</i>
Plain edge	Tags are read in the start and end nodes	A part of a product portion in the start node is selected and its proportion of resources and emissions are calculated	Object identities of sifted objects and the corresponding proportion value are stored. Cumulated values may be stored (they also can be derived later)
Division edge	Old tags (in the start node) are read and new tags (in the end node) are attached to new product in the end node	<ol style="list-style-type: none"> 1 Same operations as in the plain edge 2 Each identity in the sifted portion (from the start node) is mapped to a set of identities (produced parts) 	<ol style="list-style-type: none"> 1 Same operations as in the plain edge 2 Identity mapping between original objects and their parts is stored
Composition edge	Old tags of components are read and new tags are attached to the composed products	<ol style="list-style-type: none"> 1 Same operations as in the plain edge 2 The set of the identities of sifted components (from the start node) is mapped to a single identity of composed product 	<ol style="list-style-type: none"> 1 Same operations as in the plain edge 2 Identity mapping between original components and composed product is stored

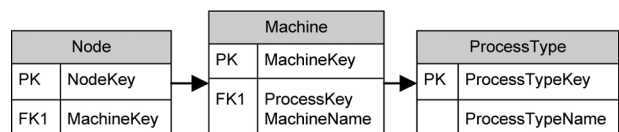
Figure 4 Product-set and attribute-set of a node



Next, we demonstrate how the traceability graph can be extended by application-specific primitives. For example, the machine used in a process is added to the database implementation (see Figure 5), and thus we can analyse the environmental burden between different machines and

processes. For this information, the relation *Machine* is implemented. The relation *ProcessType* stores the different process types of the application domain and the relation *Machine* is used to store the machines of the processes. For example, there are many harvesters doing the harvesting.

Figure 5 The relations *Machine* and *ProcessType*



By using the relations presented earlier, the implementation allows storing the information about environmental burden caused in the supply chain. The example query here illustrates how the above-mentioned relations can be used to

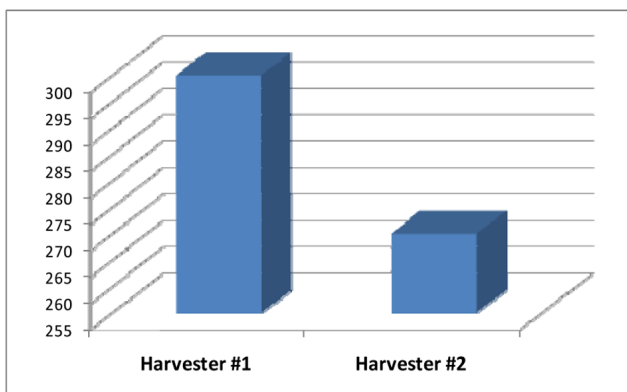
find out what is the amount of carbon dioxide emissions that different machines (harvesters) have caused in manufacturing. Figure 6 illustrates the result of the query, i.e., CO₂ emissions are grouped by machines. Information can also be analysed based on the product properties or processes.

```

SELECT
  Machine.MachineName,
  SUM(AttributeValue) AS [CO2]
FROM
  NodeAttribute
  INNER JOIN Node
    ON NodeAttribute.NodeKey = Node.NodeKey
  INNER JOIN Machine
    ON Node.MachineKey = Machine.MachineKey
WHERE
  NodeAttribute.AttributeKey =
    (SELECT AttributeKey
     FROM Attribute
     WHERE AttributeName = 'CO2')
  AND Machine.ProcessKey IN
    (SELECT ProcessKey
     FROM Process
     WHERE ProcessName = 'Harvesting')
GROUP BY Machine.MachineName

```

Figure 6 Daily CO₂ emissions grouped by machines (see online version for colours)



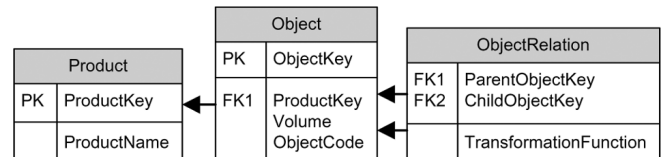
4.2 Identity shift and RFID implementation

The information about individual products, i.e., objects, are extracted from the RFID system that generates observations about object movements. The data generated from the RFID system is called an observation that contains unique object identifier, location and time.² An observation is generated when an object is detected by an RFID reader. Object identifiers are stored with the other information related to objects. In our example, the location is a symbolic location of the process and the time is a time instant when an object was detected in some process.

The relation *Object* is implemented for storing the object-specific information by the columns *ObjectKey*, *ProductKey* (Foreign Key to the Product relation), *Volume* and *ObjectCode* (physical identifier). A new row is created into the relation when an object is observed first time in the supply chain by the RFID system. The relation

ObjectRelation is used to trace the evolution path of the object (see Figure 7). Using this relation, we can store the identity shift between objects. Explicitly, division and composition are stored. For example, in the supply chain, log → board → glulam beam identity shifts among logs, boards and glulam beams are stored. The relation contains the column *TransformationFunction* for calculating cumulated values of attributes.

Figure 7 Object evolution relations



For example, a log (id1) is sawn to four boards id2000, id2001, id2002 and id2003 and one balk id3000, i.e., the relation *ObjectRelation* has five rows as shown in Table 2. The column *TransformationFunction* defines a coefficient used to calculate the value of the cumulated attribute.

Table 2 ObjectRelation example

<i>ObjectRelation</i>		
<i>ParentObjectKey</i>	<i>ChildObjectKey</i>	<i>TransformationFunction</i>
Id1	Id2000	0.1
Id1	Id2001	0.1
Id1	Id2002	0.1
Id1	Id2003	0.1
Id1	Id3000	0.5

The value of the column *TransformationFunction* is calculated by multiplying the ratio of the corresponding product portion and the ratio of shifted product portion. In this example, the ratio of the product portion of boards is 0.4 and one board is 0.25 of the product portion ($0.4 \cdot 0.25 = 0.1$). The query here illustrates the usage of this relation, in this example we query about the evolution path of the glulam beam id5000:

```

SELECT
  ParentObjectKey,
  TransformationFunction
FROM
  ObjectRelations
START WITH
  ChildObjectKey = (SELECT ObjectKey
                   FROM Object
                   WHERE ObjectCode = 'id5000')
CONNECT BY
  PRIOR ParentObjectKey = ChildObjectKey

```

The result of the query is given in Table 3, where an additional column describes the calculation of cumulated values.

From the result set of Table 3, we can calculate the cumulated values of the attributes for the glulam beam. The cumulated value is the sum ($W_1 + W_2 + \dots + W_8$) of

rows in the additional column in Table 3. In this example, we can see that the glulam beam was manufactured from four boards (id2000, id2001, id2002, id2003), and the boards were fully used to manufacture the glulam beam (*TransformationFunction* = 1). Going back in the evolution path, we can see that the board id2000 was sawn from log id1. For example, the cumulated value of the glulam beam contains 10% from the attribute values of log id2000 as described earlier.

Table 3 Evolution path for a glulam beam (id5000)

ParentObject	TransformationFunction	
Id2000	1	$W_1 = 1 \cdot V$
Id1	0.1	$W_2 = 1 \cdot 0.1 \cdot V$
Id2001	1	$W_3 = 1 \cdot V$
Id2	0.1	$W_4 = 1 \cdot 0.1 \cdot V$
Id2002	1	$W_5 = 1 \cdot V$
Id3	0.1	$W_6 = 1 \cdot 0.1 \cdot V$
Id2003	1	$W_7 = 1 \cdot V$
Id4	0.1	$W_8 = 1 \cdot 0.1 \cdot V$

4.3 Routes of objects

In the last phase of the implementation, objects are attached to the corresponding nodes. This phase also contains the implementation of edges. The relation *Route* is implemented for this purpose, i.e., by this objects can be traced through a supply chain. The path of nodes for each object is stored in the relation-based observations of RFID readers. When an RFID reader detects an object in some node, an observation is created and one row is added to the relation *Route*. The time instant of the observation is also stored into relation *Route*, in the column *Date*.

From the data management viewpoint, the relation *Node* is used to compress the traceability data by combining the group of objects that share the same amount of environmental burden. For example, during drying a product portion may consist of 10,000 boards. The environmental burden caused by drying is divided evenly (weighted by the volume of the object) for each board. In compressing traceability data, we follow the idea of Gonzalez et al. (2006a), which means that the environmental impact data is saved only once for the node instead of storing it for each object that participated in the node. The idea is to combine the paths that objects travelled together after the observation that the objects generally travel along the same path.

On the basis of the relations in Figure 8, we can find out nodes the object passed through. The next query illustrates the usage of these relations. The query lists the nodes that a glulam beam (id5000) has passed through. The query returns Nodes 1, 3 and 4.

```
SELECT
    NodeKey
FROM
    Route
WHERE
    ObjectKey = (SELECT ObjectKey
                FROM Object
                WHERE ObjectCode = 'id5000')
```

By using the relations about nodes and object movements through the nodes, an object can be traced through the supply chain and the environmental burden can be calculated for an individual object. For example, the next query sums up the current attribute value *V* allocated to the board id2000 for each node it travels through.

```
SELECT
    SUM(NodeAttribute.attributeValue /
        NodeProduct.ProductAmount *
        Object.Volume)
FROM
    NodeAttribute
INNER JOIN Route
ON NodeAttribute.NodeKey = Route.NodeKey
INNER JOIN NodeProduct
ON NodeAttribute.NodeKey = NodeProduct.NodeKey
INNER JOIN Object
ON Route.ObjectKey = Object.ObjectKey
WHERE
    ObjectKey = 'id2000' AND
    AttributeKey = (SELECT AttributeKey
                   FROM Attribute
                   WHERE AttributeName = 'CO2')
```

By combining the above-mentioned queries, we can use the traceability graph for tracing the object through the supply chain. The graph allows calculating the resource use and emissions for individual objects. A result is shown in Table 4 where we present the resources and emissions used in our example.

Figure 8 Relation for directed edge

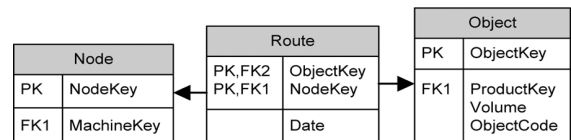


Table 4 Environmental burden of an object

<i>Product: id500</i>			
Resources	Amount	Emissions	Amount
Diesel	0.007 L	CO ₂	1.7 kg
Electricity	3.29 kWh		
Glue	0.2 kg

The graph also enables to track the chain of custody of an object through the supply chain. By using the info-type attribute *CompanyCode*, we can create a chain of custody report for an object. Table 5 illustrates the chain of custody document based on our example.

Table 5 Chain of custody document for an object

<i>Product : id500</i>		
<i>CompanyCode</i>	<i>Process</i>	<i>Date</i>
111	Harvesting	1.1.2009
311	Sawing	7.1.2009
311	Drying	8.1.2009
411	Gluing	9.1.2009

5 Discussion

We presented the traceability graph for modelling and manipulating supply chains of physical products. The present approach enables analysing cumulated properties of products based on the supply chains. The present model enables tracing of different product portions as well as single products. We gave the graphical for the traceability graph and its implementation in relational databases.

Our example was primarily designed for demonstrating the features of the traceability graph, especially for tracing of objects. The example was very simple compared with a real-life application where, for example, are thousands of harvesting nodes in different locations. This means that product portions are mixed in refining processes. In other words, the origin of raw material used in end products may possess various combinations affecting the cumulated values of attributes. For example, transporting distances of raw materials affect considerably emissions of the end products. Further, the originality of the products may be essential. Some product portions of raw materials may be certificated whereas some others may be manufactured by non-ethical methods. Our model also supports the analyses for this kind of information needs. For example, the ratio of non-certificated materials used in single products can be derived.

In the example, we manipulated product portions consisting of quite large sets of objects. However, a process may be focused to a single object. In this case, the underlying product portion would contain a single logical identity. In other words, the present approach also supports very strict modelling of single objects. Furthermore, we did not demonstrate how processes of a mass of products could be traced. Many real processes deal with this kind of material in an intermediate phase in a supply chain. These kinds of masses may also be identified by logical identities but it is also possible to manipulate them without identities, but then some average values must be used in tracing.

Tracing of objects requires strict correspondence between objects on the logical and physical levels. In some application, the identifying of physical-level objects may be a too strict demand because each individual product should be identified by RFID tags. However, if a set of physical products have same features and supply histories, then one physical identifier for a set of physical objects can be used. Now, also one logical object identity can be used to refer to

this identifier. In other words, one logical identity and physical identifier can be used for referring to a set of physical products in some processes in a supply chain. Following this approach, the related information can be maintained without a detailed presentation of the full particulars of products, which allows simpler manipulation of objects at the logical and physical levels. In general, the traceability graph can be used for different granularities of product portions. The precision of analysis depends on the granularity of used marking on the physical level.

The information of tracked objects must be fully available from all actors in the supply chain so the total environmental impact can be calculated. The development of RFID technology has reached a state where standards (EPCglobal, 2007) have been created for sharing the traceability information between the actors in a supply chain. The infrastructure of an information system needed to share the traceability information is not described in this study. Even though this work examines the case where every object is marked with an individual tag, it is also possible to do the marking with different frequencies. For example in some cases it could be more reasonable to mark only the chosen objects and track the chain with the means of estimates. In the example, we have assumed that all participants of a supply chain possess a needed RFID technology. However, if this is missing from a participant, the related process is described by an average node. This, of course, decreases the precision of tracing.

We mapped the traceability graph to relational databases. When mapping the graph to object-oriented databases (Cattell and Barry, 2000), the given relations would roughly correspond to classes in object-oriented databases, whereas cumulated values of attributes could be implemented by methods. In the context of deductive databases (Ramakrishnan and Ullman, 1995), calculation associated with cumulated values of attributes and derived attributes could be represented by rules whereas actual data could be represented by facts. One reason for selecting relational databases was that standard OLAP methods (Chaudhuri and Dayal, 1997) are in use for further analysing purposes. To be able to analyse the huge amount of data that is a result from a traceability system, we created a multidimensional model from the tables presented earlier and created an OLAP cube from the environmental traceability data. The OLAP database provides a possibility to analyse the information much more efficiently and to provide support for complex analytical queries.

The model presented in this paper is being tested in the European Commission's Sixth Framework Programme-funded project, Indisputable Key. In the research project, 10,000 logs are marked in the forest with RFID tags. The logs are then traced through the supply chain using auto-id technologies. The data gathered in the project are used to analyse the environmental impacts of forestry wood supply chain and to improve the efficiency of the supply chain.

6 Conclusions

The traceability graph enables strict tracing of cumulated resources and emissions of products in a supply chain. Processes are represented by nodes (vertexes) that involve product portions for manufactured products. Each product portion involves a ratio for describing the amount of the used resources and emissions related to the underlying process. Attributes represent information related to a process. An attribute has a current value for the underlying process, and a cumulated value derived from previous processes. Interactions among processes are presented by identity shifts (integrity constraints on the graphical level) and propagating value of attributes.

We paid attention to the correspondence between the data management model and real-world applications. That is, the objects are identified on both levels explicitly. In databases, objects are identified by a database solution whereas in the real-world physical RFID tagging. The present approach enables tracing of product portions as well as single objects. This means that for example environmental burden such as greenhouse gases can be analysed for different products and supply chains. The traceability graph is a general method for modelling supply processes, although our real-life example was from the area of forest industry. We mapped the traceability graph to relational databases and demonstrated sample analysing possibilities.

Acknowledgements

This study was supported by the Academy of Finland under grant number 115480. We also thank Swedish Environmental Research Institute for providing information regarding LCA.

References

- Attie, P., Singh, M., Sheth, A. and Rusinkiewicz, M. (1993) 'Specifying and enforcing intertask dependencies', Paper Presented at the *19th International Conference on Very Large Data Bases (VLDB'93)*, 24–27 August, Dublin, Ireland, pp.134–145.
- Bonner, A.J. (1999) 'Workflow, transactions and datalog', Paper Presented at the *Eighteenth ACM Symposium on Principles of Database System (PODS'99)*, 31 May–2 June, Philadelphia, Pennsylvania, pp.294–305.
- Booch, G., Rumbaugh, J. and Jacobson, I. (1999) *The Unified Modeling Language User Guide*, Addison-Wesley, Reading.
- Carbon Trust (2006) *Carbon Footprints in the Supply Chain: The Next Step for Business*, Report CTC616, The Carbon Trust, London.
- Carbon Trust (2007) *Carbon Footprinting*, Report CTV033, The Carbon Trust, London.
- Cattell, R. and Barry, D.K. (Eds.) (2000) *The Object Data Standard ODMG 3.0*, Morgan Kaufmann Publishers, San Francisco.
- Chaudhuri, S. and Dayal, U. (1997) 'An overview of data warehousing and OLAP technology', *ACM SIGMOD Record* Vol. 26, No. 1, pp.65–74.
- Chawathe, S.S., Krishnamurthy, V., Ramachandran S. and Sarma, S. (2004) 'Managing RFID data', Paper Presented at the *Thirtieth International Conference on Very Large Data Bases (VLDB'04)*, 31 August–03 September, Toronto, Canada, pp.1189–1195.
- Dada, A. and Staake, T. (2008) 'Carbon footprints from enterprises to product instances: the potential of the EPC network', Paper Presented at the *38. GI Jahrestagung, LNI 134*, 8–13 September, München, Germany, pp.873–878.
- EC (2000) *The Commission Regulation No. 1980/2000 of the European Parliament and of the Council of 17 July 2000 on a Revised Community Eco-label Award Scheme*, EC.
- EC (2001) *The Commission Regulation No. 761/2001 of the European Parliament and of the Council of 19 March 2001 on a European Eco-Management and Audit Scheme (EMAS)*, EC.
- EPCglobal (2007) *EPC Information Services (EPCIS) Version 1.0.1 Specification, EPCglobal Ratified Standard*, Obtained through the internet: http://www.epcglobalinc.org/standards/epcis/epcis_1_0_1-standard-20070921.pdf [accessed 1/6/2009].
- Gonzalez, H., Han, J., Li, X. and Klabjan, D. (2006a) 'Warehousing and analyzing massive RFID data sets', Paper Presented at the *22nd International Conference on Data Engineering (ICDE'06)*, 3–8 April, Atlanta, USA, p.83.
- Gonzalez, H., Han, J. and Li, X. (2006b) 'FlowCube: constructing RFID FlowCubes for multi-dimensional analysis of commodity flows', Paper Presented at the *32nd International Conference on Very Large Data Bases*, 12–15 September, Seoul, Korea, pp.834–845.
- ISO 14025 (2006) *Environmental Labels and Declarations – Type III Environmental Declarations – Principles and Procedures*, ISO 14025:2006(e), International Organization for Standardization, Geneva.
- ISO 14040 (1997) *Environmental Management – Life Cycle Assessment – Principles and Framework*, ISO 14040:1997(e), International Organization for Standardization, Geneva.
- ISO 14041 (1998) *Environmental Management – Life Cycle Assessment – Goal and Scope Definition and Inventory Analysis*, ISO 14041:1998(e), International Organization for Standardization, Geneva.
- ISO 14044 (2006) *Environmental Management – Life cycle Assessment – Requirements and Guidelines*, ISO 14044:2006(e), International Organization for Standardization, Geneva.
- Krompass, S., Aulbach, S. and Kemper, A. (2007) 'Data staging for OLAP- and OLTP-applications on RFID data', Paper Presented at the *Datenbanksysteme in Business, Technologie und Web (BTW 2007)*, 7–9 March, Aachen, Germany, pp.542–561.
- PE International (2002) *Process or LCI Result Data Set: Timber Pine (40% Water Content)*, Obtained through the internet: <http://lca.jrc.ec.europa.eu/lcainfohub/dataset2.vm?id=113> [accessed 1/6/2009].
- Puettmann, M. and Wilson, J. (2005) 'Gate-to-gate life-cycle inventory of glued laminated timbers production', *Wood and Fiber Science*, Vol. 37, pp.99–113.
- Ramakrishnan, R. and Ullman, J.D. (1995) 'A survey of deductive database systems', *Journal of Logic Programming*, Vol. 23, No. 2, pp.125–149.

- van der Aalst, W. and van Hee, K. (2002) *Workflow Management, Models, Methods, and Systems*, The MIT Press, Cambridge.
- van der Aalst, W.M.P. (1998) 'The application of Petri nets to workflow management', *Journal of Circuits, Systems and Computers*, Vol. 8, No. 1, pp.21–66.
- van der Aalst, W.M.P. and Ter Hofstede, A.H.M. (2005) 'YAWL: yet another workflow language', *Information Systems*, Vol. 30, No. 4, pp.245–275.

Notes

¹Harvester is a vehicle used in logging operation. It produces cut-to-length logs.

²Detailed description on RFID data management is found in Chawathe et al. (2004).

Sustainable ICTs and Management Systems for Green Computing

Wen-Chen Hu
University of North Dakota, USA

Naima Kaabouch
University of North Dakota, USA

Information Science
REFERENCE

Managing Director: Lindsay Johnston
Senior Editorial Director: Heather A. Probst
Book Production Manager: Sean Woznicki
Development Manager: Joel Gamon
Development Editor: Myla Merkel
Assistant Acquisitions Editor: Kayla Wolfe
Typesetter: Lisandro Gonzalez
Cover Design: Nick Newcomer

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: <http://www.igi-global.com>

Copyright © 2012 by IGI Global. All rights reserved. No part of this publication may be reproduced, stored or distributed in any form or by any means, electronic or mechanical, including photocopying, without written permission from the publisher. Product or company names used in this set are for identification purposes only. Inclusion of the names of the products or companies does not indicate a claim of ownership by IGI Global of the trademark or registered trademark.

Library of Congress Cataloging-in-Publication Data

Sustainable ICTs and management systems for green computing / Wen-Chen Hu and Naima Kaabouch, editors.
p. cm.

Includes bibliographical references and index.

Summary: "This book focuses on information technology using sustainable green computing to reduce energy and resources used"--Provided by publisher.

ISBN 978-1-4666-1839-8 (hardcover) -- ISBN 978-1-4666-1840-4 (ebook) -- ISBN 978-1-4666-1841-1 (print & perpetual access) 1. Telecommunication--Energy conservation. 2. Computer networks--Energy conservation. 3. Electronic data processing--Energy conservation. 4. Green technology. 5. Electronic apparatus and appliances--Environmental aspects. 6. Electronic data processing departments--Energy efficiency. I. Hu, Wen Chen, 1960- II. Kaabouch, Naima, 11959- TK5102.5.S925 2012
621.381028'6--dc23

2012002991

British Cataloguing in Publication Data

A Cataloguing in Publication record for this book is available from the British Library.

All work contributed to this book is new, previously-unpublished material. The views expressed in this book are those of the authors, but not necessarily of the publisher.

Chapter 10

Multidimensional Analysis of Supply Chain Environmental Performance

Antti Sirkka
Tieto Finland, Finland

Marko Junkkari
University of Tampere, Finland

ABSTRACT

Monitoring the environmental performance of a product is recognized to be increasingly important. The most common method of measuring the environmental performance is the international standards of Life Cycle Assessment (LCA). Typically, measuring is based on estimations and average values at product category level. In this chapter, the authors present a framework for measuring environmental impact at the item level. Using Traceability Graph emissions and resources, it can be monitored from the data management perspective. The model can be mapped to any precision level of physical tracing. At the most precise level, even a single physical object and its components can be analyzed. This, of course, demands that the related objects and their components are identified and mapped to the database. From the opposite perspective, the authors' model also supports rough level analysis of products and their histories. In terms of the Traceability Cube, multidimensional analysis can be applied for traceability data.

INTRODUCTION

Green computing is usually analogized with Green IT as pure technical issues closely related hardware and software solutions following

Murugesan's (2008) definition "the study and practice of designing, manufacturing, using, and disposing of computers, servers, and associated subsystems—such as monitors, printers, storage devices, and networking and communications systems—efficiently and effectively with minimal or no impact on the environment." Instead, we agree

DOI: 10.4018/978-1-4666-1839-8.ch010

with those authors (e.g. Donnellan, et al., 2011) who see that green computing not only involves the previous definition for Green IT, but also possesses methods to use ICT in business processes to reduce environmental impact of enterprises.

Our contribution is to serve fine-grain tracing approach for monitoring and analyzing life cycle data of products. As our example (forest industry) illustrates our model is general—not only for manufacturing and tracing of computers and data centers. Any physical product has its own supply history represented as a supply chain. In practice, the history of the product is the history of its parts composed in the supply chain. The precision of traceability of products depends on how detailed the history of the components can be traced. We give a logical framework for tracing and analyzing the emissions and resources of products both at the item and patch levels.

The problem with measuring the environmental impact caused by a product at the item level is that supply chains are dynamic. A manufacturer can use various subcontractors and supply various end manufacturers or retailers in different countries. For example, a product that is transported from another continent to a supermarket is bound to have different environmental impact than another product that is transported to a supermarket from a nearby producer. However, the common method of calculating the environmental impact on a product is to measure the resources used, emissions and production in some time period and calculate the average environmental impact on the product. This does not take the dynamic nature of the supply chains into account.

To be able to track the objects through the dynamic supply chain, the products/patches must be identified at the physical level. The development of an auto identification enables us to identify an object moving in the supply chain. This means that we can connect the physical world objects with their virtual counterparts in databases. With the traceability we can track the relationships among properties of processes, in this case the

environmental burden caused by processes, and actual product instances.

Unlike existing methods our model enables analyzing environmental impact on the product level – not only average values. The model supports for monitoring emissions (e.g. CO₂) and resources (e.g. Energy) at any precision level only depending on how precisely physical products and patches can be identified and monitored. Our approach is based on the Traceability Graph (Junkkari & Sirkka, 2011) that enables tracing of products on any level of precision. In practice the method, produce a huge amount of data. Multidimensional methods (e.g. OLAP, online analytic processing) enables viewing this data through several dimensions at different levels of granularity. We design a data cube, called Traceability Cube, for advanced analysis of traceability data.

WORKFLOWS

Generally, workflows are used to model the flow of materials, documents and other pieces of information from one process to another (van der Aalst & van Hee, 2002; Bonner, 1999). Modern software modeling methods such as UML contain activity diagrams for mapping real-world activities to the underlying software solution (or vice versa). There are also a number of commercial applications that have a component for drawing workflow diagrams. The common feature of these applications is that they support the illustration of different types of processes.

The workflows can be divided into two main categories, process- and data-centric. So far, the process-centric workflow modeling focusing on processes and the timing between them has been the dominant approach. However, recently the data-centric workflow modeling has gained popularity. In the data-centric workflow modeling the focus is on the transformation of data sets—initial, intermediate, and final (Akram, et al., 2006). The data sets are used as parameters to services that

consume the input data set and create output data sets. The data-centric workflows are most commonly used in scientific problem solving.

In scientific problem solving the primary feature of the scientific workflow methods is to concentrate on functionality of processes based on the underlying data (Curcin & Ghanem, 2008). Analytical steps instructions how to handle the underlying data sets, i.e. data-centric workflows are used to model, design, and execute an analysis, the capturing series of analytical steps.

Scientific workflows formalize the process of the scientific analysis by modeling the data transportation, transformation, and analytical steps between distributed computational steps. Scientific workflows are directed graphs where arcs represent the transitions between places, which represent computation steps. The main components of the scientific workflow model are:

- Workflow engine invokes the services based on the predefined instructions
- Services which form the workflow are accessed through interfaces.
- Applications implement the functionality of the services.
- Tokens of data are consumed and produced by services.

A scientific calculation process is generally geographically distributed over users and resources and involves huge data sets. Scientific workflows play an important role in Grid computing which combines multiple computer resources to solve a single task, such as geophysics, astronomy, or bioinformatics where the amount of data can be petabytes (Foster, et al., 2001; Chervenak, et al., 2001). The main principles in Grid are the non-centralized control of resources, standardized open protocols, and interfaces. A Grid network is a distributed system that enables sharing anonymous resources based on their availability. There normally is a semantic mismatch between the applications in the Grid network, and calculations

executed by applications in the Grid network can be time-consuming. These requirements must be taken into account when designing scientific workflow systems (Barker & van Hemert, 2007).

The ability to reproduce and reuse results is one of the most important requirements in scientific analysis. To be able to trace the chain of the analysis we need the provenance information, which enables the users of a scientific workflow system to share the results and reproduce the calculation process (Freire, et al., 2008; Simmhan, et al., 2005). The provenance information can be categorized into two main categories (Clifford, et al., 2008). Prospective provenance means information about the calculation steps in a workflow, i.e. the specification of the scientific workflow that needs to be executed. Information about the environment and executed calculation steps used to derive the data sets is called retrospective provenance. The provenance includes the inputs of a workflow, the outputs of the workflow, and definitions of the calculation steps and data tokens between the steps.

The provenance data that is collected from supply chain is used to connect information on processes to products. The example used in this chapter is the environmental burden caused by processes, and actual product instances. The Traceability Graph is not bound to any existing data or workflow methods. The model can be applied using existing formalisms and systems. First, the supply chain is modeled as a basic workflow model. The workflow model provides a starting point for the Traceability Graph—presented in this chapter—which can be seen as a data-centric workflow where each supply chain process is a calculation step—node—in which the life cycle data are allocated to a product. Instead of concentrating on the process functionality, which is one of the main aspects of data-centric workflows (Curcin & Ghanem, 2006), the Traceability Graph emphasizes the handling of aggregation and movement of data between processes. The life cycle information collected from each process

can be shared among supply chain stakeholders using the services as discussed in Section Future Research Discussion.

ENVIRONMENTAL PERFORMANCE MONITORING

The most common method used for measuring environmental performance in organizations is to calculate the total environmental impact for the whole company. The dominant methodologies used are Greenhouse Gas Protocol, which is a guideline for organizations to estimate their greenhouse gas emissions and the Global reporting initiative reporting framework which defines the sustainability reporting framework for the organization. Both methods result in a total environmental burden for a whole company. These resulting values cannot be used to measure an environmental impact for a certain product because the emissions are not allocated to the individual products.

The main approach for assessing the environmental impact for a product is international standard of the Life Cycle Assessment (LCA) ISO 14040 series (International Organization for Standardization, 1997). There are also specifications for the assessment of the greenhouse gas emissions of products. The Publicly Available Specification (PAS) 2050 (Carbon Trust, 2008) that builds on ISO standards for the life cycle assessment by describing the requirements for the assessment of the greenhouse gas emissions and ISO 14067 standard for Quantification and Communication of the Carbon footprint of a product.

Life cycle assessment is a standardized method for calculating the environmental impact caused by a product during its life cycle. The goal of LCA is to compare the environmental impact caused by a product so that the customer can choose the least burdensome one.

In the first phase of LCA, the scope of the life cycle assessment is defined, i.e. which life cycle

processes, inputs, and outputs are included in the assessment. In addition, the functional unit for the assessment is selected. The functional unit can be described as a meaningful amount of a product used in the assessment. For example a cubic meter of timber, one mobile phone or one tomato.

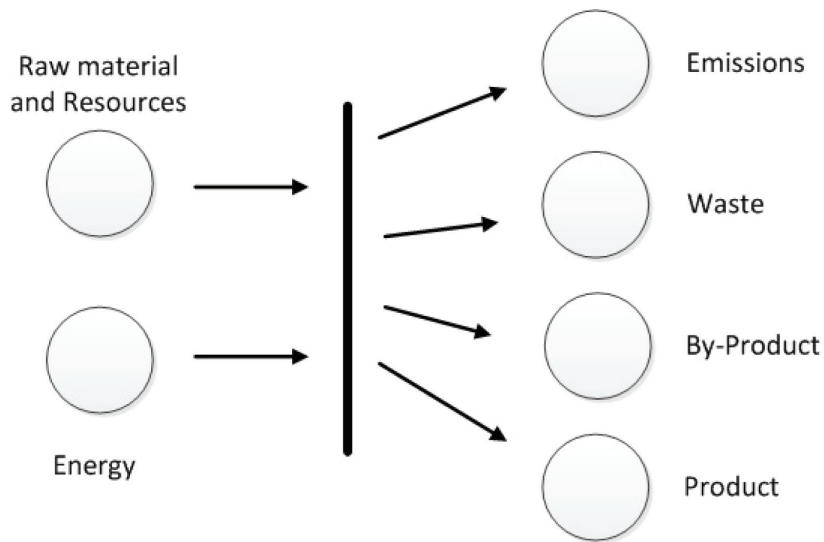
The second phase, called life cycle inventory analysis, the input and output flows for each process is defined (see Figure 1). There are two flow types. Elementary flows describe the inputs and outputs for process (e.g. raw materials and energy resources and outputs of waste and emissions respectively). Product flows are used to describe flow of products and by-products through process. If a process produces more than one product, an allocation is also needed. For example, in sawing the main product is a centreboard with certain dimensions. The sawing process also produces sideboards, wood chips, and sawdust that are considered as by-products. In this case, the emissions caused in the sawing process can be allocated to centreboard, sideboards, wood chips, and sawdust using volume or value based allocation method.

In the third phase, called impact assessment, the results of previous phase are assigned to the impact categories, which include (International Organization for Standardization, 2000):

- Climate change
- Depletion of the stratospheric ozone layer
- Acidification of land and water sources
- Eutrophication
- Formation of photochemical oxidants
- Depletion of fossil energy resources
- Depletion of mineral resources

For example, the emissions to air like carbon dioxide and methane are assigned to the Climate Change category. In the final, interpretation phase, the conclusions of the assessment are made. In this chapter, we present the model for tracing and storing the life cycle data about product manufactured in dynamic supply chain. Unlike the existing methods, our model enables analyzing resources

Figure 1. Process and flows



and emissions at the single product level—not only average values. This is achieved by allowing gathering real monitored activity data from supply chain processes.

IMPLEMENTATION OF LCA PHASES

Our approach supports the different phases the life cycle assessment as follows. In LCA Phase 1, the Traceability Graph is used to model the life cycle processes, inputs and outputs are included in the assessment. The traceability data is collected to a database based on the Traceability Graph in LCA phase 2 - Life Cycle Inventory (LCI) analysis. The phases 3 and 4 are implemented in terms of the Traceability Cube. The Traceability Cube provides abilities to use the OLAP (Online Analytic Processing) type operations (Chaudhuri & Dayal, 1997) for analyzing the information of the Traceability Graph. OLAP is a method used for describing the analysis of the complex data in the data warehouses. OLAP Council has defined three main functions that are provided by OLAP systems: multidimensional views of data, ability

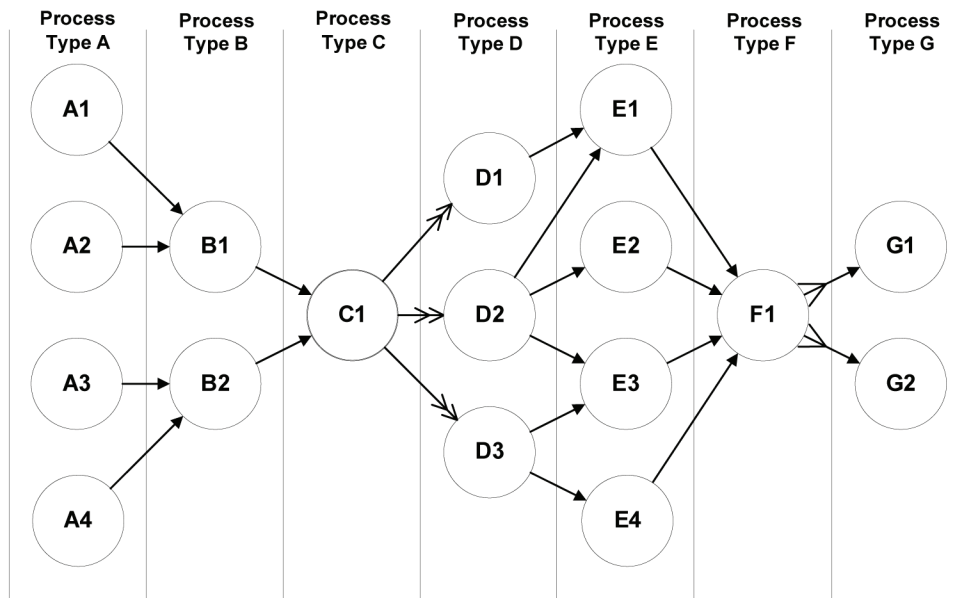
to perform complex calculations and intelligent handling of time, i.e. time intelligence.

LCA Phase 1: Traceability Graph

The Traceability Graph is used to model the supply processes of physical products and resources and emissions associated with the products and their components. The Traceability Graph has the ability to manipulate products and the transformations of the products. For example, a product may be composed from many parts or a product may be manufactured using masses of raw materials. The Traceability Graph is not bound to any existing data or workflow methods. The model can be applied using existing workflow formalisms and systems.

The Traceability Graph has also the ability to manipulate the properties of processes and to allocate them to products that are handled in that process. The Traceability Graph can be presented using nodes and edges and their properties. A node is used to describe a supply chain process. An edge is used to describe a product flow between processes. The supply chain can be viewed as supply processes following each other in a partial order. A manufacturing process is an event that

Figure 2. A sample traceability graph



transforms the input elements (raw material, energy) into output elements (product, waste, and emissions).

In a Traceability Graph, processes can be grouped based on their process types, i.e. similar processes are instances of a process type. Within a process type the specific properties of processes, such as timing, placing, etc., may vary.

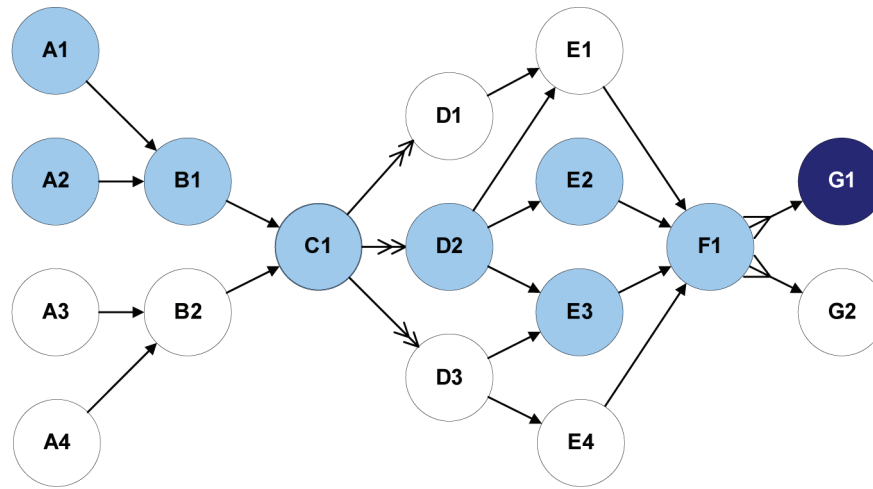
In Figure 2 there are seven process types (A, B, ..., G). Process type A has four instances. These nodes have no predecessor, which means that the traced objects have been created in these nodes. The objects are transferred forward in the graph. For example, objects from Nodes A1 and A2 are transferred to Node B1. Now objects are not changed but object sets (product portions) of A1 and A2 are unionized in B1. This also means that the resources and emissions of A1 and A2 are cumulated to the new set of objects. The objects from B1 are transferred to the Node C1 where products are classified and sent to one of the D processes.

In the D processes objects are divided into several objects. A double-headed arrow illustrates

this. For example a physical object is decomposed or divided into parts. Then, parts may be classified and sent to forthcoming processes. The E nodes receive product portions consisting of these parts. In an E process they are refined and sent to Node F1 which is a shared process for all products. The products of F1 are components for the G processes, i.e. in G1 and G2 objects are composed from the objects that F1 yields. A shared start arrow illustrates this, i.e. several component objects of the start node are needed for a single composition object of the end node.

In a Traceability Graph it is possible to trace the supply chain of an object, i.e. to find all the preceding processes where the object at hand has participated. This also means that all the information related to those processes can be attached to the object. Given the running example, let us assume that we are interested in an object that belongs to Node G1. Then, the processing history of the object is a subgraph of the main graph. In Figure 3 the colored nodes are processes in which the underlying object, its part, or a whole related to the parts has participated. In the example, parts of

Figure 3. The supply chain of an objects belonging to node G1



the underlying object have gone through F1, E2, E3, and D2, whereas the larger objects consisting of the parts have gone through C1, B1, A1, and A2. This subgraph is also the supply chain of the underlying object.

The Traceability Graph can also be used for analyzing different aspects on processes. For example a process type can be selected and we can see how much some process causes the environmental burden. Further, this analysis can be done in a supply chain of one object or a set of objects.

Next, we introduce the properties of edges and nodes of the Traceability Graph.

An *object* is a unit of tracing in a phase of the related supply chain. This can be a single product or a patch depending on the precision of tracing in the underlying supply system.

A process node contains the identity of a process, the set of product portions (patches) and the set of attributes associated with the process. A *product portion* involves the quantity of products, the identifiers of objects and the ratio of the emissions and resources compared with the total ones in the process node. The ratio is calculated by an application specific method. It can be based on the portion of mass or used time of machines, for

example. Product portions of a process are viewed through the end products of a process.

An *attribute* of a node determines information associated with a process. Input attributes describe the resources of a process whereas output attributes can be used for determining the emissions of a process. Each attribute has two values: one for the underlying process and the other for containing the cumulated values from the previous nodes. A cumulated value is calculated based on the ratios of product portions and quantity that is sifted from the previous nodes via edges.

Via edges, products are sifted from a process to another more precisely from a product portion to another. An edge also determines the mapping of objects between two processes. The mapping can be:

1. Equivalence: Objects from a start node of a product portion are sifted to the related product portion of the end node.
2. Subsetting: Only some objects are sifted to the related product portion of the end node.
3. Supersetting: All the objects are sifted to the related product portion of the end node but the product portion of the end node contains similar objects from another process node.

4. Division: Objects of a start node are divided into smaller objects. If an object represents a single product, this is portioned.
5. Composition: Products of the start nodes are components for the end node.

In 1-3 the objects maintains their identities but in 4 and 5, the identities must be changed. In case 4 the identity of a product is mapped with the identities of parts that are produced from the product. In case 5 several objects are needed for one composition object, i.e. the identities of components are mapped with the identity of the related composition. It is worth noting that a product of an end node may contain components from several start nodes.

Through an edge, the information of sifted products from a node to another node is transferred to an end node following the mapping of objects. An edge involves those objects that are sifted from a start node to the end node (only some products of a product portion may be selected from other processes). This part of the product portion of the start node is called a *sifted product portion*. In transferring products from a process to another, the attributes must be re-calculated for corresponding to the sifted product portion. This is based on the ordinary and *derived attributes*. The derived attribute is associated with an edge and it determines the amount of an ordinary attribute that is related to the sifted product portion.

This information collected throughout supply chain processes—nodes—enables to connect the products and processes using the Traceability Graph that can be thought as a provenance model, where provenance is collected as a set of nodes—discrete activities through supply chain processes—that describe the route of an object traveling through a supply chain. Each node includes the following information: location, the time interval when an object was present in the node and elementary flows connected to the process.

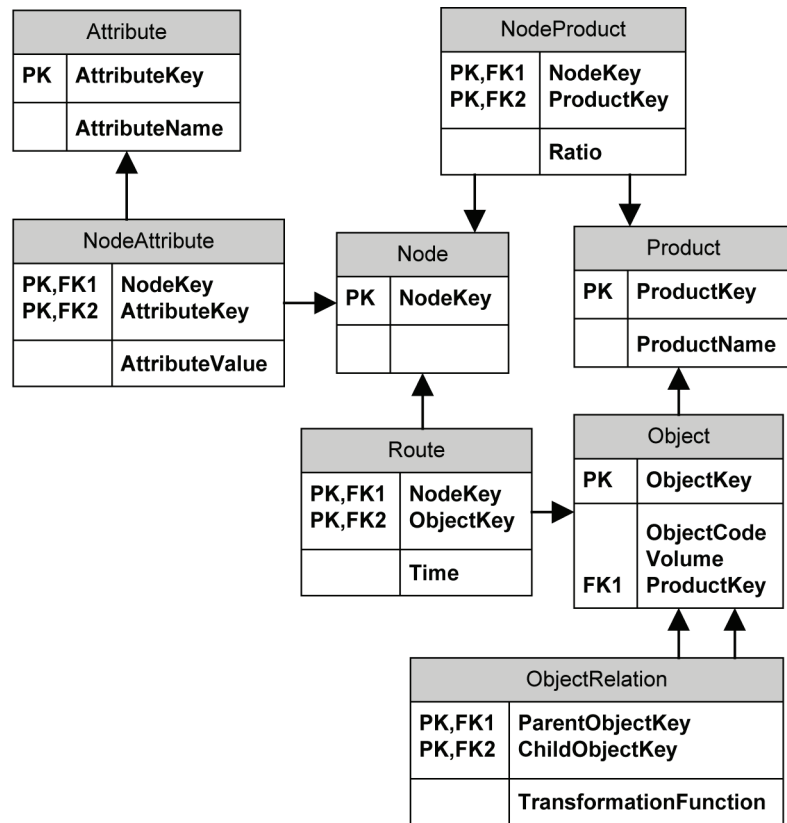
LCA Phase 2: Storing Traceability Data

The Traceability Graph is mapped to the relational database as presented in Figure 4. We selected the relational database because the standard OLAP methods are used to further analyze the huge amount of data that is a result for tracing the individual objects. In Figure 4 PK means primary key and FK means foreign key.

The information of the Traceability Graph is stored into eight relations:

- Node relation is used to store the identities of process nodes.
- Attributes (e.g. raw materials, energy) are stored into Attribute relation.
- Product types are stored into Product relation.
- The relation NodeAttribute is used to store the process (Node) specific attributes. For example □Process#1, Electricity, 100 kWh□ specifies the use of electricity of Process#1.
- The relation NodeProduct is used to store the information about product portions of a specific process (Node). The column Ratio is used to allocate the environmental burden between the portions of products and by-products. For example □Process#1, Product#1, 0.6□ specifies that Product#1 is an end product of Process#1 and the related ratio is 0.6.
- The relation Object is used to store the object specific information like physical code of the object and its volume.
- The relation ObjectRelation is used to store the object mapping when object identities are changed. The column Transformation function is used to calculate the cumulated attribute values.
- The route of the objects through a supply chain is realised by the Route relation. This corresponds to the sifted product portion.

Figure 4. Database schema for the traceability graph



The relation model in Figure 4 can be easily extended to include more product and supply chain specific information. For example, we can implement an organisational hierarchy by creating Process, Site and Organisation relations (Node→Process → Site→ Organisation). This kind of extension enables analysis of environmental data by using the hierarchy as a dimension in multidimensional OLAP model.

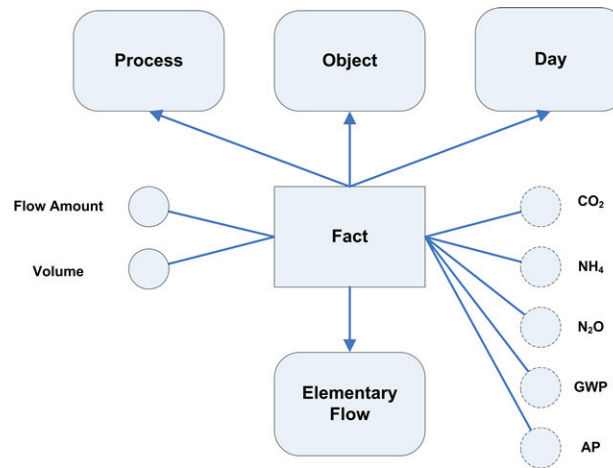
LCA Phases 3 and 4: Traceability Cube

There are three different type of OLAP systems, the relational, multidimensional and hybrid OLAP. The data and dimensions are stored as relational tables in the database in the relational OLAP. The multidimensional OLAP stores the data in the

optimized multi-dimensional array called OLAP Cube. OLAP Cube is a data structure that allows fast analysis of data from multiple perspectives (dimensions). The cube consists of facts that are called measures and dimensions that categorise the facts. The hybrid OLAP mixes the previous types, i.e. the usage of relational tables and cubes can be selected for each case separately (Chaudhuri, 1997). To be able to use the OLAP type operations for analyzing the information of the Traceability Graph we must combine the previous tables as a data cube. Common operations include:

- *Slicing and dicing* the data providing multi-dimensional view of data based on subsets corresponding to the selected dimensions.
- *Drilling down/up* rises or lowers the level of aggregation. For example we can view

Figure 5. Traceability cube schema



the data in day, week, month or yearly level.

- *Pivot/rotate* operation provides alternative view of data. The pivot operations can be used to change the dimensional orientation of the data cube.

In this work, we use the multidimensional data model “MD” that is presented in Torlone (2003). In figures, the dimensions are presented as round-cornered boxes, the facts are presented as boxes, the description of dimensions are presented as small diamonds, and the measures as circles. The

circles drew with dashed line presents calculated measures.

Figure 5 presents the Traceability Cube with some example dimensions. The Process dimension can be used to compare the environmental impact between manufacturing sites and manufacturers. The Object dimension is used to aggregating the environmental data for different product groups. The measure Flow Amount is the amount of elementary flow used in a process in a certain date allocated to specific object. The measure Volume specifies the volume of an object. In Figure 6 some sample instances over the Traceability Cube are

Figure 6. A sample instance over the traceability cube

EPC	Process	Site	Product	Day	Month	Year	Flow	Amount	Volume
1	Kiln Drying	Mill#1	Timber	1	1	2010	Electricity	0,5 kWh	0,30 m ³
2	Kiln Drying	Mill#1	Timber	1	1	2010	Electricity	0,5 kWh	0,35 m ³
3	Kiln Drying	Mill#1	Timber	1	1	2010	Electricity	0,5 kWh	0,33 m ³
1	Transporting	Truck#1	Timber	2	1	2010	Lorry Transport	100 km	0,30 m ³
2	Transporting	Truck#1	Timber	2	1	2010	Lorry Transport	100 km	0,35 m ³
3	Transporting	Truck#1	Timber	2	1	2010	Lorry Transport	100 km	0,33 m ³
4	Kiln Drying	Mill#1	Timber	1	1	2010	Electricity	0,53 kWh	0,40 m ³
5	Kiln Drying	Mill#1	Timber	1	1	2010	Electricity	0,53 kWh	0,42 m ³
6	Kiln Drying	Mill#1	Timber	1	1	2010	Electricity	0,53 kWh	0,38 m ³
4	Transporting	Train#1	Timber	2	1	2010	Rail Transport	200 km	0,40 m ³
5	Transporting	Train#1	Timber	2	1	2010	Rail Transport	200 km	0,42 m ³
6	Transporting	Train#1	Timber	2	1	2010	Rail Transport	200 km	0,38 m ³

Figure 7. Emission and impact calculation: climate change

EPC	Machine	Process	Day	Elementary Flow	Flow Amount	Volume
2	Truck#1	Transport	2.1.2010	Lorry Transport	100 km	0,35 m2
4	Rail#1	Transport	2.1.2010	Rail Transport	200 km	0,40 m2

↓

EPC	Emission	Emission Amount	Unit		Impact Category	Amount	Unit
2	CO ₂	19,9102	kg	→	GWP	19,9102	as kg of CO ₂ eq
2	CH ₄	0,0078544	g	→	GWP	0,19636	as kg of CO ₂ eq
2	N ₂ O	0,000777828	g	→	GWP	0,15401	as kg of CO ₂ eq
4	CO ₂	22,1193	kg	→	GWP	22,1193	as kg of CO ₂ eq
4	CH ₄	0,0055732	g	→	GWP	0,13933	as kg of CO ₂ eq
4	N ₂ O	0,043629899	g	→	GWP	8,63872	as kg of CO ₂ eq

presented. The EPC column is the unique identity of an object.

The Flow Amount is used for calculating the calculated measures – amount of emissions (e.g. carbon dioxide, methane, nitrous oxides) and amount of key environmental performance indicators (see e.g. Lim & Park, 2009). The emission amount is the amount of emissions caused when using elementary flow (raw material or energy) in some process. For example, carbon dioxide emissions when using electricity from Tampere electricity station in Finland were 194 g / kWh in the year 2008. There are many environmental databases that comprise life cycle inventory data from different supply chain processes. For example, the ELCD core database by European Commission - DG Joint Research Centre - Institute for Environment and Sustainability comprises more than 300 process datasets (e.g. key materials, energy carriers, transport, and waste management).

In Figure 7 some emissions and impact category global warming potential for objects with code 2 and 4 are. Key environmental performance indicators are calculated based on the emissions.

The impact category climate change is calculated based on carbon dioxide, methane, nitrous oxide and several other emissions. The measurement unit for the climate change is kg of carbon dioxide equivalent which means that all the other emissions are converted by using a conversion factor. For example the conversion factor of Methane is 25. Full list of emissions and factors can be found from PAS 2050 (Carbon Trust, 2008).

In Figure 8 the impact category acidification of land and water resources and some of the emissions affecting it are presented. The acidification potential is represented in terms of sulphur dioxide (SO₂) equivalents. Thus, all the other emissions are converted to sulphur dioxide equivalents using conversion factors: NO = 1.07, N₂O = 0.7, NO_x = 0.7, NH₃ = 1.88, HCl = 0.88, HF = 1.6 (Heijungs, et al., 1992).

Process Dimension

The process dimension contains the organisational hierarchy and other process specific information, in this example Machine → Process → Site → Organisation. The machine level includes the

Figure 8. Emission and impact calculation: acidification of land and water resources

EPC	Machine	Process	Day	Elementary Flow	Flow Amount	Volume
2	Truck#1	Transport	2.1.2010	Lorry Transport	100 km	0,35 m2
4	Rail#1	Transport	2.1.2010	Rail Transport	200 km	0,40 m2

↓

EPC	Emission	Emission Amount	Unit		Impact Category	Amount	Unit
2	NO _x	8	g	→	AP	5,6	as kg of SO ₂ eq.
2	N ₂ O	0,0048	g	→	AP	0,00336	as kg of SO ₂ eq.
2	NH ₃	0,0072	g	→	AP	0,013536	as kg of SO ₂ eq.
4	NO _x	64	g	→	AP	44,8	as kg of SO ₂ eq.
4	N ₂ O	0,07	g	→	AP	0,049	as kg of SO ₂ eq.

spatial description location that can be used for spatial analysis of environmental burden.

The process dimension can also be used to benchmark between different processes or manufacturers, i.e. it can be used to compact workflow. For example we could merge the transporting nodes of our running example and get the total amount of transporting in our sample supply chain. In an enlarged example, presented in Figure 9, we analyze the effect of different transporting methods by merging the transporting nodes based on the transporting type – the machine type in example dimension.

By comparing the environmental performance of same type of processes companies can see if the performance metrics of their process compare to industry best practices. This allows companies to notice where they could make the biggest improvements in their environmental performance.

Object Dimension

The object dimension is used to describe the properties of object and can be used to benchmark the environmental burden between different products and product types. The total environmental impact

caused when manufacturing a single object is a result when slicing the cube using object level. The slices contain environmental impact from the processes that are extensionally connected with the product (see Figure 10).

By using the slice, we can also get the information about the supply chain of a product. This accurate item level information improves the visibility of the supply chain enabling organizations to improve their product life cycle management. For example in forest industry, the companies are certificating their product using PEFC (www.pefc.org) chain-of-custody certification (Forests area: 229 million ha, Forest owners: > 475,675). The PEFC chain-of-custody certification is a method for tracing wood from forest to the final product to ensure the wood or wood-fibre can be traced back to certified forest. The certification has two methods of realizing this:

- Percentage based method – the method allows mixing certified and non-certified raw materials taking into account that the percentage of certified raw material must be known. Company can sell as certified the

Figure 9. Process dimension and benchmarking

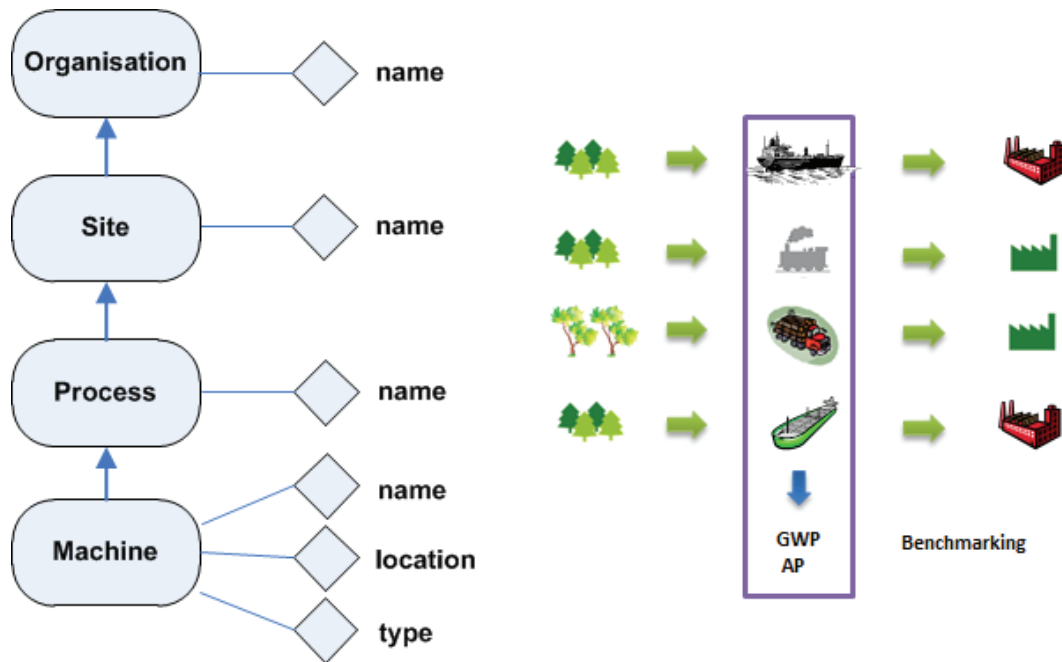
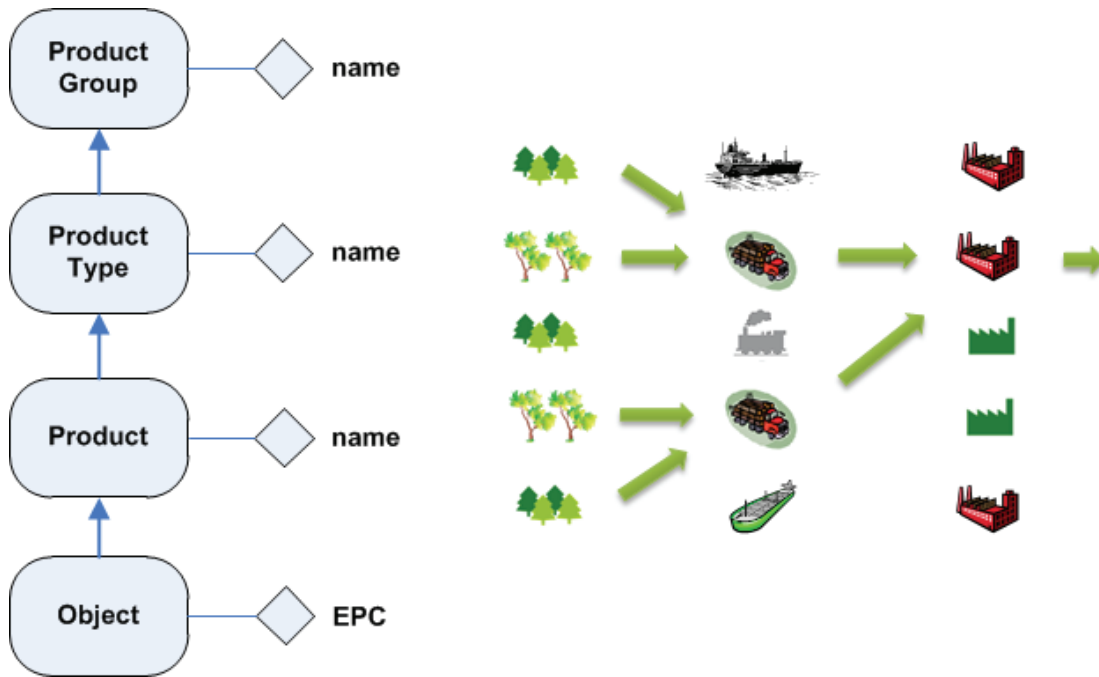


Figure 10. Object dimension and a supply chain of a product



- proportion of its production, which equals the proportion of the certified raw material
- Physical separation method – the method requires certified and non-certified raw material to be physically separated throughout the supply chain.

As seen, the methods can be integrated and improved using the Traceability Cube. The accurate item level information of the origin can be used to create a transparent and trustworthy certification system for origin.

Other usage example is recalling unsafe, defective, or hazardous products. The item level traceability allows manufacturers to recall just the particular products that contain the unsafe elements. For example, Toyota recalled approximately 9 million cars in 2009-2010 for three related recalls pedal entrapment/floor mat problem and accelerator pedal problem. If the reason of recall is for example a faulty set of components, a manufacturer could have only recalled those cars that include the faulty components, not the whole set of manufactured cars in some time period.

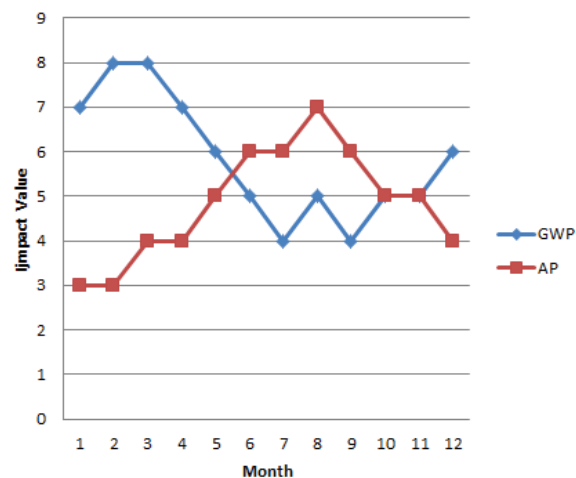
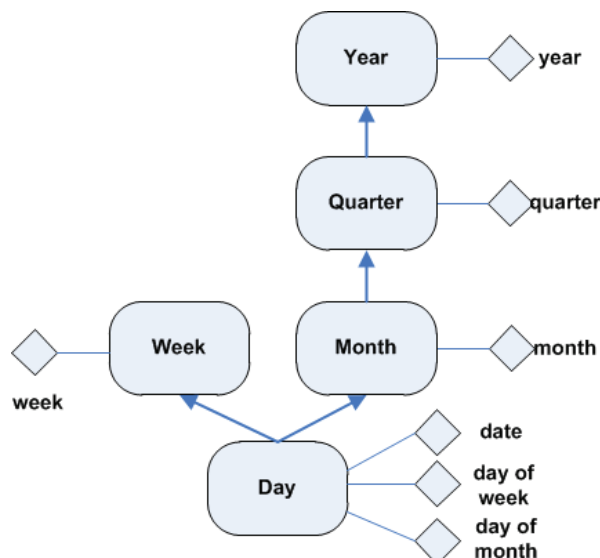
The item or patch level traceability can also be used in food industry to avoid total recalls. The Traceability Cube enables manufacturers, suppliers and resellers to identify the products where some suspected raw material was used and all customers whom the products were delivered. Using the system companies are better equipped to retrieve the affected products and to protect their reputation and brand value.

Date Dimension

In data warehousing the date dimension is the most frequently used dimension. The date dimension is used for analysing the effect of time regarding the environmental impact. For example by using the date dimension, we can perform trend analysis, for detecting a pattern of behaviour in a data series. Also historical and period comparison analysis can be made.

For example, we can track if the environmental burden of a manufacturing facility is diminishing or growing. The trend information can also reveal if environmental burden is affected by seasons or other time related matters. For ex-

Figure 11. Date dimension and trend report



ample the Figure 11 presents a situation where acidification potential is higher during summer months and global warming potential is higher during fall, winter, and spring. The trend information can easily be used for forecasting also the future environmental burden using standard algorithms as time series analysis.

Elementary Flow Dimension

The elementary flow dimension describes the resources and raw materials used to manufacture an object in each process (see Figure 12).

In this chapter, we have used the environmental values as an example. The Traceability Cube can be easily extended to include also economic and social information. By adding a price for an elementary flow we can also monitor and analyze the economic performance of the processes in similar methods than presented above. The social elementary flow can also be added to the system. For example work related accidents happened during the process. This enables us to monitor and analyze the products and their manufacturing process in all three viewpoints of sustainability and thus allowing us to choose the most sustainable way of manufacturing.

The analytics capabilities of the Traceability Cube can be used for analyzing the environmental data. For example, environmental data can

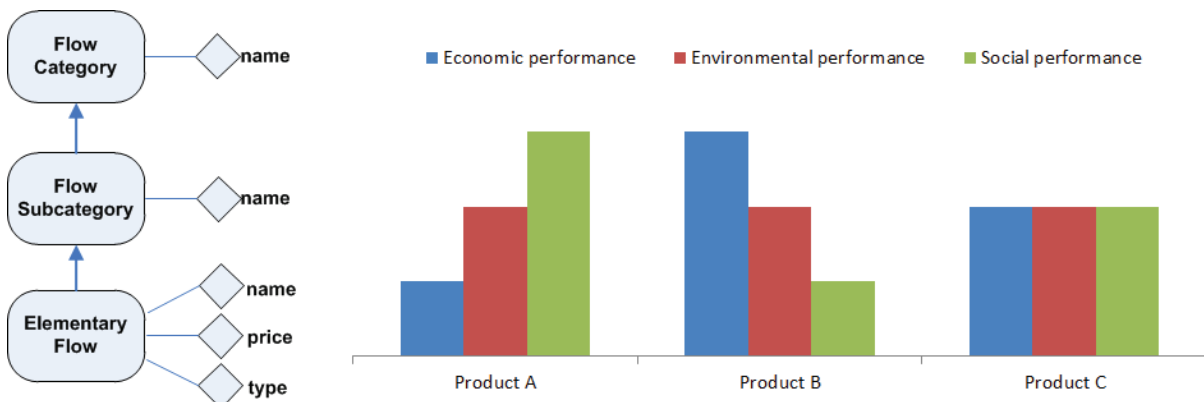
be summed up to create the total environmental impact for the whole life cycle of the product. The data can also be used for comparing the performance between different manufacturers or manufacturing sites. The possibility to analyze the supply chain on the process and item level allows the end users to select a product, which creates least environmental burden. This creates pressure for the manufacturers to improve the eco-efficiency of their supply chains.

FUTURE RESEARCH DIRECTIONS

The precision of traceability of the resources and emissions depends on the underlying data model and ability how strictly physical products and their components can be identified. Our model can be applied to any granularity of tracing. For applications, it is required physical identity mechanism that can be mapped to their logical counterparts in the database.

One option for marking the objects is Radio-Frequency Identification (RFID) technology, which can be compared to the bar code identification: an identification code is embedded to an object. The RFID system can be divided into three components:

Figure 12. Elementary flow dimension and sustainability performance comparison report



- RFID-tag. The main parts of the tag are a chip, an antenna, and a packaging. The chip holds the individual code attached to the physical object. The antenna transmits the information to the reader and the packaging holds the antenna and the chip.
- RFID-reader. RFID-readers read the object identifier from the RFID-tag without needing a clear line of sight. The RFID-reader can be integrated to a machine or it can be just a hand held device.
- Host. The Host computer hosts the application that maps the RFID-data stream from RFID-readers to more understandable form that is useful for the manufacturing execution and enterprise resource planning systems.

Unlike bar code identification, the RFID system does not require a line of sight and the orientation of reader and tag is not significant for communication. Moreover, the reading is automatic and each item is individually labeled with a tag that is more difficult to counterfeit than a simple barcode. With RFID technology a vast amount of tags can be read simultaneously. RFID tags can be active, passive, or semi-passive. The RFID technology and data management has been researched extensively during the last decade—e.g. RFID data staging to OLTP- and OLAP-applications (Krompass et al., 2007) provides a method that answers to different kinds of needs that transactional and analytical applications poses to RFID data, and the RFID data warehousing (Gonzalez, et al., 2006). Dada and Staake (2008) describe a method of presenting the carbon footprint of the product instance to the customers.

The example above presented how environmental impact information can be analysed. To be able to cover the whole supply chain the environmental monitoring system must be modular, i.e. each actor is responsible for generating data from a part of the supply chain of the product. The main challenge of analysing the environmental

impact of a product is the precision of underlying environmental data. In the optimal situation, all the products in the world would be traced through the full supply chain where all processes have smart metering systems for measuring the emissions generated when manufacturing products and this information would be shared through standard interfaces. However, this is not realistic as a starting point.

As discussed in Usva et al. (2009) there is a need for a modular approach where the starting point is to use the environmental impact values produced by current approaches, varying from expert judgments and partial estimates to the usage of aggregated product group data generated from national input-output tables or data from life cycle assessments that is calculated using averages. These default values would be used when the part of the supply chain does not have environmental monitored traceability data available. These values should be defined so that the stakeholders would be encouraged to produce more accurate data. This is achieved by using somewhat higher values than average of the industry.

The modularity of a supply chain means that each actor is responsible for generating data from a part of the supply chain of the product and to share it with other stakeholders. To be able to share product related information in the complex supply chains the organizations have to agree on a common standard. One of the most promising is EPC global Architecture Framework¹ standards, which are generally accepted methods for sharing product data in supply chains. They enable supply chain stakeholders to capture, store and share product related data. The EPCGlobal architecture includes EPC Information Services specification (EPCGlobal, 2007) that defines storing and sharing the traceability data that is created when a product marked with an RFID-tag passes an RFID-reader in a process in a supply chain. This event data normally contains unique identification code, location, and time. By extending the EPC Information Service specification also environmental data

can be included in event data. For example: ‘At location X in time Y the object Z was observed with the environmental data [elementary flow #1, elementary flow #2]’. In Björk et al. (2011), we have given more detailed technical description of sharing environmental data and scalability of the system in real world application.

To be able to generate a total carbon footprint for a product the organizations must share environmental information of the products that were handled by them in their part of the supply chain. For example, in our running example, some organizations are responsible for harvesting the timber; sawmill companies handle the sawing and glued laminated timber manufactures are using the boards sawn in sawmills. All these stakeholders own a part of the final product’s life cycle information. To be able to share environmental information each stakeholder must implement an EPC Information Service that implements the extension for environmental data. To be able to handle the object transformation (division or composition) in the supply chain, the stakeholder responsible for the transformation part of the supply chain is also responsible for aggregating the environmental information from up to that point. In other words, when a manufacturer is further processing products, the manufacturer is responsible for calling the EPC Information Services of a supplier and to add this (derived attribute in the Traceability Graph) information to the environmental information of the further processed product.

To be able to present the environmental data coherently and reliably there has to be a standard set of rules which are enforced by an external auditing party. The ISO 14025 (International Organization for Standardization, 2006) specifies a method for creating environmental declarations and programmes for creating them. One programme operator is an international EPD system, which maintains a specific set of product category rules that provides specifications for creating an environmental product declaration for products.

The rules could be extended so that the best level of the environmental product declaration would be a monitored real-time traceability based environmental data.

When the environmental impact of the whole supply chain is visible to the customers in a reliable and visible way, they can make educated choices of selecting the product with least environmental impact, which would encourage companies to produce data for their part of the supply chain and to optimize their production in a sustainable way.

CONCLUSION

We presented a model how emissions and resources can be monitored from the data management perspective. The model can be mapped to any precision level of physical tracing. At the most precise level, even a single physical object and its components can be analyzed. This, of course, demands that the related objects and their components are identified and mapped to the database. From the opposite perspective, our model also supports rough level analysis of products and their histories. We showed how multidimensional analysis can be applied for OLAP analysis based on the Traceability Graph.

In short, our approach supports the different phases the life cycle assessment as follows. The Traceability Graph is used to model the processes (LCA phase 1). The traceability data is collected to a database based on the graph (LCA phase 2). The Traceability Cube is used to represent the impact assessment and analyzing the results (LCA phases 3 and 4).

We also presented an infrastructure, which allows tracing the environmental burden of an individual product in a complex and dynamic supply chain. The system can be used with any level of physical tracing from using the yearly values to the component level of the product. The model presented can be used for creating a reliable system for measuring an environmental impact of

a single product and to present this information to stakeholders, so that they are able to choose the products, which create less environmental burden

REFERENCES

- Akram, A., Allan, R., & Kewley, J. (2006). Data centric approach for workflows. In *Proceedings of the Tenth IEEE International Enterprise Distributed Object Computing Conference Workshops (EDOCW 2006)*. IEEE Press.
- Barker, A., & van Hemert, J. I. (2007). Scientific workflow: A survey and research directions. In *Proceedings of the Third Grid Applications and Middleware Workshop (GAMW 2007)*, (pp. 746-753). ACM.
- Björk, A., Erlandsson, M., Häkli, J., Jaakkola, K., Nilsson, Å., & Nummila, K. (2011). Monitoring environmental performance of the forestry supply chain using RFID. *Computers in Industry*, 62(8-9), 830–841. doi:10.1016/j.compind.2011.08.001
- Bonner, A. J. (1999). Workflow, transactions, and datalog. In *Proceedings of the Eighteenth ACM Symposium on Principles of Database System (PODS 1999)*, (pp. 294-305). ACM Press.
- Carbon Trust. (2008). *Specification for the assessment of the life cycle greenhouse gas emissions of goods and services, PAS 2050*. Carbon Trust.
- Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *SIGMOD Record*, 26(1), 65–74. doi:10.1145/248603.248616
- Chervenak, A., Foster, I., Kesselman, C., Salisbury, C., & Tuecke, S. (2001). The data grid: Towards an architecture for the distributed management and analysis of large scientific datasets. *Journal of Network and Computer Applications*, 23, 187–200. doi:10.1006/jnca.2000.0110
- Clifford, B., Foster, I., Hategan, M., Stef-Praun, T., Wilde, M., & Zhao, Y. (2008). Tracking provenance in a virtual data grid. *Concurrency and Computation*, 20(5), 565–575. doi:10.1002/cpe.1256
- Curcin, V., & Ghanem, M. (2008). Scientific workflow systems - Can one size fit all? In *Proceedings of the 5th Biomedical Engineering Conference (CIBEC 2008)*, (pp. 18-20). CIBEC.
- Dada, A., & Staake, T. (2008). Carbon footprints from enterprises to product instances: The potential of the EPC network. *GI Jahrestagung*, 134, 873–878.
- Donnellan, B., Sheridan, C., & Curry, E. (2011). A capability maturity framework for sustainable information and communication technology. *IEEE IT Professional*, 13(1), 33–40. doi:10.1109/MITP.2011.2
- EPCglobal. (2007). *EPC information services (EPCIS) version 1.0.1 specification, EPCglobal ratified standard*. Retrieved May 10, 2011, from www.epcglobalinc.org/standards/epcis/epcis_1_0_1-standard-20070921.pdf.
- Foster, I., Kesselman, C., & Tuecke, S. (2001). The anatomy of the grid: Enabling scalable virtual organizations. *The International Journal of Supercomputer Applications*, 15(3), 200–222. doi:10.1177/109434200101500302
- Freire, J., Koop, D., Santos, E., & Silva, C. (2008). Provenance for computational tasks: A survey. *Computing in Science & Engineering*, 10(3), 11–21. doi:10.1109/MCSE.2008.79
- Gonzalez, H., Han, J., & Li, X. (2006). Flow-Cube: Constructing RFID FlowCubes for multi-dimensional analysis of commodity flows. In *Proceedings of the 32nd International Conference on Very Large Data Bases*, (pp. 834-845). Very Large Data Bases.

Heijungs, R., Guinée, J. B., Huppes, G., Lankreijer, R. M., Udo de Haes, H. A., & Sleeswijk, A. ... de Goede, H. P. (1992). *Environmental life cycle assessment of products*. Leiden, Germany: CML.

International Organization for Standardization. (1997). ISO 14040, environmental management - Life cycle assessment - Principles and framework. *ISO, 14040*, 1997.

International Organization for Standardization. (2000). ISO 14020, Environmental LABELS AND DECLARATIONS -- General principles. *ISO, 14020*, 2000.

International Organization for Standardization. (2006). ISO 14025, environmental labels and declarations -- Type III environmental declarations -- Principles and procedures. *ISO, 14025*, 2006.

Junkkari, M., & Sirkka, A. (2011). Using RFID for tracing cumulated resources and emissions in supply chain. *International Journal of Ad Hoc and Ubiquitous Computing*, 8(4), 220–229. doi:10.1504/IJAHUC.2011.043582

Krompass, S., Aulbach, S., & Kemper, A. (2007). Data staging for OLAP- and OLTP-applications on RFID data. In *Proceedings of Datenbanksysteme in Business, Technologie und Web (BTW 2007)*, (pp. 542-561). BTW.

Lim, S.-R., & Park, J. M. (2009). Environmental indicators for communication of life cycle impact assessment results and their applications. *Journal of Environmental Management*, 90(11), 3305–3312. doi:10.1016/j.jenvman.2009.05.003

Murugesan, S. (2008). Harnessing green IT: Principles and practices. *IEEE IT Professional*, 10(1), 24–33. doi:10.1109/MITP.2008.10

Torlone, R. (2003). Conceptual multidimensional models. In Rafanelli, M. (Ed.), *Multidimensional Databases: Problems and Solutions* (pp. 69–90). Hershey, PA: IGI Global. doi:10.4018/978-1-59140-053-0.ch003

Usva, K., Hongisto, M., Saarinen, M., Katajajuuri, J. M., Nissinen, A., Perrels, A., ... Koskela, S. (2009). Towards certifies carbon footprints of products – A road map for data production. *VATT Research Reports*, 143(2).

Van der Aalst, W., & van Hee, K. (2002). *Workflow management, models, methods, and systems*. Cambridge, MA: The MIT Press.

ADDITIONAL READING

Carbon Trust. (2008). *Guide to PAS 2050 – How to assess the carbon footprint of goods and services*. Carbon Trust.

Dada, A., Rau, A., Konkel, M., Staake, T., & Fleisch, E. (2010). *The potential of the EPC network to monitor and manage the carbon footprint of products: Dynamic carbon footprint demonstrators*. White Paper WP-BIZAPP-054. Auto-ID Labs.

Dada, A., Staake, T., & Fleisch, E. (2009). *The potential of the EPC network to monitor and manage the carbon footprint of products: Carbon accounting*. White Paper WP-BIZAPP-047. Auto-ID Labs.

Defra. (2006). *Environmental key performance indicators: Reporting guidelines for UK business department for environment*. London, UK: Queen's Printer and Controller.

Finkenzeller, K. (2004). *RFID handbook fundamentals and applications in contactless smart cards and identification* (2nd ed.). London, UK: Wiley.

Ilic, A., Staake, T., & Fleisch, E. (2009). Using sensor information to reduce the carbon footprint of perishable goods. *IEEE Pervasive Computing / IEEE Computer Society and IEEE Communications Society*, 8, 22–29. doi:10.1109/MPRV.2009.20

International Organization for Standardization. (2006). *ISO 14064, greenhouse gases -- Part 1: Specification with guidance at the organization level for quantification and reporting of greenhouse gas emissions and removals. ISO 14064-1:2006.*. International Organization for Standardization.

Johansson, O., & Pålsson, H. (2009). The impact on Auto-ID on logistic performance: A benchmarking survey of Swedish manufacturing industries. *Benchmarking International Journal (Toronto, Ont.)*, 16(4), 504–522.

Junkkari, M., & Sirkka, A. (2011). Using RFID for tracing cumulated resources and emissions in supply chain. *International Journal of Ad Hoc and Ubiquitous Computing*, 8(4), 220–229. doi:10.1504/IJAHUC.2011.043582

Lim, S.-R., & Park, J. M. (2009). Environmental indicators for communication of life cycle impact assessment results and their applications. *Journal of Environmental Management*, 90(11), 3305–3312. doi:10.1016/j.jenvman.2009.05.003

Ranganathan, J., Corbier, L., Phatia, P., Schmizt, S., Gage, P., & Oren, K. (2004). *The greenhouse gas protocol—A corporate accounting and reporting standard*. WBCSD and WRI.

Simmhan, Y. L., Plale, B., & Gannon, D. (2005). A survey of data provenance in e-science. *SIGMOD Record*, 34(3), 31–36. doi:10.1145/1084805.1084812

KEY TERMS AND DEFINITIONS

RFID: Radio-Frequency Identification (RFID) is a technology that is used to identify an object by using radio wave technology to exchange the identity data between the tag attached to an object and RFID-reader

EPCglobal Architecture Framework: EPC-global Architecture Framework is a collection of standards of hardware, software and data which aim to exchange product level supply chain information transfer through the use of Electronic Product Codes.

EPCIS: Electronic Product Code Information Services standard specifies the EPC data sharing between the supply chain stakeholders. The goal is to enable supply chain stakeholders to gain shared view of the EPC-bearing objects in the supply chain.

LCA: Life cycle assessment is a standardized method for calculating the environmental impact caused by product during its life cycle. The goal of LCA is to compare the environmental impact caused by a product so that the customer can choose the least burdensome one.

LCI: Life cycle inventory is an inventory input and output flows of a certain system. Input flows include input of energy and raw materials. Output flows include releases to land, water and air. The inventory data is related to the functional unit defined in the life cycle assessment.

Environmental Indicator: Environmental Indicator is a measure of performance used to measure the state of the environment

OLAP: Online Analytical Processing is a method used for describing the analysis of the complex data in the data warehouses.

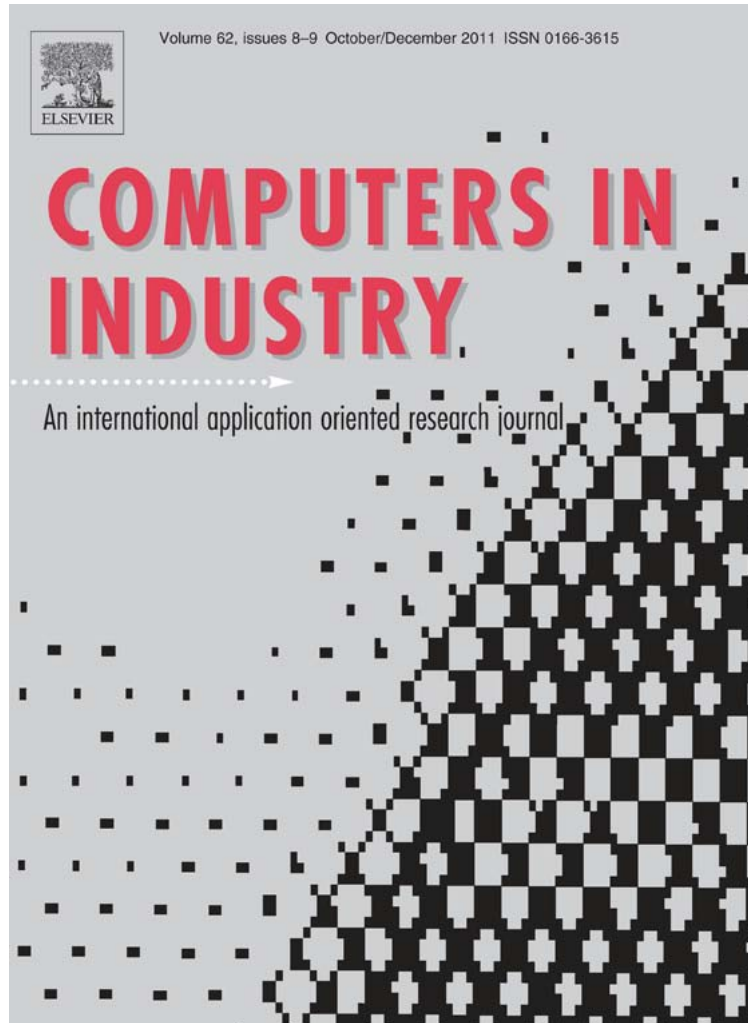
Traceability Graph: A workflow model for tracing an allocating emissions and resources of products in supply chains.

Traceability Cube: A multidimensional model for analyzing life cycle information based on the Traceability Graph.

ENDNOTE

¹ See EPC Global (2011).

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and educational use, including for instruction at the author's institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.elsevier.com/locate/compind

Monitoring environmental performance of the forestry supply chain using RFID

Anders Björk^c, Martin Erlandsson^c, Janne Häkli^b, Kaarle Jaakkola^b, Åsa Nilsson^c, Kaj Nummila^b, Ville Puntanen^a, Antti Sirkka^{a,*}^a Tieto Finland Oy, Hatanpäänvaltie 30, 33101 Tampere, Finland^b VTT Technical Research Centre of Finland, P.O. Box 1000, FI-02044 VTT, Finland^c IVL Swedish Environmental Research Institute Ltd., P.O. Box 21060, SE-100 31 Stockholm, Sweden

ARTICLE INFO

Article history:

Received 9 May 2011

Accepted 2 August 2011

Available online 3 September 2011

Keywords:

RFID

Forestry industry

Environmental monitoring

EPCIS

LCA

ABSTRACT

It is estimated that wood raw material worth of approximately €5 billion is wasted annually in Europe. The major reason for this is that the raw material is not used in the most efficient way as information needed regarding the wood raw material is not available throughout the supply chain. An automatic traceability system makes it possible to utilise raw material information efficiently throughout the forestry-wood production chain and to maximize the raw material yield, and to optimise and to monitor the environmental impact, by linking the relevant information to the traced objects. This paper describes novel RFID technology and traceability solutions that have been developed for the wood products industry. RFID-marking connects the physical objects with their database counterparts thus allowing automatic tracing of the objects. The architecture is needed to the dynamic and decentralised nature of the wood industry. The developed novel RFID based technology allows tracing of individual logs from the tree felling to the sawing of the logs at the saw mill. By combining the traceability and process information systems, new methods are enabled for analysing the performance of the supply chain. As an example, the environmental performance of a product can be traced and analysed even on an individual level. This means that not only the performance from the own production of a manufacturer will be accessible, but also the upstream processes that constitute the product value chain and the life cycle performance for the product leaving the manufacturer.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Radio frequency identification (RFID) provides means to automatically identify objects using radio frequency signals. Passive RFID is based on modulating the backscattered signal from the transponder to the reader. This principle has been used since the 1940s [1] with radar and transponders. More recently passive RFID technology based on the same general principle has found numerous applications in marking and identifying objects, for example in logistics, in access control and in anti-theft devices. RFID is commonly used at low frequencies (LF), high frequencies (HF) and ultra high frequencies [2].

UHF RFID transponders have gained popularity in several applications as they enable long reading ranges up to several meters with inexpensive passive transponders without batteries. Passive transponders are powered by the signal transmitted by the

reader. Even longer read ranges are possible utilising semi-passive or active transponders with a power source (battery) at the cost of increased transponder price. Typically passive transponders consist of an antenna and an integrated circuit (microchip) on a plastic substrate. This inlay is attached to marked object either directly or inside a protective casing. The antenna design for UHF RFID transponders is discussed for example in [3].

The use of UHF RFID technology in the forestry industry is a novel application for this technology. Previously the use of commercially available RFID transponders in the marking of trees has been experimented [4–6]. The conclusion in the trials was that the commercial transponders were not very well suited for large scale tree or log marking as they were not designed for this purpose. HF transponders have been used to mark logs [7,8] but the limited read range of the HF transponders presents challenges in the automatic identification of the logs. In the Indisputable Key project, UHF RFID technology for the use in the forestry industry for automatic and manual log marking was developed [9].

The Indisputable Key project, an integrated project funded by EU in the sixth framework program [10], develops tools and knowledge to enable a significant increase in raw material yield and in utilisation of production resources in the forest and wood industry, thus decreasing the environmental impact. The main

* Corresponding author. Tel.: +46 358 400 268 630.

E-mail addresses: anders.bjork@ivl.se (A. Björk), martin.erlandsson@ivl.se (M. Erlandsson), janne.hakli@vtt.fi (J. Häkli), kaarle.jaakkola@vtt.fi (K. Jaakkola), asa.nilsson@ivl.se (Å. Nilsson), kaj.nummila@vtt.fi (K. Nummila), ville.puntanen@tieto.com (V. Puntanen), antti.sirkka@tieto.com (A. Sirkka).

means to achieve the benefits is the utilisation of traceability at item level, i.e. at log and board level. This traceability requires novel technology solutions. The results are also applicable to other biological raw materials at large extent, thus opening up opportunities for a wider use. The project consortium includes 29 partners from 5 countries: Estonia, Finland, France, Norway, and Sweden.

The traceability systems are implemented at different nodes of the forestry-wood production chain: in forest in harvesters, at saw mills at multiple production steps and at secondary manufacturers using the sawn boards. The traceability at item level – logs and boards, allows monitoring and optimising the environmental impact by linking the relevant information to the individual logs and boards by using Traceability Services (TS) software – the repository for item level traceability data and process level data, and services to access and utilise this information by using selected parts of the EPCglobal EPCIS interfaces [11] (Electronic Product Code, EPC and respectively EPC Information Services, EPCIS).

Passive EPC Class 1 Generation 2 UHF transponders [12] which are pulping compatible and biodegradable were developed to achieve item level traceability for logs [9]. Also robust RFID readers were developed for use in the forest in the harvesters and at the saw mills. Printed ink marking methods were used for the boards.

The developed technology and the benefits of its use are demonstrated in Sweden in a case covering the complete wood supply chain – from the tree felling in the forest to a secondary manufacturer using the boards. Approximately 20,000 logs are to be marked with RFID transponders. Parts of the system are also demonstrated in France and used to study and develop the manufacturing processes in Finland and Norway.

In this paper a prototype traceability system for the forestry industry is presented. Section 2 deals with RFID (Radio Frequency Identification) identification required to trace the wood objects automatically through the supply chain. Section 3 describes the IT architecture that was designed and implemented to capture and share the traceability information. Section 4 presents an example case how the traceability can be used to monitor and analyse the environmental performance of an organisation. Finally, the conclusions are given in Section 5.

2. Traceability in the forest and wood industry

Making information available at different stages along the forestry-wood production chain requires automatic traceability systems. The developed systems are based on the Individual Associated Data (IAD) concept; the measurement and processing data are related to the individual logs or boards and the traceability covers also the traceability of the data associated with the items. For complete traceability, the items have to be automatically identified at all processing steps and all the associated data has to be stored and be retrievable. The associated data – traceability data – is stored and made available with the traceability service which is described in Chapter 3.

Automatic and reliable identification of each item requires a highly readable unique ID-code for each log or board. The forestry and wood products industry sets additional requirements on the item marking technologies: operation in harsh outdoor conditions and industrial environments, suitability for the processing steps of the items, etc. For the logs the selected technology is EPCglobal Class 1 Generation 2 compatible passive UHF RFID-transponders that have a long reading distance and allow a globally unique ID-code for the logs with 96 or 198 bits of data with SGTIN-96 or SGTIN-198 (serialized global trade item number). In the board marking inexpensive ink marking is used and sufficient uniqueness of the ID-code is achieved in each process step for the boards in the production chain at the saw mill.

2.1. RFID system for the log identification

The RFID system consists of the transponders, their applicator to the logs, readers and the middleware. The following requirements are set to the transponder used for log marking:

- High readability
- Pulping compatibility
- Automatic applicability
- Low price

The transponders are attached to the logs automatically by the harvester during the log cutting after the tree felling or manually at later stages of the supply chain, and the logs are identified using an RFID reader and data collected by the harvester associated with them. RFID readers are used also at the saw mill in the log sorting at the log reception and in the saw intake. The special requirements for the harvester reader include the following:

- Tolerance to four-season Nordic weather conditions in the forest.
- Tolerance to extreme shocks and vibration in the operating harvester during tree felling and log cutting.
- Tolerance to liquids, dirt and impacts.
- Operation in proximity of large metallic bodies in the harvester.

The readers at saw mills are subject to industrial conditions including outdoor temperatures, dust, dirt, vibration, shocks, and impacts.

The requirements for the RFID systems are such that no commercial UHF RFID solutions for log tracing satisfying these requirements existed prior to the developed technology.

2.1.1. Transponders

The high readability for automatic log identification requires a long reading distance that is achieved with UHF technology. In addition to the long reading distance, high transponder survivability is needed as only functioning transponders can be read. Therefore, the transponder is inserted into the log so that it is protected by the surrounding wood at different log processing steps before the sawing. Wood as a natural material is challenging environment for the transponder; its electrical properties are strongly affected by the varying moisture content. For automatic application into the wood, the transponder size and shape need to be optimised for penetration into the wood while maintaining a long reading distance.

Wood chippings made from the parts of the logs, which are not sawn into boards are used as raw material at pulp mills. In principle, the material to be pulped may contain no plastics, metal or coal. Transponders for the forestry industry need to be made of pulping compatible materials. The casing of the developed novel transponder for log marking is made of durable artificial wood material that is suitable for the pulping processes, has reasonably low electrical losses, and is relatively inexpensive.

The developed transponder (patents pending) is approximately an 80 mm long wedge that is inserted into the log end in the harvester. The transponder is shown in Fig. 1. A reading range of approximately 2–3 m was measured at the frequency range of 860–930 MHz for the transponder when inserted into fresh moist logs [9].

2.1.2. Transponder applicator

The transponders are inserted into the log using an applicator – a manual insertion tool or an automatic device in the harvester. Both manual and automatic applicators were developed. Manual marking of logs with the transponders is shown in Fig. 2.

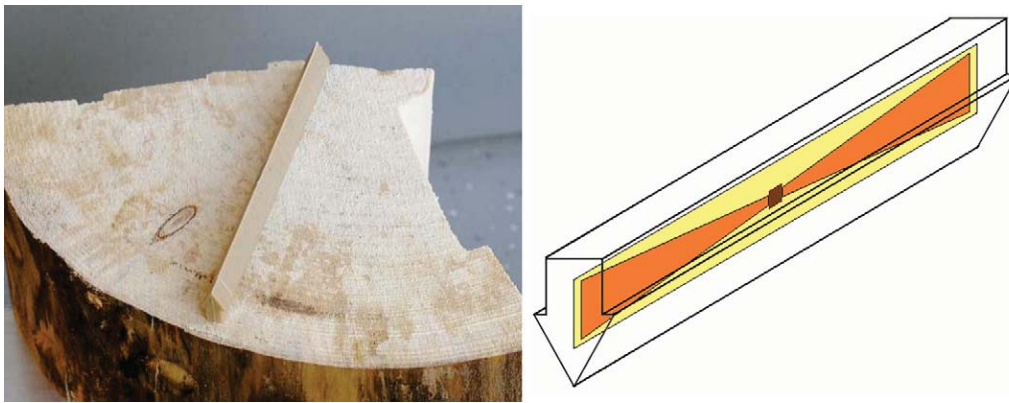


Fig. 1. Developed UHF transponder for marking logs.

2.1.3. RFID readers

The logs are marked and identified when the tree is cut into logs by the harvester. A robust RFID reader with a patented adaptive RF front-end [13] was developed for installation on the harvester head. The reader was tested to survive the outdoor temperatures and to operate under vibrations and shocks at the levels specified in ISO 15003 [14] (2 G vibration at 10–2000 Hz, 50 G shocks) for environmental resistance testing for electronic devices for agricultural machines. The reader is enclosed into robust cast aluminium IP67 casing. A photograph of the reader prototype is shown in Fig. 3 together with a reader set-up at a saw mill.

The RFID reader is controlled over a CAN-bus in the harvester using EPCglobal's Reader Protocol [15]. For implementing the Reader Protocol over the CAN-bus a new Messaging/Transport Binding (MTB) was developed.

At saw mills commercial RFID readers were used with specially developed software for log identification and singulation. The readers were placed into robust aluminium casings to protect them from possible impacts, dirt and dust. Robust metal antennas were integrated to the reader casings. The integrated reader set-ups were positioned over the conveyor at the log sorting where the logs are received at the saw mill and at the saw intake before the logs are sawn into boards at two saw mills; one in Sweden and another one in Finland. The reader position over the conveyor allows the reader installation at the saw mills with minimal changes to the existing conveyors.

2.2. RFID data management

The volume of data generated by storing complete history of movements of individual wood objects (tree–log–board–upgraded product) throughout the supply chain is enormous. To be able to use this information effectively, a data structure that supports effective information retrieval is needed. In order to improve the efficiency of the traceability data warehouse the event data captured from RFID readers and process systems must be cleansed and compressed.

The RFID-reader application reads the RFID-tag constantly when the tag is within the readers range. So the raw RFID data is a stream of observation tuples {reader, EPC, timestamp} where reader is an ID of a reader; EPC is an Electronic Product Code that is defined by the EPCglobal Tag Data Specification [16] and is used to identify individual object; and timestamp is the time instant of the observation.

The normal method used in RFID middleware applications to reduce the amount of the generated RFID data is aggregation and filtering. Filtering includes normally the removal of certain readings based on the reader id or EPC as described as in aggregation types listed by Floerkemeier and Lampe [17].

- Entry and exit. This aggregation is used to reduce the successful reads of the tag to two readings – in and out – i.e. when the object appeared into readers' range and when the object disappeared from the readers' range.



Fig. 2. Manual application of the transponder into the log end.



Fig. 3. Developed RFID reader prototype for the harvester and a test reader set-up at a saw mill.

- Count. This aggregation type can be used to return only a total amount of the similar objects. Similarity can be based to some object property.
- Passage. In some cases the entry and exit observations may be compressed to simple passage observation.
- Virtual readers. The RFID readers may be grouped to function as one “virtual” reader. This enables the application to see them as one reader with bigger read range.

Fig. 4 illustrates entry and exit observation detection. The row A shows the timeframes where the object was in the RFID readers read range. The row B shows the time frames in which the RFID tag was detected by the RFID reader. The raw RFID data must be cleansed to include only the entry event and exit event of an observation. Cleansed RFID observations will be stored in a stay tuple {reader, EPC, time_in, time_out} where time_in and time_out attributes describe the time that item stayed in the location of the reader.

By using the stay table the amount of RFID data will be compressed by proportion of object reads by each reader.

RFID data may be compressed by using the object containment. This idea is presented in [18,19]. The wood objects move in groups inside some containment object during some steps in the forestry wood supply chain. For example logs are grouped as truck loads and transported to a saw mill. Boards are packed to package and transported to customers or green boards are stacked as stick package for drying and storing. In a saw mill that uses oak trees, an oak log is sawn up to 20 boards. The boards are stacked as a boule and the boule is dried for up to four years. During the drying the boule is moved between different locations. If RFID tag is used to mark the entire boule, the movement information of the each board does not need to be tracked and stored every time that the boule is moved.

Same idea was presented by Harrison in [20], where he called it as symbolic location. It means that the object can be contained by parent object. In his paper a parent can also be a symbolic location like certain warehouse or forklift truck. In this paper the parent object is called as containment object and it is separated from the location concept.

The readings made about containment objects like boule, package or truck load can be stored to the same stay tuple {EPC,

location, time_in, time_out} where EPC is a identifier of containment object. However to be able to trace the object to the containment object, a tuple containment {co_EPC, EPC, valid_from, valid_to} is needed, where co_EPC is a code of an containment object, EPC is a code of an object contained. Also the validity of containment must be stored so the valid_from shows when the containment starts and the valid_to shows when the containment relation ceases to be valid.

Usage of containment tuple will compress the amount of saved RFID data by the proportion of objects in containment object and movements recorded for each containment object. For example a boule may contain 20 boards and it could be moved about 20 times during drying process, which means that the amount of rows in the database will be reduced from 400 records to 40 records. By using the methods described above the amount of data that Traceability Services stores to TS repository can be reduced.

3. Monitoring supply chain performance using RFID

In this paper, the traceability is more than just knowing the location of an object. In addition to location information, it is possible to connect the information about movements of the products with the information about processes. In other words, it is possible to allocate the properties of the processes – for example, the environmental burden caused by the process – to the actual product instances involved in each process. To be able to store and derive this information systematically there is need for a model that enables the allocation of process information throughout the supply chain to the traced products. Theoretically, Traceability Services data model, which enables traceability and visibility of products in the distributed supply chain, is based on a general tracing data model called the Traceability Graph [21] developed for strict allocating any information of batches and individual products in supply chains.

By rolling out the Traceability Services the organisation can monitor and analyse the efficiency of its processes and value chain in real time.

The solution connects the steps of the supply chain together and provides a common data model for the whole supply chain. The solution offers services for monitoring environmental, economic and social performance of an organisation.

3.1. System architecture of Traceability Services

The purpose of the Traceability Services is to act as a repository for item level traceability data and process level data and to provide services based on this information. The traceability data is created when object marked with RFID-tag passes a reader and an



Fig. 4. Entry and exit observation detection/passage observation detection.

event is recorded from supply chain. An event data contains unique object identification, location and time, for example: 'At location X in time Y the object with code XXX was observed'. The master data provides context for the event data. For example by using master data the previous RFID-event can be interpreted as: 'At saw mill A log yard slot B in 13.10.2009 the log with EPC code urn:epc:-tag:sgtin:735004747.0001.22605 was observed'.

The system architecture of Traceability Services follows the specifications of EPCGlobal Architecture Framework and is presented in Fig. 5. In order to add process measurement information to events, the EPCIS Standard was extended by utilising the extension mechanism provided by the EPCIS Specification [11]. For example length of the log, volume of the log and electricity usage during process.

Traceability Services Application is an application realizing the EPCIS Accessing Application role and provides organisations the information to carry out overall enterprise business processes aided by traceability data (Tables 1–3).

TS Event Creators implement the role of MeasurementTransaction Capturing Application which is a new role that was developed to unify the handling of events from RFID Readers, Barcode Readers and different Manufacturing Execution Systems. This role uses the new interface, CaptureMeasurementTransaction, specified in Table 4, to deliver MeasurementTransaction Events to EPCIS Capturing Application. MeasurementTransaction Event is a new data type used for exchanging measurement data associated to observed

Table 1
Raw RFID data.

Reader	EPC	Timestamp
r1	epc1	t1
r1	epc1	t2
r1	epc1	t3
r1	epc2	t2
r1	epc2	t3
r1	epc2	t4
r1	epc1	t5
r1	epc2	t5
r2	epc1	t6
r2	epc2	t6

Table 2
Stay table.

reader	EPC	time_in	time_out
r1	epc1	t1	t6
r1	epc2	t2	t6

objects, for more detailed description, see Chapter 3.2.2. The MeasurementTransaction Event type is used as a new event field in the EPCIS Events, see Fig. 6. By using the ALE interface MeasurementTransaction Capturing Application allows integrat-

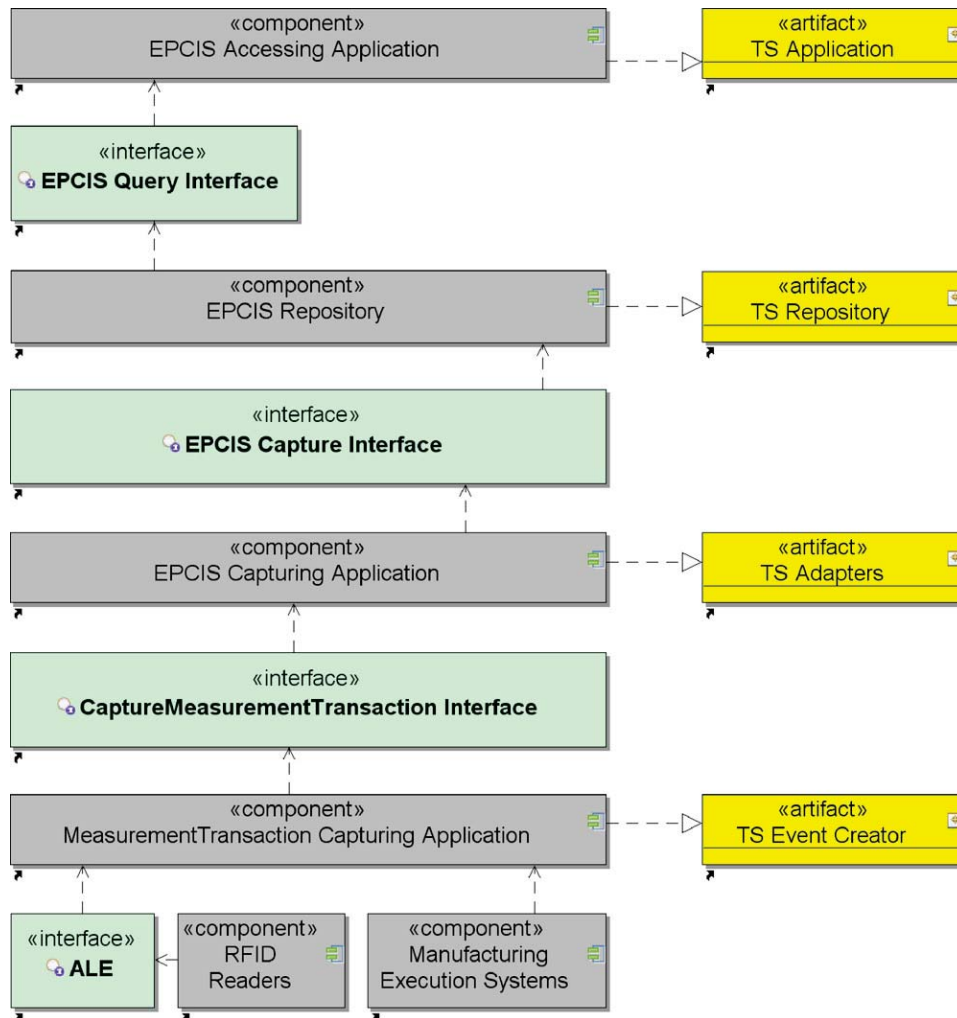


Fig. 5. System architecture of the Traceability Services.

Table 3
Containment table.

co_EPC	EPC	validFrom	validTo
stick_package#1	board#1	1.1.2007	10.1.2007
stick_package#1	board#2	1.1.2007	10.1.2007
stick_package#1	board#3	1.1.2007	10.1.2007
-	-	1.1.2007	10.1.2007
stick_package#1	board#100	1.1.2007	10.1.2007
stick_package#2	board#101	1.1.2007	Null

Table 4
CaptureMeasurementTransaction interface.

```

<<Interface>>
CoreCaptureMeasurementTransactionService
-----
CaptureMeasurementTransaction (measurementTransaction:
List<MeasurementTransaction>):void
<<extension point>>
    
```

ing of EPC compliant readers into the system by supporting the Reader Protocol standard [15].

TS Adapters implement the EPCIS Capturing Application role and capture the occurrences of MeasurementTransaction Events and deliver them to EPCIS Repositories by using the EPCIS Capture Interface. A new EPCIS Event type, Modification Event was created to support the evolving raw material, see Fig. 6. TS Adapter combines the data received via the CaptureMeasurementInterface; reader observations with business information acquired from measurement systems and business applications via. Any Application specific format can be used to connect the business data to the object observations.

3.2. Data Definition Layer

The Data Definition Layer defines the format of the data that is used to exchange data between Capturing applications, Data Repositories and Accessing Applications. To be able to exchange the measurements associated to object observations the Data Definition Layer is extended by introducing a new event type and new event fields to the existing subtypes of EPCIS Event.

3.2.1. ModificationEvent

The Modification Event is used when an object is used to create new objects, for example a log is sawn to multiple boards. The event has rawMaterialEPCs; list of EPCs that were used to create the new object and the result EPC that is EPCs of the created objects. As an example, when three boards (id2, id3, id4) are sawn from one log (id1), three ModificationEvents are created by the EPCIS Capturing Application. Each event has the same value (id1) in the rawMaterialEPCs list as they were manufactured from the same log. The result EPCs has the EPC of the created board. Measurements information related to the created board is contained in Modification Event's Measurement Transaction List field. Description for the types and fields defined by EPC Global can be found in [11]. In the tables below, the new fields are explained (Tables 5 and 6).

3.2.2. MeasurementTransaction

MeasurementTransaction is a new Event Field developed in the project. It contains the measurements made on the physical object. MeasurementTransaction can be either information about Code reads related to read events or measurement information of physical object. MeasurementTransactionItem holds one value of the related to the MeasurementTransaction, this could be for example the length of the log measured by the harvester's

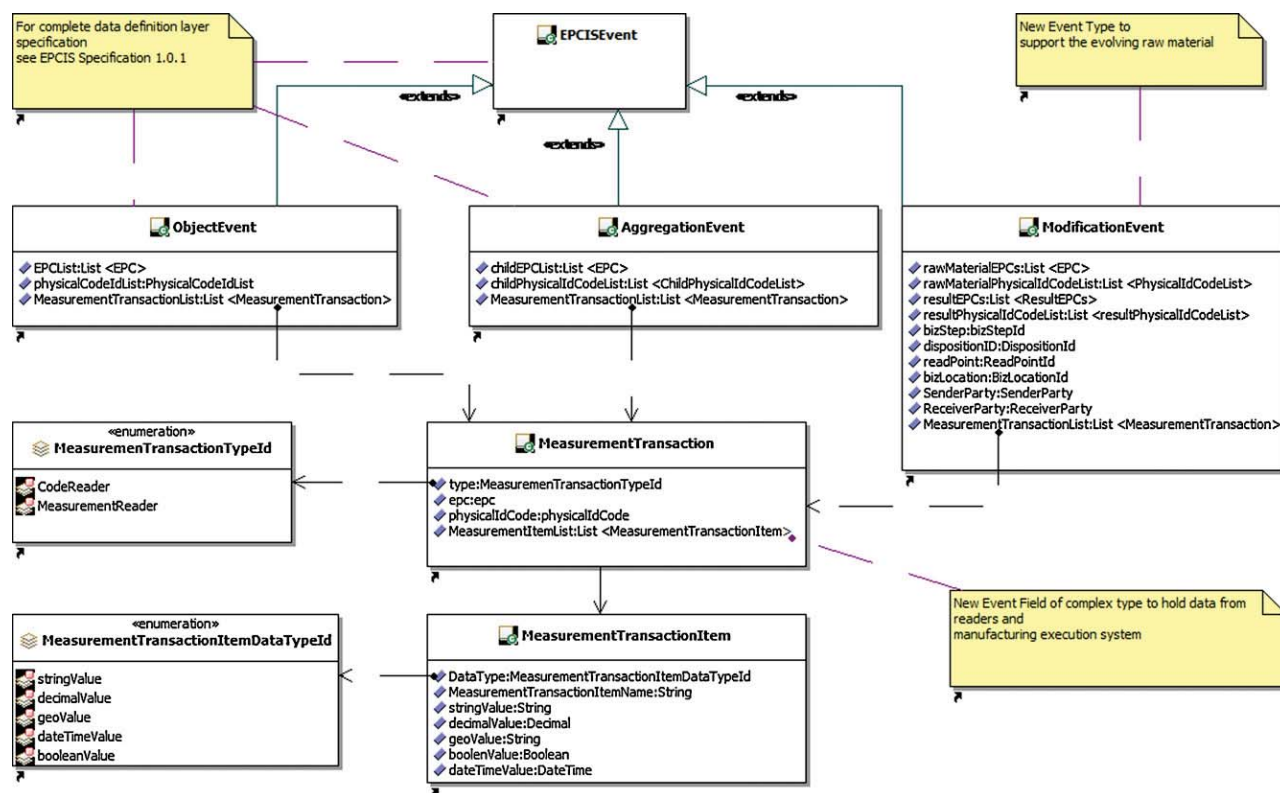


Fig. 6. Data Definition Layer.

Table 5
Action field for ModificationEvent.

Action value	Meaning
OBSERVE	The EPCs named in the raw material list have been used to create the parent object(s) during this event.

Table 6
New event fields for ModificationEvent.

Field	Type	Description
rawMaterialEPCs	List<EPC>	An unordered list of the EPCs of the raw material objects. More detailed specification can be found from [2].
resultEPCs	List<EPC>	An unordered list of the EPCs of the resulting objects. More detailed specification can be found from [2].

measurement device. The fields of the MeasurementTransaction are described in Tables 7 and 8.

The “type” field has two values which are described in Table 7, as enumeration and it describes the type of the Measurement Transaction. This is used for logic in the EPCIS Capturing Application. It enables handling of different Code Readers and Manufacturing Execution Systems as measurement devices, thus making it possible to easily combine both the EPC codes from the readers with measurement data from the measurement devices.

3.3. Traceability data repository

The TS repository provides a common data model for whole supply chain. It is used for storing traceability data and master data from the supply chain. The master data acts as context to event based observation data. The data model for the TS repository extends the traditional RFID traceability data models [18,20,21] by allowing system to store the event data together with the associated process measurement data.

Table 7
MeasurementTransaction type.

Field	Type	Description
Type	MeasurementTransactionType	See Table 8.
physicalIdCode	physicalIdCode	The physical ID code (object marking) is physically attached to the object. The physical ID code can be an RFID tag, punched marks, ink, barcode labels etc.
Epc	EPC	URI denoting the unique identity for a physical object. More detailed description, see [2].
MeasurementTransactionItemList	List MeasurementTransactionItem	Collection of data gathered by the reader.

Table 8
MeasurementTransactionType fields.

type value	Meaning
MeasurementReader	Any device or system that measures physical properties of the object
CodeReader	Any device or system that reads the object marking from the physical object

The theoretical background of traceability data repository is based on the Traceability Graph – detailed introduction in [21] – that enables to model the supply processes of products and patches with associated resources and emissions. The Traceability Graph has the ability to manipulate the transformations of the products. In a supply chain, products or patches may be divided or composed, and, thus, the related resources and emissions must be divided or aggregated, respectively. Resources and emissions are accumulated through a supply chain. The Traceability Graph has an ability to allocate and accumulate these properties of processes through supply chains. In the present paper, the theoretical and technical details of the Traceability Graph are bypassed.

In the visual representation, a process node of the Traceability Graph is illustrated as circle whereas steps (edges) between processes are illustrated by arrows. Three types of edges are distinguished: (1) Plain edge (illustrated by plain arrow), where products are sifted from a process to another as such. (2) Composition edge (illustrated by multi-line start), where the products of the start node are components for the end node. (3) Division edge (illustrated by double head arrow) means that the products of the start node are divided in the end node.

Fig. 7 presents the running example as the visual symbols of the Traceability Graph. The processes included in the example are felling the trees (harvesting), transporting the logs to the saw mill, sawing logs to boards, drying the boards, packaging the boards and transporting packages to secondary manufacturer, where boards are cut to pieces that are used to manufacture a window frame.

In Fig. 7 the coloured nodes (For interpretation of the references to colour in this figure legend, the reader is referred to the web

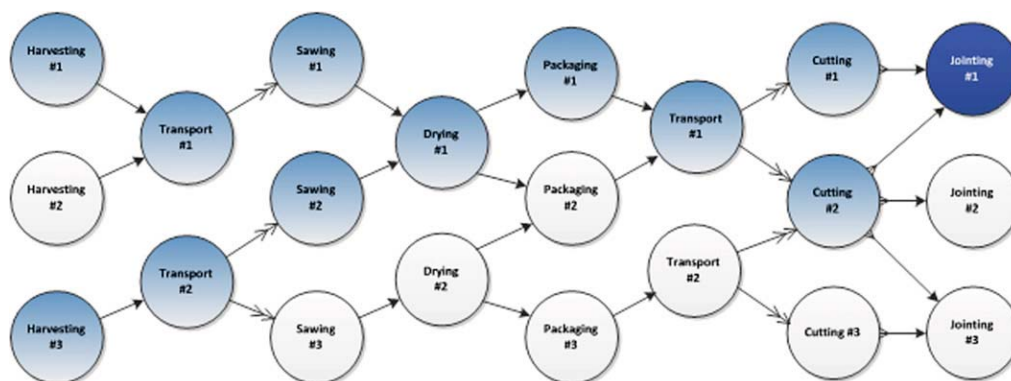


Fig. 7. Example Traceability Graph.

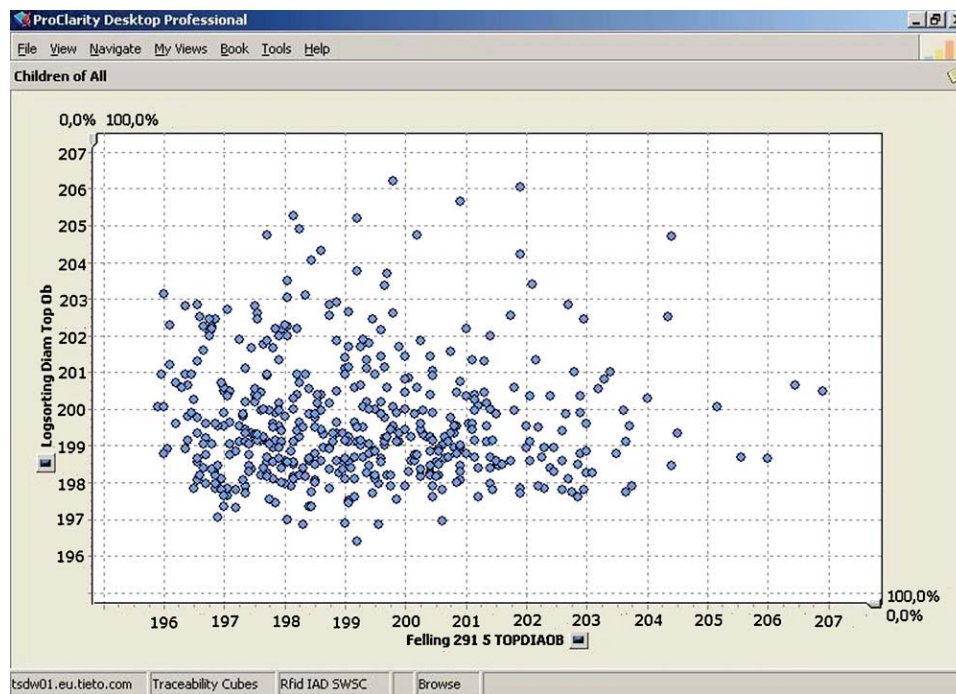


Fig. 8. Analysing the associated data between processes.

version of the article.) are the processes that constitute the supply chain of the window frame. This sub graph of the total Traceability Graph is the supply chain of the window frame.

To be able to analyse the huge amount of data that is stored to the traceability data repository a multidimensional model is created from the tables presented in [21] and an online analytical processing, OLAP cube [23] was created from the traceability data. The OLAP database provides a possibility to analyse the information much more efficiently and to provide support for complex analytical queries.

By combining the process information of the supply chain and the traceability data about products travelling through the supply chain the Traceability Services enables new methods for analysing the performance of the organisation. The properties of products can be compared between different steps; e.g. the measured log length in harvesting vs. the log length in the log sorting at the saw mill. Another possibility is to analyse how a certain product property affects another product property. For example, how the area of origin of the log affects the board quality.

In Fig. 8 there is an example where the user compares log diameter measurements in the forest and in the saw mill. Based on this information a company can calibrate its measurement devices automatically.

In addition to item level property correlation the traceability information can be used to monitor the environmental performance of a supply chain. By combining the information about raw material use, resource consumption, energy use, generation of waste, production volumes and the observations of the object movement through supply chain the environmental impact can be analysed at the item level.

4. Monitoring environmental performance

Continuous monitoring and control of processing conditions and product quality is normal procedure in most of today's industrial production. Introduction of the necessary tools for a

similar follow-up on the environmental performance of the production processes and the generated products gives the industry the opportunity to be more actively involved in environmental protection. Such process application that accounts for environmental performance was developed and applied on a municipal wastewater treatment plant in the EU project HIPCON [22]. This is one part that also is developed in Indisputable Key traceability system and applied and tested in the wood manufacturing industry. The environmental performance related to the products in this system is registered in the traceability system and accessible to the industrial users.

The innovative in this context in the Indisputable Key system scope is that it is now possible to trace environmental data from the value chain to an individual product. This means that not only the performance that the manufacturer can evaluate on-line from the companies own production unit are accessible, but also the upstream processes that constitute the product value chain and the life cycle performance for the product leaving the manufacturer. This is one dimension beyond conceptual pioneer earlier work at Chalmers University, where the scope was to collect data on-line from different industries to a common data domain to make it possible to share the information [23]. The scope elaborated here furthermore supports the full traceability that is essential to proof e.g. origin of harvesting and the legal condition as well as the performance.

On-line monitoring of the environmental performance brings the potential to detect appropriate improvements in the process to limit the environmental impact caused by the production. The system developed for on-line monitoring of environmental performance is not restricted to the wood products industry. The same system layout can be applied in any production chain or process.

The conventional approach for analysing the environmental impact of product manufacturing is to calculate the yearly average impact based on annual information. Compared to that approach the more detailed information recorded in the traceability system, collected from different parts of the supply chain and related to an

individual product or a product batch, provides a much more effective base for proactive environmental management in the industry.

4.1. Environmental key performance indicators, KPI

Key performance indicators (KPI:s) are measurable and quantifiable indicators that reflect the environmental performance of a business or a process. Specific indicators on environmental performance have been suggested by several researchers, and vary both in number and in the way they are calculated. Singhal et al. [24] suggest a number of KPI:s that are different from the ones suggested by the Department for Environment, Food and Rural Affairs in the UK (DEFRA) [25]. The latter suggest a number of KPI:s to facilitate environmental reporting of a business. Other suggestions of environmental KPI:s exist that for instance propose a further integration of the indicators. The number of suggested KPI:s varies widely and while DEFRA suggests twenty-two (22) indicators, Steen et al. [26] suggest that a set of three (3) indicators is sufficient in order to interpret the result. Another way to constitute the indicators is to follow the Eco-Efficiency concept i.e. to define a quota based indicator where the environmental performance is divided by an economical aspect [27].

Within Indisputable Key, it was desirable to select a common set of sector specific environmental KPI:s that makes it possible to benchmark the environmental performance between different companies [28]. In order to support cooperative and supply chain approaches to minimise environmental impact, it was found convenient to use such KPI:s that are possible to sum together from

different steps in the value chain. Such KPI:s makes it possible to calculate the integrated performance in the value chain. Therefore indicators based on established Life Cycle Assessment (LCA) methodology defined in ISO 14040, -44 [29,30] was found adequate for the system.

It was decided to follow the approach established for environmental product declarations, i.e. a set of environmental indicators for the major impact categories found on LCA methodology. The set of environmental KPI:s defined here makes it possible to keep the information presented to the industrial user at a comprehensible level, while still meeting the project objective to create a tool that includes environmental impact from a life-cycle perspective. The majority of the proposed KPI:s were established in line with international LCA practice as it is applied for Environmental Product Declarations (EPD) [29,30], with some additions of specific KPI:s of particular interest for the wood products industry. Eleven (11) indicators were used in the project with the aim to be in line with the forthcoming set of indicators valid for EPD for building products within EC [31]. The selected indicators are listed below:

1. Climate change
2. Acidification
3. Eutrophication
4. Stratospheric ozone depletion
5. Ground level photochemical ozone
6. Depletion of non-renewable resources
7. Human and ecological toxicity
8. Biodiversity

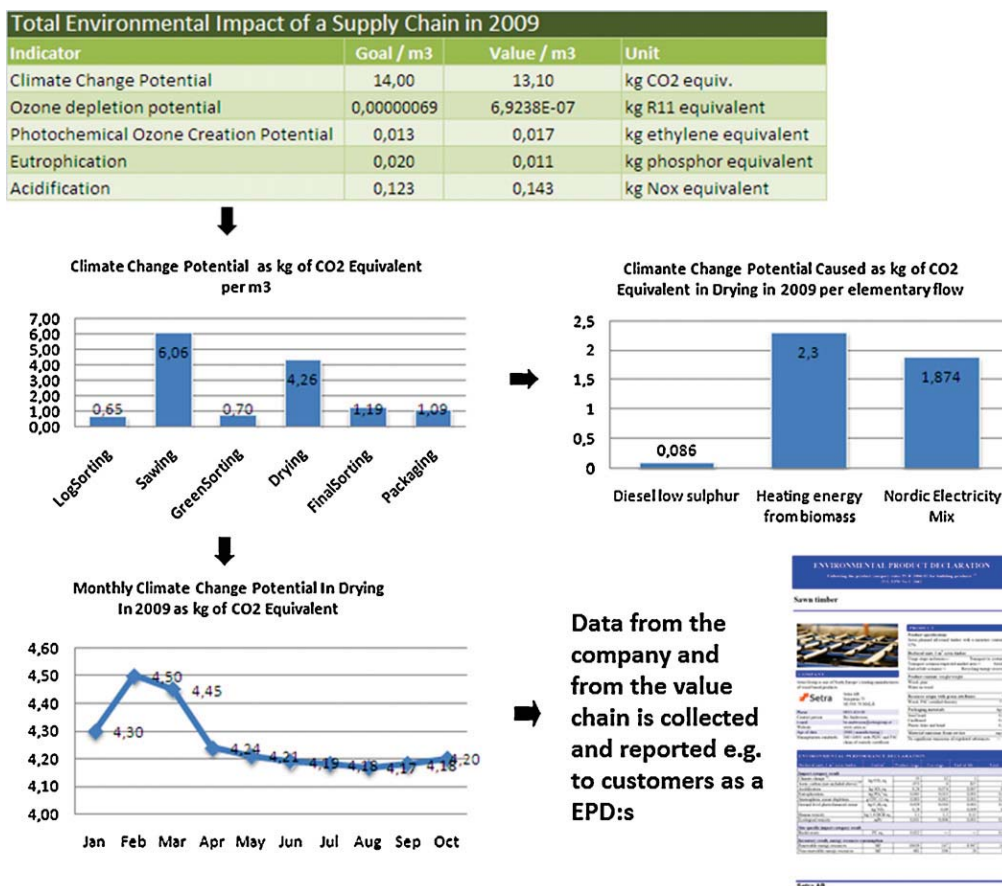


Fig. 9. Environmental impact drill through. Environmental KPI performance may be followed on-line and summarised with upstream information so that e.g. EPD may be produced.

9. Resource use
10. Generated waste
11. Water emissions

An overview of the LCA procedure, creating the base for the selected KPI:s, is presented in [29].

The KPI:s reflect the environmental impact of production. Thus the aim for an industrial user is to keep the values of the KPI as low as possible. In Fig. 9 the analytic potential of the traceability data repository is presented. By using the OLAP methods the traceability data in the traceability data repository can be analysed in different granularity levels.

The user can view the result with different levels of detail. The first table presents the most aggregated level of information which is the total accumulated KPI value for all the stages of the entire production chain. From that it is possible to drill down and get information on the contribution from separate stages (e.g. log sorting, sawing, green sorting, drying, final sorting and packaging). From there it is possible to analyse which of the used resources has the largest contribution to the selected KPI for each production stage. The user can also drill down to a time trend information about emissions and see how the environmental impact has developed.

4.2. Environmental benefits

The environmental KPI:s should be viewed as a tool for the industrial user to enhance the company's environmental management and thus taking an active part in contributing to environmental protection. As such, the tool brings great potential benefits. The actual magnitude of the environmental benefits for an industrial user depends on the involvement and interest of the personnel. Company policies can play an important role. A company with a clearly stated policy for environmental management, with the goal to reduce its impact on the environment, has already the necessary incentive to motivate the personnel into active use of the KPI:s.

In order for a company to take actions towards a more environmental friendly production, the first step is to assess the current environmental impact of the product. When this is known, potential improvements can be identified and actions can be taken. The current status of environmental impact is documented in the initial inventory that precedes the configuration of KPI:s in the Traceability Services. The industrial user can then follow-up on the KPI:s in the TS and benchmark against the inventory results,

aiming for improved production compared to the annual averages that were the outcome of the initial inventory. The important extra benefit from a more detailed follow up on the environmental performance is the possibility to study and take action on variation over time. Allocation of KPI:s to individual items, such as logs and boards, which is the novel idea developed in Indisputable Key adds an extra dimension to the production control and refinement. The information collected in the traceability system enables a continuous improvement and fine-tuning of the production stages.

The methodology developed in Indisputable Key, which introduces traceability in the wood supply chain and a tool for calculating environmental KPI:s, has a major advantage in preventing the risk of sub-optimisation caused by overlooking effects in other parts of the supply chain. Traceability enables the complete view of the entire supply chain, as exemplified in Fig. 9. With this kind of information available the industry can easily identify which parts of the supply chain are the "hot spots" with respect to the environmental impacts monitored by the KPI:s. Collaborative action can then be taken by the actors in the supply chain to reduce the impact, starting with the most critical stages (Fig. 10).

The system architecture developed in Indisputable Key creates a link between the final product and its origin. This makes it possible to trace information up-streams in the supply chain all the way to the harvesting of the tree. In other words, the origin of a wood product (e.g. when and where the tree was cut) is possible to read from the traceability system. This feature can help prevent illegal cutting and it can provide the information needed to guarantee that a product is made from wood coming from certified forests.

The environmental KPI:s harmonize with the required content of an Environmental Product Declaration (EPD) [31], thus facilitating the creation of EPDs for the products of the industrial user. EPDs describe the environmental characteristics of a product or a service from a life cycle perspective. The overall goal is to provide relevant, verified and comparable information to meet the customer and market needs. An EPD can be used for benchmarking the product against other products that could serve the same purpose. In a time with increasing environmental awareness, the fact that a product or service is "green" creates an added value amongst the customers.

Active monitoring and control of the environmental impact of production has the additional advantage that it can raise attention to process improvements that might otherwise be overlooked. The

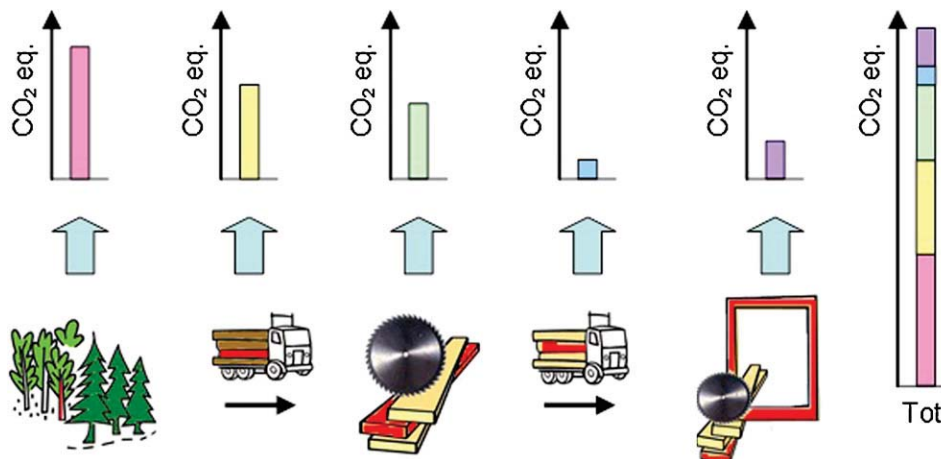


Fig. 10. Information about direct emissions contributing to the KPI Climate Change (in the unit CO₂ equivalents) from production of a window frame is collected from the different stages in the supply chain. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.) The tracked item is illustrated with the colour red.

fact that the KPI:s take into account the environmental burden from manufacturing of resources used by the wood products industry (e.g. electricity, diesel and glue) makes it possible to detect excessive use of external resources. Limiting the use of resources has a positive effect on the environment as well as the economy of production, which is sometimes of higher priority for the industrial user.

5. Economic benefits

Current experiences from different manufacturing industries indicate RFID systems in general are more efficient than other identification systems concerning logistic performance [32]. In the Indisputable Key project the traceability system was implemented at industry gate-to-gate in France. This implementation improved the stock management by reducing errors from 0.6% to 0.3% and the costs of the stock management is estimated to be 70% less using traceability than the old system. Also, the average fill rate of dryers was increased from 77% to 88% [33].

A greater economic potential is supposed if the entire or larger part of the value chain business-to-business (B2B) is covered by the tractability system, covering typically forestry harvesting to a saw mill or a secondary wood manufacturer. The cost of the traceability can be reduced by marking only some portion of the logs. For example by marking few logs per day in harvesting companies can compare log diameter measurements in the forest and in the saw mill. Based on this information a company can calibrate its measurement devices automatically, thus reducing manual work. This kind of information can also be used for research purposes like examining which area provides the best yield at sawmill or boards with best quality.

It has been estimated that the biggest problem in forestry wood supply chain is downgrading of wood, because the information is not available throughout the supply chain. RFID traceability system could reduce this amount remarkably. In the Indisputable Key project in the Swedish supply chain by marking of only 0.5% of the logs with the transponder in forest would enable saw mill to produce more of the most valuable product, thus adding value of the product with 8.5 €/m³ [33]. This improvement is based on using the information of wood properties from the forest to assort logs in a quality classes optimised to fit the final product requirement, closing the information gap between the forest and the sawmill, i.e. the entire value chain B2B.

The cost of the RFID marking consists of the costs of the RFID reader installations and their maintenance and of the price of RFID tags, which is main constituent of the overall cost of the system. The pricing of the tags depends strongly on manufacturing volumes and the typical price range for the tags can be estimated to be of the order of 0.1–1 USD when purchased in large numbers.

6. Conclusions

Novel technology for traceability in the forest and wood industry was developed; pulping compatible UHF transponders for marking logs, robust RFID readers for harvesters and saw mills. The software architecture for acquiring and sharing the traceability information was designed and implemented. The Traceability Services connects the steps of the supply chain together and allows new methods to analyse the performance of the whole chain and any process within it. The developed architecture which is based on EPCglobal architecture allows sharing traceability data within and across enterprises.

The development in the Indisputable Key project has provided the tools necessary for the wood industry to become more proactive with respect to environmental protection. Monitoring of

environmental KPI:s makes it possible to achieve the following advantages for the industrial user:

- Status of environmental performance.
- Detection of potential improvements.
- Basis for optimisation of environmental impact.
- On-line monitoring and control of environmental impact.
- Environmental benchmarking.
- Environmental management.
- Marketing purposes such as EPDs or support for eco-labelling.

The magnitude of the achieved benefits are very much dependant on the industrial users engagement and interest in making use of the developed technology. This study demonstrated the readiness of RFID technology application in the forestry wood industry.

Acknowledgements

This study is performed as part of the project 'Indisputable Key – Intelligent distributed process utilisation and blazing environmental key' with grants from the European Commission within the Sixth Framework Programme, contract No 34732. The RFID development was in part funded by TEKES – the Finnish Funding Agency for Technology and Innovation. We would like to thank all project partners that have contributed to the work.

References

- [1] H. Stockman, Communications by means of reflected power, Proc. IRE (1948) 1196–1204.
- [2] K. Finkenzeller, RFID Handbook Fundamentals and Applications in Contactless Smart Cards and Identification, 2nd ed., Wiley, 2004.
- [3] K.V. Seshagiri Rao, et al., Antenna design for UHF RFID tags: a review and a practical application, IEEE Transactions on Antennas and Propagation 53 (12) (Dec 2005) 3870–3876.
- [4] N. Mohamed, et al., A case study on impact of RFID adoption in tree inventory management, in: 2nd IEE International Conference on Computer Science and Information Technology (ICCSIT), Beijing, (2009), pp. 624–628.
- [5] T.-W. Jung, et al., Research on RFID system recognition test under the forestry environment for tree management, in: Fifth International Joint Conference on INC, IMS, and IDC, Seoul, (2009), pp. 1180–1184.
- [6] S.-S. Lin et al., "POC of RFID application in forest sample zone investigation," in Global Perspective for Competitive Enterprise, Economy and Ecology, S.-Y. Chou, A. Trappey, J. Pokovski, S. Smith, Eds., London, United Kingdom, Springer, 2009, Part 10, Proceedings of the 16th ISPE International Conference on Concurrent Engineering, pp. 541–552.
- [7] R. Uusijärvi, "Environmentally Sustainable and Efficient Transformation processes (LINESET)," SP Technical Research Institute of Sweden, Wood Technology (SP Trätek), QLRT-1990-01476 Final Report, Research Report no. P 0309034, 2003.
- [8] S. Korten, C. Kaul, Application of RFID (radio frequency identification) in the timber supply chain, Croatian Journal of Forest Engineering 29 (1) (2008) 85–94.
- [9] J. Häkli, et al., UHF RFID based tracking of logs in the forest industry, in: Proceedings of the IEEE conference on RFID (IEEE RFID 2010), April, (2010), pp. 245–251.
- [10] Intelligent distributed process utilisation and blazing environmental key (Indisputable Key), EU FP6 project, Project no. 34732, Available from: <http://www.indisputablekey.com>. (accessed 27.10.10).
- [11] EPC Information Services (EPCIS) Version 1.0.1 Specification, EPCglobal Ratified Standard, 2007. Available from: http://www.epcglobalinc.org/standards/epcis/epcis_1_0_1-standard-20070921.pdf. (accessed 27.10.10).
- [12] Class 1 Generation 2 UHF Air Interface Protocol Standard "Gen 2". EPCglobal Ratified Standard, 2008. Available from: http://www.epcglobalinc.org/standards/uhf1g2/uhf1g2_1_2_0-standard-20080511.pdf. (accessed 27.10.10).
- [13] P. Pursula, et al., UHF RFID reader with reflected power canceller, IEEE Microwave and Wireless Components Letters 19 (1) (2009) 48–50.
- [14] Agricultural engineering – Electrical and electronic equipment – Testing resistance to environmental conditions. International Organization for Standardization, ISO 15003:2006, 2006. Available from: http://www.iso.org/iso/catalogue_detail.htm?csnumber=38800. (accessed 27.10.10).
- [15] Reader Protocol Standard Version 1.1, EPCglobal Ratified Standard, June 2006. Available from: http://www.epcglobalinc.org/standards/rp/rp_1_1-standard-20060621.pdf. (accessed 27.10.10).
- [16] Tag Data Standard v. 1.4, EPCglobal Ratified Standard, June 2008. Available from: http://www.epcglobalinc.org/standards/tds/tds_1_4-standard-20080611.pdf. (accessed 27.10.10).

- [17] C. Floerkemeier, M. Lampe, RFID middleware design – addressing both application needs and RFID constraints, *GI Jahrestagung 1* (2005) 277–281.
- [18] H. Gonzalez, et al., Warehousing and analyzing massive RFID data sets, in: *Proceedings of the 22nd International Conference on Data Engineering*, Atlanta, GA, USA, (2006), p. 83.
- [19] L. Shaorong et al., "A Temporal RFID Data Model for Querying Physical Objects", A TIMECENTER Technical Report, February 2007. Available from: www.timecenter.cs.aau.dk/TimeCenterPublications/TR-88.pdf. (accessed 27.10.09).
- [20] M. Harrison, "EPC information service – Data Model and Queries", White paper, Auto-ID Centre for Manufacturing, University of Cambridge, February 2003.
- [21] S. Sudarshan Chawathe, et al., Managing RFID data, in: *Proceedings of the Thirtieth international conference on Very large data bases (VLD'04)*, Toronto, Canada, (2003), pp. 1189–1195.
- [22] Holistic Integrated Process Control (HIPCON), EU FP6 project, NMP STRP 505467-1, available from: <http://www.hipcon.org>. (accessed 27.10.09).
- [23] R. Carlson, et al., System for integrated business environmental information management, *Advances in Environmental Research* 5 (4) (2001) 369–375.
- [24] P. Singhal, et al., Key environmental performance indicators (KEPIs): a new approach to environmental assessment, in: *Proceedings of Electronics Goes Green 2004 (Plus): Driving Forces for Future Electronics*, Berlin, (2004), pp. 697–702.
- [25] Environmental Key Performance Indicators: Reporting Guidelines for UK Business. Department for Environment, Food and Rural Affairs (DEFRA). Queen's Printer and Controller, Product Code PB 11321, January 2006.
- [26] B. Steen, et al., Development of interpretation keys for environmental product declarations, *Journal of Cleaner Production* 16 (5) (2008) 598–604.
- [27] S.-R. Lim, J.M. Park, Environmental indicators for communication of life cycle impact assessment results and their applications, *Journal of Environmental Management* 90 (11) (Aug 2009) 3305–3312.
- [28] M. Erlandsson et al., "D3.8: Environmental product declarations (EPD) for selected products," Indisputable Key Project no: 34732, European Commission within the Sixth Framework Programme, Second release.
- [29] Environmental management – Life cycle assessment – Principles and framework, International Organization for Standardization, ISO 14040:2006, 2006. Available from: http://www.iso.org/iso/catalogue_detail?csnumber=37456 (accessed 27.10.10).
- [30] Environmental management – Life cycle assessment – Requirements and guidelines, International Organization for Standardization, ISO 14044:2006, 2006. Available from: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=38498 (accessed 27.10.10).
- [31] Sustainability of construction works – Environmental Product Declarations – Product category rules". European Committee for Standardization (CEN), prEN 15804. CEN technical committee CEN/TC 350, working draft.
- [32] O. Johansson, H. Pålsson, The impact on Auto-ID on logistic performance: a benchmarking survey of Swedish manufacturing industries, *Benchmarking: An International Journal* 16 (4) (2009) 504–522.
- [33] Å Nilsson et al., "D3.11b. Environmental and economic impact of implementation of developed approach and on further possibilities and challenges in wood chain traceability". Indisputable Key Project no. 34732, European Commission within the Sixth Framework Programme, Second release. 2009.