

## Managing LCI Data from Different Workgroups within the same Instance of an LCA Database

Clemens Döpmeier<sup>1</sup>, Claudia Greceanu<sup>1</sup>, Oliver Kusche<sup>1</sup>, Christian Schmitt<sup>1</sup>

### Abstract

Life Cycle Thinking (LCT) and Life Cycle Assessment (LCA) are the scientific approaches behind modern environmental policies and business decision support related to Sustainable Consumption and Production. For the efficient and reliable application of LCA, it is necessary that LCA studies are based on consistent, transparent, and quality-assured Life Cycle Inventory (LCI) data. To manage and provide such consistent and quality assured Life Cycle data to users the Institute of Applied Computer Science of the Karlsruhe Institute of Technology developed an open source service oriented LCA database software (soda4LCA) that can be accessed directly from within LCA tools to store, manage and extract LCA datasets via the Internet. Different soda4LCA instances can even be connected together via the Internet to form a LCA data network which can then be used by users via one of the network nodes. Because of user requirements, the storage model of soda4LCA is enhanced in the current development version to support the management of largely separated collections of datasets (called data stocks) by different working groups. This can be seen as a first step to support multitenancy in soda4LCA. It's also possible in the current development version to build other logical collections of datasets for such use cases as compiling collections of datasets for review or release management. Data stocks are also the key element of soda4LCA to provide dataset replication via nodes of a soda4LCA data network. Further details of the data stock concept and its usages will be described in the following chapters.

### 1. Introduction

Life Cycle Thinking (LCT) and Life Cycle Assessment (LCA) are the scientific approaches behind modern environmental policies and business decision support related to Sustainable Consumption and Production (Europe 2011a, 2009, 2003). For the efficient and reliable application of LCA, it is necessary that LCA studies are based on consistent, transparent, and quality-assured Life Cycle Inventory (LCI) data. Because of this, several activities in Europe and worldwide have been established, which are aimed at providing national and international LCI databases. Examples in Europe are the Suisse Ecoinvent database (Ecoinvent 2011) and the European Platform on LCA (Europe 2011b) for a Europe-wide data supply.

As already described in a paper presented at the EnviroInfo 2011 conference (Döpmeier 2011a), the Institute of Applied Computer Science (IAI) of the Karlsruhe Institute of Technology (KIT) developed an open source service oriented LCA database software for the International Reference Life Cycle Data System (ILCD) Data Network of the EU (Wolf 2010) that allows to connect different autonomously installed database nodes together using a RESTful Service API to form an integrated network of LCI databases with harmonized data (see Figure 1).

---

<sup>1</sup> Karlsruhe Institute of Technology, Institute for Applied Computer Science, Postfach 3640, 76021 Karlsruhe;  
e-mail: {clemens.duepmeier, oliver.kusche}@kit.edu

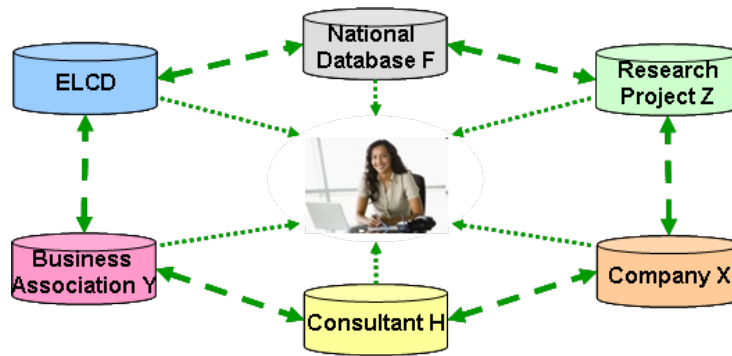


Figure 1

Data network with different LCA databases, like the European Life Cycle Database (ELCD) or other national, research or company databases

Within such a network, the datasets provided by different parties can reference common background datasets and nomenclature standardized by the corresponding network community and provided by one or more central systems within the network.

While the exact details of how such a network can be organized within the soda4LCA software would be an interesting topic on its own, which will be presented at this EnviroInfo conference in another paper (Kusche 2012), this paper will focus on a related and somewhat opposed functionality highly demanded by users: the possibility to store on the one hand separated but on the other hand interrelated LCI data collections created and managed by different working groups within the same database instance of soda4LCA.

In practice, a soda4LCA database instance may not only contain different data collections of LCI data which are a work in progress of certain groups creating and maintaining such data, but may also contain released datasets which likely are already used by other parties. It's even a quite common scenario that a certain released version of a dataset is used in studies by LCA experts, while the maintainers already work on an updated version of this dataset. Therefore, concepts for version and release management of datasets and an accompanying work flow and functionalities for the management of largely separated collections of datasets (called *data stocks*) created and maintained by independent working groups were developed and implemented in soda4LCA.

## 2. Root Data Stocks and Dataset Management

The data stock concept of soda4LCA is based on the idea to group datasets together to form specific sets of interrelated datasets within a soda4LCA instance. At the moment two types of data stocks are defined for soda4LCA: *root data stocks* and *non-root data stocks*.

Root data stocks allow to create working databases within a soda4LCA instance which contain datasets owned and maintained by a certain group of LCA experts. For example, an industrial association could use a central soda4LCA instance administrated by an IT-provider to maintain their own root data stock of LCI datasets related to plastic products (datasets within the left sub-cylinder of Figure 2) while another association could use the same soda4LCA instance but their own and different root data stock to create and maintain datasets related to aluminium products (datasets within the right sub-cylinder in Figure 2).

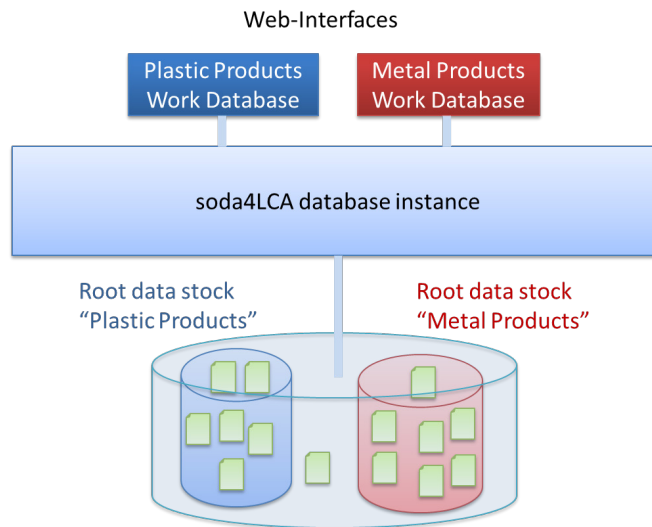


Figure 2  
Soda4LCA instance containing different root data stocks for different groups of LCA experts

Each dataset within a soda4LCA instance and all of its versions belong to exactly one root data stock which is the natural home of this dataset and its different versions. The handling and maintenance of datasets in root data stocks is largely separated from each other within the same soda4LCA instance. While a soda4LCA instance contains only one user database, the access right management system of soda4LCA allows it to assign different access rights to a user or a user group related to different data stocks. Thus, while one user or group may have rights to add, change or manage datasets in a certain root data stock, the same user or group may only have rights to view datasets of another data stock or they even don't have any access to the datasets in another data stock at all. Using the concept of release data stocks, which will be explained in a later section of this paper, a user or group can even be given different access rights for different versions of the same dataset. So, a certain released dataset within a data stock can be made publicly available while a newer updated version being worked on is only viewable and editable by the group of maintainers of the root data stock to which the dataset belongs.

The version management system for datasets within a soda4LCA instance allows the management of several distinguished versions of one and the same logical dataset within its root data stock. According to the ILCD format standard, the version number of a dataset (for instance, 01.01.005) consists of three parts, called the major, minor and sub-minor version numbers of the dataset, respectively. soda4LCA does not enforce any rules on how the different parts of a version number should be used by teams organizing datasets, but a common approach would be to use sub-minor version numbers for error corrections to datasets which don't change the semantics of datasets while minor and major version numbers should be incremented for smaller or larger semantically changes to the datasets, respectively.

The dataset management of soda4LCA treats each version of a dataset as single entity within the database. While some important metadata of a dataset will be extracted into a separate database schema for easy management, search and navigation between datasets, soda4LCA will always store the complete XML representation of each dataset version in the database as well. Because soda4LCA is designed to be a service oriented database system that treats dataset versions as single entities, editing operations on datasets are foreseen to mainly happen in LCA tools outside of the database system itself and these tools ex-

change complete versions of datasets with the soda4LCA service. Thus, the main operations related to datasets within soda4LCA are import, export, check-in and checkout operations as shown in Figure 2.

According to the state diagram shown in Figure 3, datasets can be in different states related to main operations on datasets. In *standard mode* an *import* of an external dataset into soda4LCA will result in a new dataset entity (creation of new entities) within soda4LCA if the database does not already contain a dataset entity with the same identifying UUID and version number. Otherwise an error will be thrown that a dataset entity with the same version number and UUID already exists in the database. If needed, the import mode can be switched to *overwrite mode*. In this import mode, an already existing dataset version will be overwritten within the database but only if it is in the state UNRELEASED.

An import operation will assign new dataset entities always to exactly one root data stock determined by the import request. The client performing the import operation needs import rights on this root data stock to perform the operation; otherwise an error will be thrown. Datasets will be in a state UNRELEASED when imported into the database for the first time. Until they are switched to the state RELEASED by a data stock manager they can be overwritten by a modified version or more commonly checked out for changes.

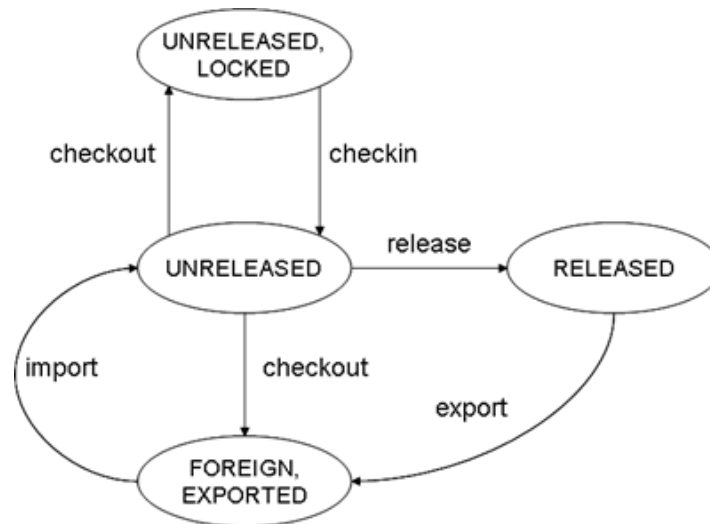


Figure 3  
Authoring workflow for dataset versions in a root data stock

A *checkout* operation exports a dedicated version of a dataset (dataset entity) and locks the entity within the database for further operations. The dataset will then be in the state “UNRELEASED, LOCKED”. Tools can use this checkout operation to extract a specific version of the dataset for editing within a tool. While a dataset version is checked out by one client user for editing, other clients trying to check out the same dataset will get an error notification stating that the dataset is already checked out by another client user. To make this message more precise, the soda4LCA service stores information about the owner of each lock which can be shown to other users requesting a checkout for the dataset. The lock on a dataset will be freed when the lock owner performs a *check-in* operation on the dataset.

During a check-in operation, the content of the LOCKED dataset will be replaced by the modified content after editing and the state of the dataset will then be reverted to UNRELEASED again. Checkout and check-in implement a transaction concept on datasets which is called *long transaction*. Contrary to common database transactions, long transactions are often used on document oriented content because the

changes to such data entities can take a very long time from several minutes, to hours, several days or even weeks.

At any time a client having rights to do so can request a dataset using an *export* operation. The export operation extracts a copy of the requested version of a dataset from the soda4LCA instance and returns it in ILCD format to the client. There is no state change associated with an export operation within the database. Note, that an exported dataset cannot be returned to a soda4LCA instance by a check-in operation because no lock is associated with the exported dataset. If an exported dataset should be stored beside the exported version in the same database after some modifications have been applied, this can only be achieved by an import operation. Normally, the dataset's version number should be increased before such an import because otherwise the soda4LCA system will complain that the dataset version already exists within the database. The option to overwrite the existing version while doing an import should be used carefully and can only be done if the original dataset is in the state UNRELEASED (i.e. not LOCKED or RELEASED). Versioning datasets when they are modified is a better approach because different versions can be compared afterwards. It is far easier to revert to a former version if modifications prove wrong when the previous version has not been overwritten.

Dataset versions which are quality controlled and have enough quality to be ready for publication can be switched to the state RELEASED by a dataset manager of the root data stock to which the dataset belongs. A released dataset can't be overwritten or checked out anymore for editing. Released datasets are foreseen to be used in LCA studies by third parties and therefore it's extremely important to keep them immutable. Otherwise, results of studies using released datasets cannot be reproduced by others when a dataset with a certain version number changes semantically between two usages. As can be seen in the next chapter released datasets are often grouped together to a set of harmonized "released datasets" which is a special version of a release data stock.

### **3. Non-root Data Stocks**

While each version of a logical dataset belong to a certain root data stock for creation, changing and administration of the datasets, further (non-root) data stocks can be created in a soda4LCA instance to build arbitrary logical collections of datasets. As described earlier, a soda4LCA instance can host two different root data stocks (see Figure 2 and Figure 4): one with LCI datasets related to products made out of plastics by a plastic working group and the other with LCI datasets related to metal products managed by the metals working group. Then, a dataset manager which has data management rights on both root data stocks could create a new non-root data stock which logically groups certain released versions of datasets from each root data stock together to form a LCI materials data collection which could then be made available to the public for general usage (Figure 4).

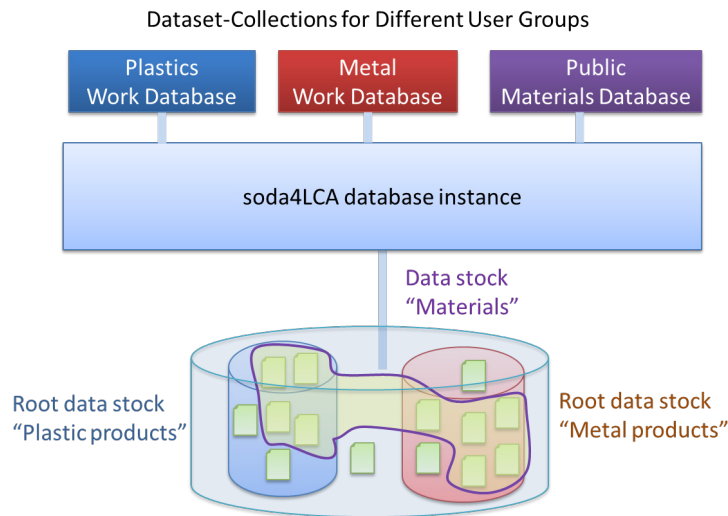


Figure 4

Soda4LCA instance containing two root data stocks and one non-root data stock

Typically, the public materials dataset collection will contain only versions of datasets which are in the state RELEASED. That means work on these dataset versions has been finished, the quality of the datasets was checked and the data management team is sure that these dataset versions have release quality. Moreover, all the specific versions of datasets contained in the materials collection should have been proved to form a harmonized set of datasets which could be used together in related LCA studies. A data stock with these characteristics is called release data stock.

A *release data stock* contains only released versions of datasets and at most one version of a certain dataset. Data stocks may be given the state "released" by a datasets manager if he is sure that he wants to release a data stock for general usage by the public or a group of authorized users. Soda4LCA will then check that the data stock only contains at most one released version of each dataset belonging to the release data stock. In the future it is foreseen that a set of conformity and quality rules, which datasets in a data stock must fulfil, and quality control mechanisms will be implemented which allow for proving these rules semi-automatically. A data stock can then only be switched to the state released if the quality mechanism has proved that the data stock fulfils the conformity and quality rules.

A release data stock may, but does not necessarily need to contain datasets from different root data stocks. I.e. in the case of the European Life Cycle (ELCD) database or the database with biomass energy datasets of the German BioEnergieDat project, the soda4LCA instances only contain one root data stock which contains all datasets as work database. For the BioEnergieDat project, this root data stock is called "BioEnergieDat" in Figure 5.

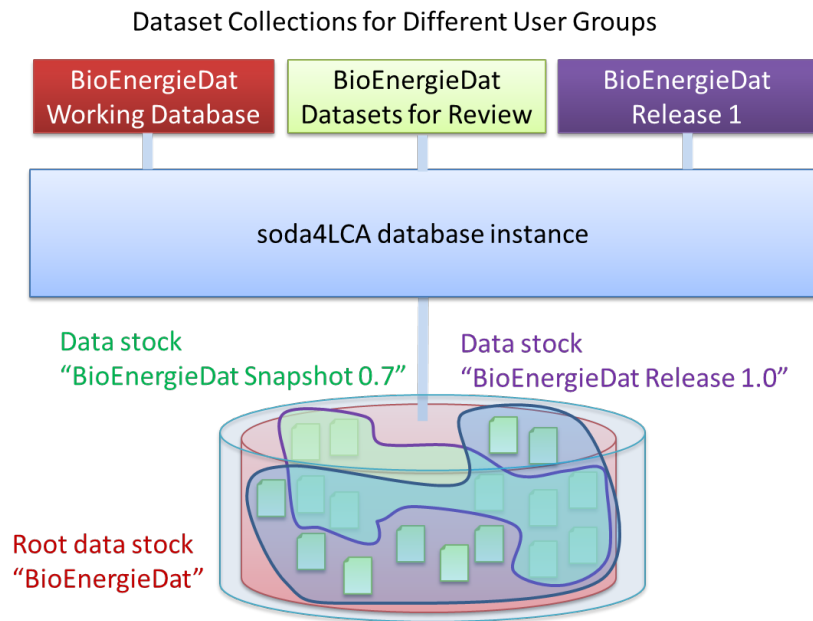


Figure 5

Soda4LCA for the BioEnergieDat project containing a work database, a data stock for review and a release data stock

The compilation of the whole BioEnergieDat database is carried out in several time periods in which different sub-packages of datasets are first prepared, then reviewed by independent experts and then finalized according to the input of the experts. To support the review, these sub-packages of datasets were grouped into pre-release data stocks for review (i.e. data stock "BioEnergieDat Snapshot 0.7"). Such a snapshot does not contain every dataset within the database, but only dataset versions which are to be reviewed and are in a good enough quality for such a review. Datasets which are not yet complete or whose quality is not good enough at the time of the review, or datasets which don't belong to the sub-package of datasets which should be reviewed by the experts for this review, are not assigned to the review data stock.

The authorization of users and groups on data stocks can now be set up so that the reviewers have full read permissions on all datasets contained in the review data stock but won't have rights to see any other datasets in the root data stock. They won't even see that there is a root data stock as work database beside the review data stock. After performing changes which were recommended by the reviewers to the datasets of data stock "BioEnergieDat Snapshot 0.7", another data stock "BioEnergieDat Snapshot 0.7.1" can be created which contains these changes.

Figure 6 shows the web user interface of the BioEnergieDat soda4LCA instance as seen by an administrator which already logged into the system. As shown at the top right the database contains three data stocks with names default (the root data stock), Release1 and Review according to the scenario just described.

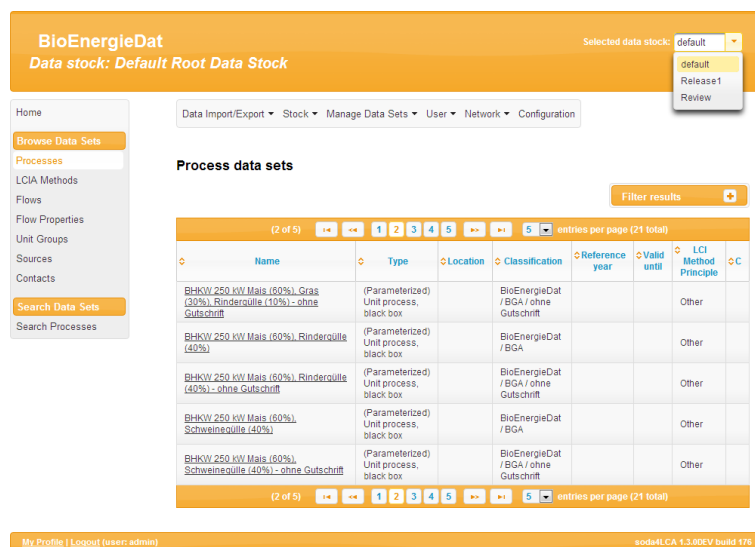


Figure 6  
BioEnergieDat database user interface showing available data stocks for the admin user

#### 4. Summary and Outlook

As described in the preceding chapters, the current development version of soda4LCA already contains the concept of data stocks for grouping datasets together to implement different usage scenarios. The concept of root data stocks allows an administrator to configure a soda4LCA instance as a service oriented database application which contains more than one work database for different work groups. Each such root data stock can be assigned different access rights for users and groups. If desired by the different working groups, root data stocks can be assigned access permissions which completely separate the access to the datasets for the different groups. In future versions of soda4LCA, this functionality will be enhanced further towards complete multitenancy support.

The concept of non-root data stocks allows dataset managers to build different collections of datasets for several distinct use cases. Release data stocks can be used to provide collections of harmonized and quality controlled datasets to users for doing LCA studies. Such release data stocks can be formed out of the datasets of several distinct root data stocks managed by different working groups. In the future it will even be possible to combine datasets from several release data stocks on different soda4LCA server instances to form one logical release of datasets (Kusche 2012).

Another usage scenario is to group a certain collection of datasets together for an internal or external review and for internal or external release management. Preview and review data stocks will in general not be public and will also contain datasets which are not ready for release. But they document the state of certain datasets at a given point of time. Release data stocks mark dataset collections which are available for general usage by others.

To help dataset creators and reviewers to compare different versions of datasets, a future version of soda4LCA could contain functionality to visualize the differences between dataset versions as is common nowadays in software version control systems. Also functionality to define sets of machine readable conformity and quality control rules can be implemented in soda4LCA, which can then be attached to data



stocks for semi-automatic quality control and conformity checking. This would help dataset managers and reviewers to determine if a set of datasets assigned to a data stock has release quality and is conform to conformity rules like the ILCD network conformity rules.

## **Bibliography**

- Ciroth, A.; Srocka, M.; Döpmeier, C.; Kusche, O.; Schebek, L. (2011b) - Interacting with the ILCD data network from an LCA software - the example of openLCA. Life Cycle Management Conf.2011 (LCM 2011), Berlin, August 28-31, 2011
- Döpmeier, C.; Kusche, O. (2011a): Concept and implementation of a service API for the remote access to life cycle data assessment databases. Pillmann, W. [Hrsg.] Innovations in Sharing Environmental Observations and Information: EnviroInfo 2011 - 25th Internat.Conf.on Environmental Informatics, Ispra, I, October 5-7, 2011 Aachen: Shaker Verl., 2011 Part 2 S.729-736 ISBN 978-3-8440-0451-9
- Ecoinvent Centre (2011): Ecoinvent LCI Database, Version 2.2, <http://www.ecoinvent.org/database/>
- European Commission (2011a): Communication from the Commission to the European Parliament, The Council, The European Economic and Social Committee and the Committee of the Regions: Europe 2020 Flagship Initiative - A Resource-Efficient Europe.
- European Commission - Joint Research Centre (JRC) - Institute for Environment and Sustainability, (2011b): European Platform on LCA. <http://lct.jrc.ec.europa.eu/assessment/projects>
- European Commission (2009): EU, Directive of the European Parliament and of the Council establishing a framework for the setting of Ecodesign requirements for energy-related products.
- European Commission (2003): Communication from the Commission to the Council and the European Parliament, Integrated Product Policy: Building on Environmental Life-Cycle Thinking (COM (2003)302).
- Kusche, O.; Döpmeier, C. (2012): Creating LCA Data Networks with soda4LCA. Pillmann, EnviroInfo 2012: 26th Internat.Conf.on Informatics for Environmental Protection, Dessau, D, October 29-31, 2012 Dessau
- OpenLCA (2011): OpenLCA project homepage, <http://www.openlca.org>
- Schebek, L., Ciroth, A.; Döpmeier, C.; Eltrop, L.; Simon, S.; Viebahn, P.; Wagner, H.-J.; Zschunke, T. (2010): Life Cycle Inventories for bioenergy: A German database for policy support. EcoBalance 2010: Towards & Beyond 2020; Tokio, 2010
- soda4LCA Open Source Project (2011): <http://www.iai.fzk.de/www-extern/index.php?id=2219&L=>
- Wolf, M.A.; Chomkamsri, K.; Döpmeier, C.; Kusche, O.; Pant, R.; Pennington, D. (2010): International reference life cycle data system (ILCD) data network. International Conference on EcoBalance, Tokyo, Japan, November 9-12, 2010