Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

ALIGNED D1.2 Description of scientific methods (Task 1.2 Framework for foreground life cycle inventory of bio-based sectors - LCI data template)

Pizzol, Massimo; Ghose, Agneta; Lancz, Kíra

Link to publication from Aalborg University

Work Package 1 – Shared modelling framework and learnings

D1.2 – Description of scientific methods
Task 1.2 – Framework for foreground life cycle inventory of bio-based sectors

Guide for structuring and sharing LCI data

Lead Contractor: Aalborg University (AAU)
Author(s): M. Pizzol, A. Ghose, K. Lancz

This document is a part of the ALIGNED project (grant no. 101059430) deliverable D1.2. It contains a description of how to share foreground inventory data within the ALIGNED project.

TABLE OF CONTENTS

List of tables

List of figures

Acronyms and abbreviations

ABBREVIATIONS	Description
WP	Work Package
LCA	Life cycle assessment
LCI	Life cycle inventory
UUID	Universally Unique IDentifiers

Executive summary

Step-by-step procedure to compile LCI inventory documentation in ALIGNED.

1) Draw a flow-diagram sketch or matrix of the foreground product system to identify the number of new and existing exchanges.

2) For each new exchange create universally unique identifiers (UUID). For each existing exchange find the UUID from the corresponding background database.

3) Compile the inventory using the wide-data table format and filling as many fields as possible. Remember to add notes.

4) Compile metadata using the metadata template.

1. The need for universally unique identifiers

Before talking about how to structure Life Cycle Inventory (LCI) data it is important to specify why we need universally unique identifiers (UUID) for every exchange in such inventory (https://en.wikipedia.org/wiki/Universally_unique_identifier).

Without identifying univocally each exchange in the dataset, it is not possible to:

- understand how activities are linked with each other, i.e. what is the structure of the product system model. How do we know that activity A requires an input from activity B if we don't know what exactly is exchanged between A and B?
- even more critical, understand to what specific concept each exchange refers to. There are many possible types of « electricity » (high- vs low-voltage, and country-specific) and many possible types of « carbon dioxide » (fossil, non fossil, biogenic, etc.) so how do we know which one should be included in the inventory without specifying exactly which one are we referring to?

These points are essential to properly communicate and reproduce the LCA model, for example in existing LCA software. Therefore, the very first step in this tutorial for reporting LCI data is about how to uniquely identify each exchange.

Suggested procedure to define UUIDs for each activity.

- For each **new foreground exchange** (both product and environmental exchange) in your product system, create a UUID using a generator, for example this one: https://www.uuidgenerator.net/ We recommend version 4 UUID. If the foreground product system includes four activities, it will require four UUIDs. See figure below.

- For each **existing background exchange** (both product and environmental exchange) in your product system, choose from a list of predefined UUIDs. Since existing UUIDs are database-specific, database version and the approach (consequential or attributional, for instance) should also be described to ensure the reproducibility of the LCA model.

In Figure 1. Two foreground activities, one linked to a background activity. UUIDs are provided for all exchanges. Foreground exchanges are identified with new, user-defined UUIDs, while background exchanges are identified with existing UUIDs
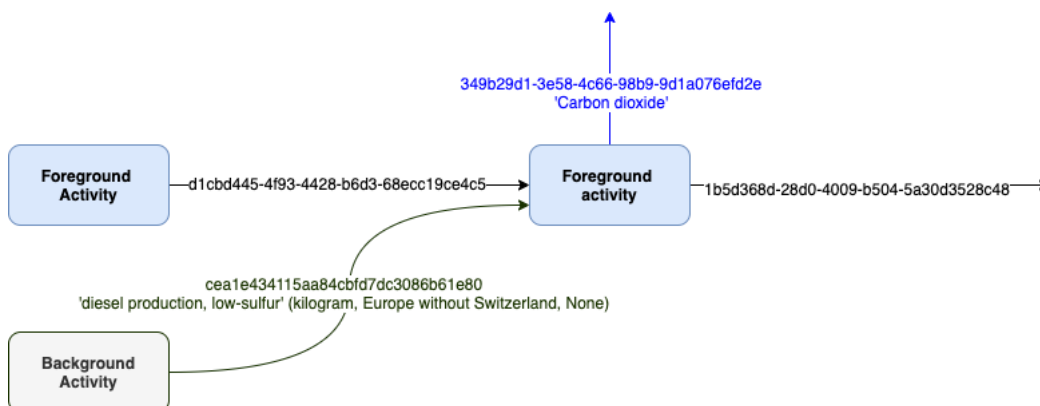
Figure 1 Exchanges between foreground and background activities and their UUIDs

Note: shouldn't databases also have UUIDs? In theory, yes. In practice, unfortunately no database provider has yet provided these to the best of our knowledge. The ideal solution would be to generate UUIDs for each new foreground database. To keep this manageable, however, we recommend using readable names that are sufficiently unique, e.g. "ALIGNED-WP5-LCI-database" for new datasets, while using the suggested database name for existing background datasets.

Note: on using exisiting background exchanges. Depending on the database version used and system model of this database (e.g. ecoinvent version 3.9.1 and consequential system model), the UUIDS for each exchange will be different. For, example diesel production low-sulfur in Europe has a different UUID in ecoinvent v.3.9.1 cutoff system model compared to the UUID shown in Figure 1 which is from the consequential system model of ecoinvent v.3.9.1. Also, the UUIDs are generated using the brightway2 software and are different form other UUIDs that ecoinvent might have provided for these datasets or from UUIDs used by another LCA software such as SimaPro or OpenLCA. So, UUIDs generated for existing background exchanges are database-specific.

2. General approach and terminology

The LCI template table includes information about <u>exchanges</u>. It is important to highlight the perhaps counterintuitive notion that exchanges are the primary focus, not activities as such. Activities should be literally intended as simply organized groups of exchanges that have one thing in common, they all belong to the same activity. Thus, the smallest unit in the template is the description of an exchange. Each exchange and its respective UUID belongs to a specific database. Exchanges can be grouped into activities. Exchanges can be of different types. Exchanges have units, values, and uncertainty.

Exchange: the smallest unit of analysis. Can be the link between two activities (product exchange), or between an activity and the system boundary (environmental exchange, although the term environmental is misleading, it could in principle also be an economic exchange or any other type). Can have uncertainty.

Database: where the exchange is located. A database is a structured collection of exchanges. It is structured because the exchanges can be used to describe a structure of interlinked activities. From the perspective of a user who intends to document a new LCA model we distinguish between foreground databases (created by the user) and background databases (created by others than the user).

Activity: an organized group of exchanges. It is organized because – from the perspective of the activity itself - some exchanges are considered inputs and some other outputs. One specific exchange is the reference flow or production exchange. We don't use the term "dataset" in this context as it could be misleading. In other context, a dataset could be referred to as the data concerning a single activity. However, in this context a "database" can contain one or more activities so the use of dataset is not recommended.

Note on system models: the LCI should be provided already organized into a system model. While some database providers provide separate versions of the data (unallocated and with different system models) the data here should be provided already organized into a system model – because each database is a model and making choices about the structure of the model is part of the LCI step. This means that a multifunctional system should be solved using either partition or substitution, and that only the LCI of the system with solved multifunctionality should be included. This in practice means either including inputs of negative sign when using the substitution method or reporting exclusively exchange amounts scaled after allocation when using the partitioning method. Additional information about the allocation procedure used and the system model can be reported using the "metadata" or should be reported in separate documentation.

3. Structuring the product system as a table in wide-data format

There is not a universal format to report LCI data. Some formats are software-specific, others are designed mainly for being handled by a computer and not a human. Any choice will be a compromise between its utility for various purposes and its user-friendly-ness. In the context of the ALIGNED project, we need a template for collecting and organizing LCI data from different case studies, by different users with a variable level of LCA expertise, and for easy review and use by internal project members and external LCA practitioners. Thus, we need a template that is designed for two purposes: allowing manual compilation of the inventory by the user (instead of automatic from a computer) and maximizing reproducibility. The template is based on the framework provided by the Brightway2 software for LCA as this is open source.

In ALIGNED, the LCI is reported as a table in wide-data format. A wide format is a format where *"unstacked data is presented with each different data variable in a separate column"* (https://en.wikipedia.org/wiki/Wide_and_narrow_data). There are both advantages and disadvantages in doing this.

Main advantages are that tabular data are simple and easy to inspect and manipulate with a variety of software, and several types of information can be organized in a table, which is useful in the case of LCA where each numerical data point is associated with a lot of contextual information. The main disadvantage is that of poor readability because some information needs to be repeated, so it's not compact. Also, this format can result in large files and is therefore not optimal at database scale. We consider it a sufficient compromise at foreground system scale (a foreground system might include activities in the order of the hundreds, not in the order of the thousands as a database), which is the main use-case of the ALIGNED project.

The structure of the wide-format table is presented in the following.

Table 1 List and description of items needed for LCI documentation

Item	Description
ROWS	
ε	Important, each row in the template refers to an **exchange**, either product exchange or environmental exchange. **This is hereby referred to as exchange "ε"**
COLUMNS	
Activity database	An identifier for the dataset. In principle, a UUID can be used. However, we recommend a readable but sufficiently unique name.
Activity code	The UUID of the production exchange of the foreground activity where ε is located. Or the UUID of the environmental exchange
Activity name	The name of the foreground activity where ε is located. While multiple activities can in principle have the same name, e.g. "electricity production" as produced by different technologies. We suggest being as specific as possible. Name should be readable and understandable.
Activity unit	The unit of the production exchange of the foreground activity where ε is located. Or the unit of the environmental exchange
Activity type	Two types: "process" for product exchanges (technology matrix) and "biosphere" for environmental exchanges (intervention matrix)
Exchange database	The identifier of the database where the exchange ε is located.
Exchange input	The UUID of ε. A computer generated UUIDs created by the user for foreground exchanges and taken from an existing list for background exchanges.
Exchange amount	The value of ε
Exchange unit	The unit of ε
Exchange type	"production" for the main output (reference flow, there must be one production exchange for each activity). "Technosphere" for product exchanges other than the main output (e.g. Inputs and co-product outputs). "Biosphere" for environmental exchanges.
Exchange uncertainty type	See the stats_arrays package documentation here: https://stats-arrays.readthedocs.io/en/latest/#id22
Exchange loc	See the stats_arrays package documentation here: https://stats-arrays.readthedocs.io/en/latest/#id22
Exchange scale	See the stats_arrays package documentation here: https://stats-arrays.readthedocs.io/en/latest/#id22
Exchange negative	"True" unless the "amount" field is negative then use "False"
Notes	Any additional note. Be as extensive as possible.
Other	It is possible to include additional columns if needed, the ones indicated above are the minimum recommended set. Other columns may include software or database-specific data points, for example columns for "SimaPro process name" and "SimaPro unit" if the database is meant for importing into a SimaPro project.

4. Metadata

Metadata refers to information about specific data or "data about data". In ALIGNED, the metadata needed are coherent with the data provided in the inventory documentation format, as well as with the current indications of the GLAD (Global CA Data Access) network and their framework (https://www.globallcadataaccess.org/). For each dataset provided (cf. previous sections) a metadata file needs to be provided as well. The following table lists the metadata required together with a description.

Table 2 List and description of items needed as metadata

Item	Description
ROWS	
	Important, each row in the template refers to an **activity** (referred in GLAD as process). Hereby referred to activity "α". The metadata file can thus describe multiple activities at once and should mirror the content of the LCI documentation file.
COLUMNS	
Unique identifier of dataset	This is the UUID of the activity (process/production process). The UUID can be generated from https://www.uuidgenerator.net/
Process name	The name of α. While multiple activities can in principle have the same name, e.g. "electricity production" as produced by different technologies. We suggest being as specific as possible. Name should be readable and understandable.
Sector(s)	The sector of the dataset Use comma to separate multiple values. e.g. Mining, Other
Process type	The type of process (unit, system, etc.)
Description	A narrative description of the data set
Valid from	The year of the start of the validity of the data set.
Regional Code	The geographical location of the data set (e.g. DE, GLO, ROW)
Data format	The data format the data set is available in (e.g. ECOSPOLD1, ECOSPOLD2, ILCD, .csv)
LCI modelling approach	The LCI modelling type (attributional, consequential, before modelling)
Method used to handle multifunctionality	Substitution OR Allocation
Available for free ?	Indicates if the data set is available for free
URL to dataset or database	A url to download the complete data set, or alternatively, the website of the data provider. With respect to the ALIGNED project this should be the link to the Zenodo repository.

5. Step-by-step procedure to compile LCI inventory

Templates in excel for both data and metadata are available together with this guide in the dedicated project repository.

Instructions on how to use the templates to compile and document an inventory are explained in the following section. An example of filled template is available in the project dedicated repository, together with an "explainer" guide that refers to the same example.

Code to convert from the template into a product system in the software brightway2 is available here in the dedicated project repository as well as in the GitHub repository: https://github.com/massimopizzol/aligned-converter

The procedure for compiling LCI data is in five steps and described in the following.

5.1. Draw a flow-diagram

Sketch of the foreground product system. For example, using drawing tools like draw.io or similar. Advanced users can prepare a matrix version of the product system. This will be the first step to identify the number of new and existing exchanges.
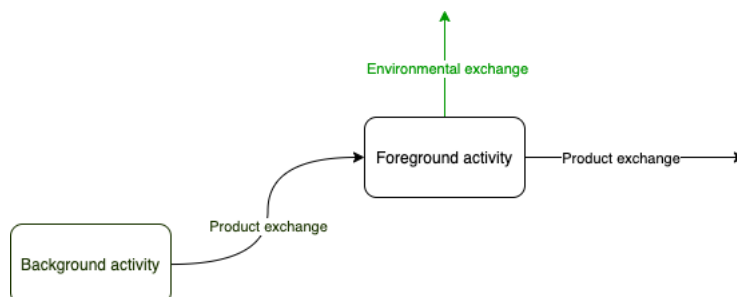

Figure 2 A product system with three exchanges.

5.2. Associate each exchange to a UUID.

Create new UUIDs for foreground exchanges. For each existing (background) exchange find the UUID from the corresponding database. SimaPro users can use the exact full name of the dataset or environmental exchange used, or any other ID if available in the documentation of this dataset. Brightway users can use the Brightway-generated UUID for the databases available (including biosphere).
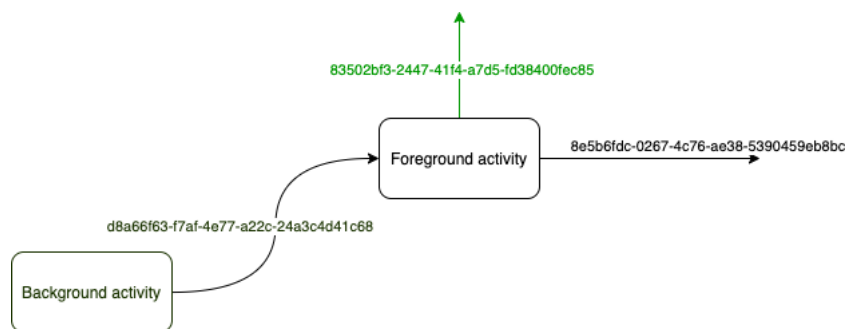
Figure 3 Adding UUIDs to all exchanges.

5.3. Compile the inventory table

Use the wide-data table format and fill in as many fields as possible.

Table 3 Inventory table compiled for the three exchanges.

Activity database	Activity code	Activity name	Activity unit	Activity type	Exchange database	Exchange input	Exchange amount	Exchange unit	Exchange type
foreground-db-name	8e5b6fdc-0267-4c76-ae38-5390459eb8bc	Foreground activity	some-unit	process	foreground-db-name	8e5b6fdc-0267-4c76-ae38-5390459eb8bc	1	some-unit	production
foreground-db-name	8e5b6fdc-0267-4c76-ae38-5390459eb8bc	Foreground activity	some-unit	process	background-db-name	d8a66f63-f7af-4e77-a22c-24a3c4d41c68	1	some-unit	technosphere
foreground-db-name	8e5b6fdc-0267-4c76-ae38-5390459eb8bc	Foreground activity	some-unit	process	background-db-name	83502bf3-2447-41f4-a7d5-fd38400fec85	1	some-unit	biosphere

It is strongly recommended to add notes. Suggestion for names:

a. For the "Activity database" identifier you can use an UUID or a more readable name. In the latter case use hyphens and several words to univocally define the name. For example: "ALIGNED-WP5-Pulp-and-paper-LCI-2023"

b. For the "Exchange database" we recommend names that identify as univocally as possible the database. For example: "ecoinvent v3.9 conseq".

c. To identify different alterative scenarios we recommend either creating multiple databases and thus having multiple "Activity database" identifiers (e.g. "ALIGNED-WP5-scenario1" and "ALIGNED-WP5-scenario2") or simply keeping all scenarios in the same database and specifying the scenario name either in the "activity name", e.g. "Biomass-processing-scenario-1" or in a new column.

5.4. Compile metadata

Use the GLAD metadata template.

5.5. Create a datapackage

At this point, it is recommended to create a 'datapackage' with the LCI data. In simple words this means to upload the data in a repository, ideally openly accessible, and organize the files in a specific structure so that it is clear what is data and what is metadata. Instructions for doing this manually or automatically in python are provided at the GitHub repository: https://github.com/massimopizzol/aligned-datapackage