The openLCA format converter – new release May 2013

Andreas Ciroth, Imo Graf, Michael Srocka, GreenDelta GmbH, May 16 2013
gd@greendelta.com

Content

1　The format converter – introduction & motivation

The format converter is created and maintained by GreenDelta since 2006. Motivation for creating it was already described in earlier documentation [1]:

"The converter aims to provide a surrogate for a format consensus. It provides the technical means for loss-less data exchange between the most important LCA data formats today, and thus enables users to switch between them as they need. In that sense, it resembles very much an electrical adaptor. […] it might even be a more convenient solution since it allows using different formats in parallel."

The format converter was the first released element in the openLCA project, [2]. Since its beginning, it was released as free, open source software, under the Mozilla Public Licence. The latest version is available from the openLCA website [3]. Up to now, there are four releases:

Home / openlca_converter	
Name ⬍	**Modified** ⬍
⬆ **Parent folder**	
📁 **openlca_converter_2.1**	2010-09-30
📁 **openlca_converter_2.0**	2010-06-04
📁 **openlca_converter_1.1**	2007-08-15
📁 **converter1.0**	2007-05-04

Figure 1: Previous releases of the format converter, with release date

This text is about the release for a format converter version 3.0, which is then the fifth release.

2　Release March 2013 – what's new

The new release contains the following main changes and updated:

- new formats have been added: SimaPro CSV and SimaPro EcoSpold (i.e. the EcoSpold 1 'dialect' that is used by the LCA Software SimaPro);

- for the ILCD format, the new flows from the ELCD III database have been taken into account;

- the user interface is updated and extended.

We are grateful for support that we received from PRé Consultants (www.pre-sustainability.com) for this converter release.

3 A quick documentation for the new converter

3.1 Converter requirements

The format converter has only moderate hardware requirements that should be met by all modern computers. The software itself requires less than 20 MB of hard disc space. If larger amounts of data need to be converted in one go, the conversion is faster if more RAM is available.

Since the openLCA converter is a Java software, it runs under Windows, Linux, and MAC OS. A Java version of 1.6 or higher is required. If you are unsure about whether Java is installed on your computer, you can follow these steps to find out: www.openlca.org/documentation/index.php/Verify_Java_version.

3.2 Converter features

The converter is **a tool for converting LCA data sets from one LCA data format to another**.

It currently covers the following data formats:

- **EcoSpold 1**: data format released by the ecoinvent centre, Switzerland [4]

- **EcoSpold 1 SimaPro** (new): The SimaPro LCA software uses a slightly different EcoSpold format than the ecoinvent centre; it has now been integrated in the format converter as well

- **EcoSpold 2**: data format released also by the ecoinvent centre, Switzerland [5]; designed to be a successor of EcoSpoldy 1, will be broadly available with the upcoming ecoinvent 3 database.

- **ILCD**: Data format released by the Joint Research Centre (JRC) of the European Commission, in version 1.1 [6]

- **CSV SimaPro** (new): A data format used by the SimaPro software for import and export; in contrast to EcoSpold 1, it supports parameters.

Not all data formats are available for all conversion directions; Table 1 gives an overview.

Table 1: Supported LCA data formats in the openLCA converter, version 3

from/to	EcoSpold 1	EcoSpold 1 (SimaPro)	EcoSpold 2	ILCD 1.1	CSV (SimaPro)
EcoSpold 1		x	x	x	-
EcoSpold 1 (SimaPro)	x		x	x	-
EcoSpold 2	x	x		x	x
ILCD 1.1	x	x	x		-
CSV (SimaPro)	-	-	x	-	

In a conversion, the format converter applies a **mapping** of elements in one format and format reference to elements in the other format. For example, flow nr. 188,"Carbon dioxide, biogenic" in a data set in EcoSpold 1 format becomes, when this data set is converted into ILCD, a flow with UUID 08a91e70-3ddc-11dd-9241-0050c2490048,"carbon dioxide (biogenic)". As a result of the mapping, the converted data set integrates much better with original data sets from the target format. This is especially important for the application of Life Cycle Impact Assessment methods.

Finally, the converter offers various **tools for analysing the conversion results**. These will be explained later more in detail.

3.3 Using the converter

3.3.1 Overview of conversion steps

Using the format converter is straightforward:

0. Download the converter, unzip the archive, and start the converter application, in Windows by double-clicking on the converter jar file.
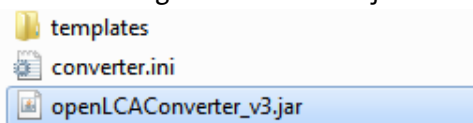
   Download and extraction of the downloaded zip archive needs to be done only once. As soon as the converter is open, follow these principal steps (see also Figure 2):

1. Select a source file.
2. Select a target directory.
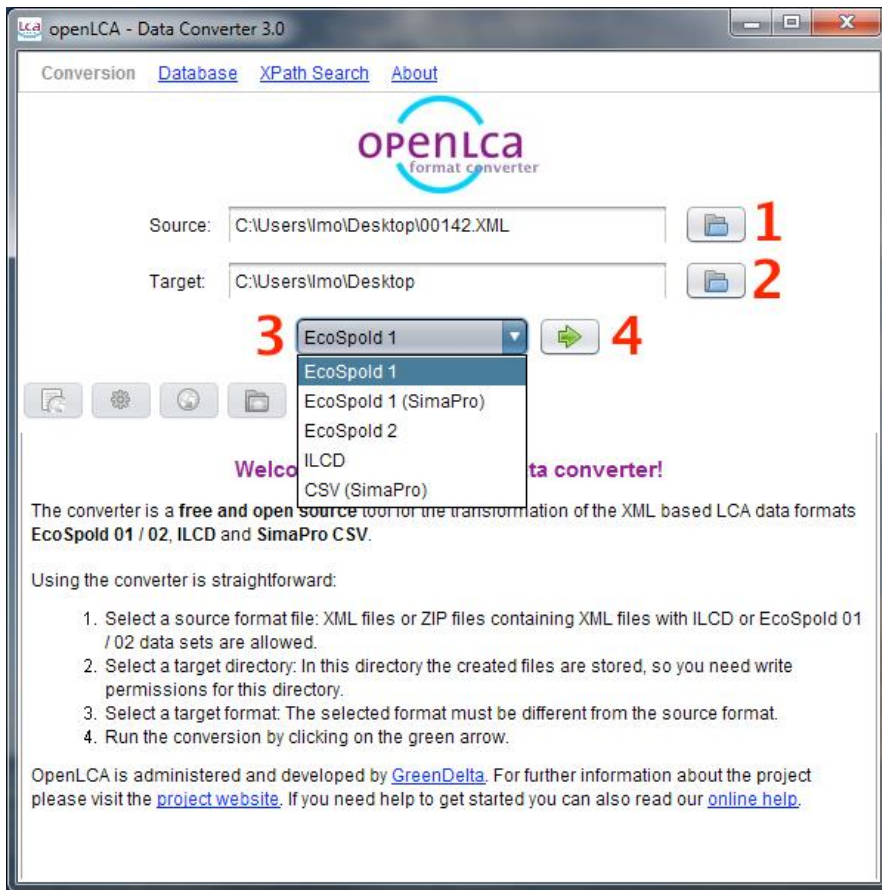3. Select a target format.
4. Run the conversion.

Figure 2: Four main steps of data conversion in the format converter.

When the conversion is finished, you should inspect the results.

3.3.2 Converting data sets

The steps for converting data sets have already been introduced in the previous section.

1) Select the source file

In the first step, the source file needs to be selected. To do so, click on the folder symbol in the source line of the converter:


Depending on the data format, the file ending can vary (zip, csv, xml).

- for EcoSpold 1, the file extension needs to be XML or zip (an archive of EcoSpold XML files)

- for EcoSpold 2, the file extension needs to be spold or zip (an archive of EcoSpold 2 spold files)
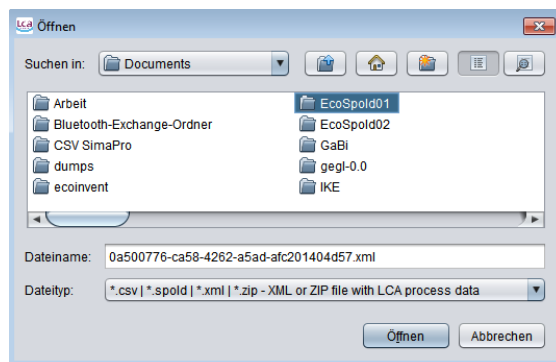
- For ILCD, the file extension needs to be zip (an archive of ILCD 1.1. XML files, in the characteristic ILCD folder structure)

- For csv, the file extension is (obviously) csv.

You can convert one single or also multiple process data sets in one go – but you can select only one file as source file. Therefore, for converting multiple data sets, you can combine them in one zip archive, or, for EcoSpold 1, also append all in one XML file.

The source format is automatically detected. If no suiteable format is found, the conversion will not start, and an error message will be thrown instead.
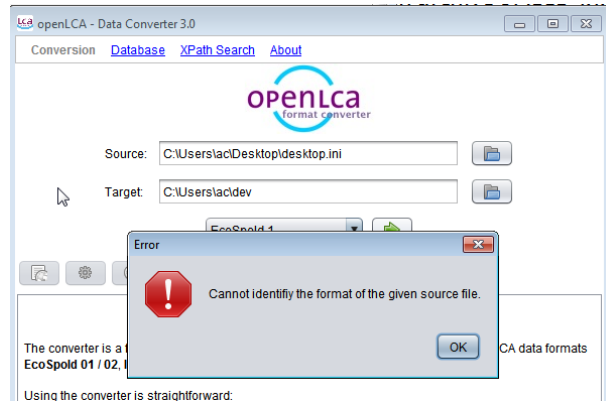
2) Select the target directory

As a next step, the target location needs to be selected. The converted file(s) will be created in the selected folder, with the name of the target format (e.g, ILCD, or EcoSpold_01). To select, click on the folder symbol in the 'target' line.

3) Select the target format

Next, the target format needs to be selected, via the drop-down menu. The target format must b different from the source format, otherwise an error is thrown.
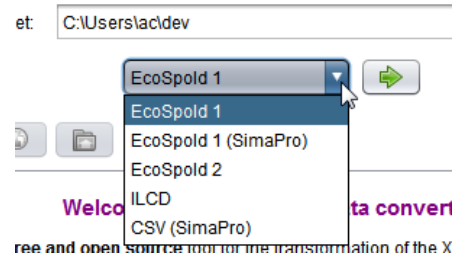
4) Start the conversion

Pressing the green arrow (right near the target format, 4 in Figure 2) starts the conversion.

3.3.3 Converting data sets – specific settings for EcoSpold 2 to SimaPro CSV conversion

For converting Ecospold 2 to SimaPro CSV, additional settings are necessary.

The SimaPro CSV format is a format that is used by the SimaPro LCA software for data exchange. It contains more information than EcoSpold 1 for a process data set. For example, it includes parameters in the process. SimaPro CVS files are obtained as an export of SimaPro (Figure 3).
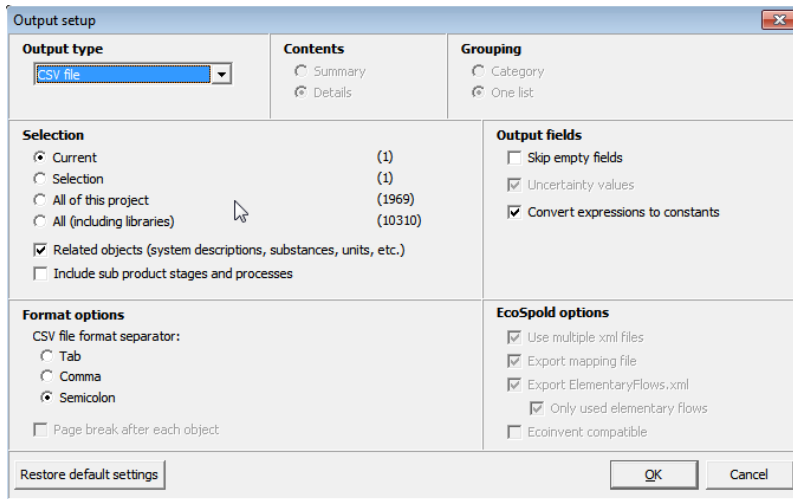
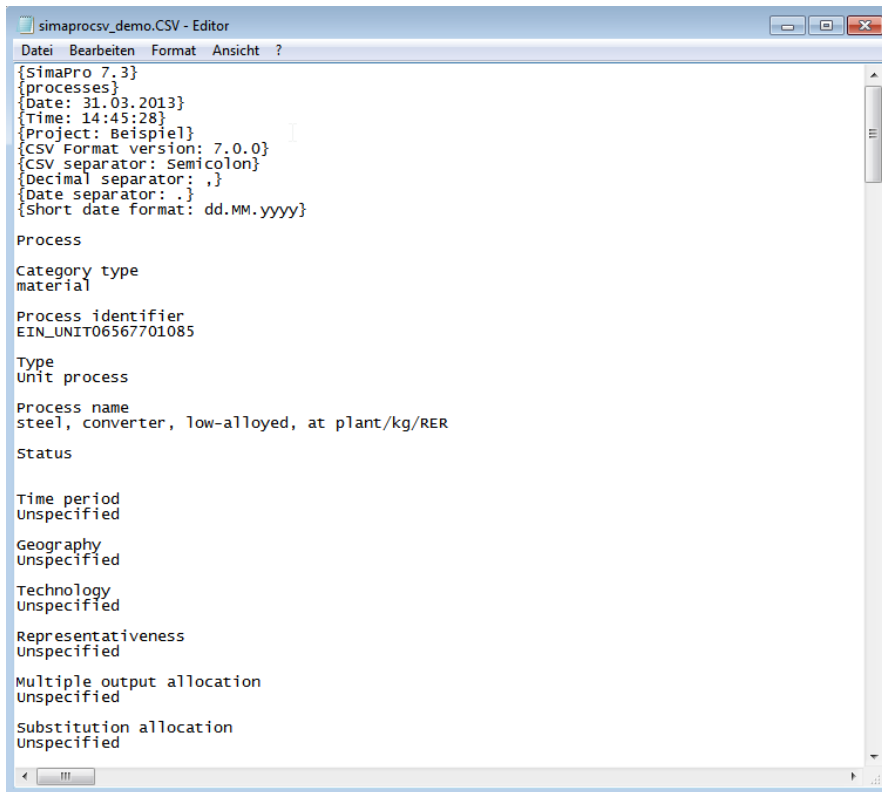Figure 3: Creating a SimaPro CSV file (SimaPro 7.3.3 screenshot).

Figure 4: An example SimaPro CSV file

If you select SimaPro CSV as target format, the user interface changes (Figure 5), and a new menu entry appears, 'Settings'.

Figure 5: Converter user interface (upper part) for converting to SimaPro CSV format

3.3.3.1   Preferred language

EcoSpold 2 files may contain text in different languages. In the CSV file, always one language is possible. Therefore, a selection of the preferred language is required prior to the conversion. This preferred language will then be entered into the csv file. The default language is English, it is also always provided in an EcoSpold 2 file. "Scan languages" will scan through the files that are selected as conversion source. After the scan is completed, all available languages can be selected in the "preferred language" combo box.
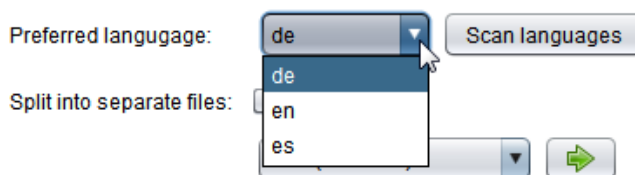
Figure 6: After the available languages have been checked in the source files, a preferred language can be selected

For a single data entry, the preferred language will be taken the first. If the preferred language is not available for the specific data entry, then the first available language entry is taken. This principle is also followed when the dataset are scanned in the settings, which will be explained later.

3.3.3.2   Split CSV files

Checking „Split into separate files" will create one csv file for each process data set; the name of the csv file is then the EcoSpold 2 field activityId. Otherwise, one csv file for all process data sets is created.

3.3.3.3   Additional settings

For converting to CSV format, further settings are available from the "Settings" menu entry: The product name can be "designed", mappings for geography and compartments can
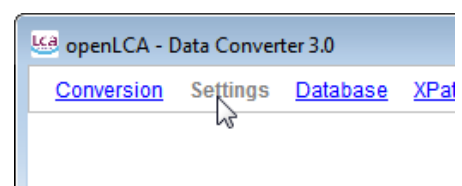
be managed, and units can be edited.

3.3.3.3.1 Product name

SimaPro uses the product name as a unique identifier for linking process data sets. The converter creates unique names as a combination of different attributes; the order of these attributes can be specified in the product name dialogue. To do so, just type into the "order" number entries the order of the respective EcoSpold 2 attributes. You can save your modification, and you can also revert back to the default order ("restore defaults" button near the save button).

Save the preferences

Set the preferences to default
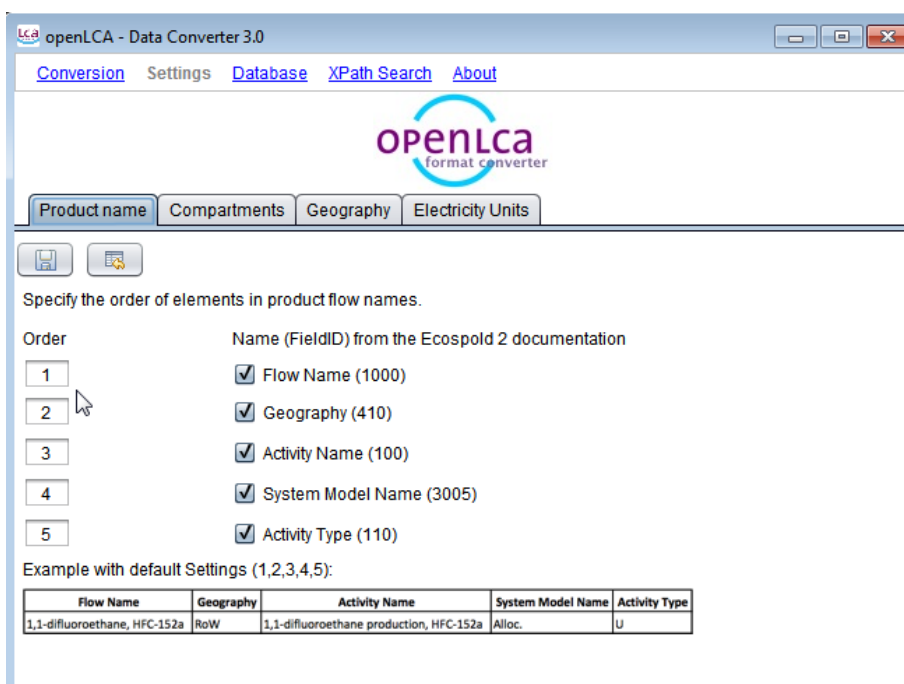
Figure 7: Settings for CSV format, product name, default
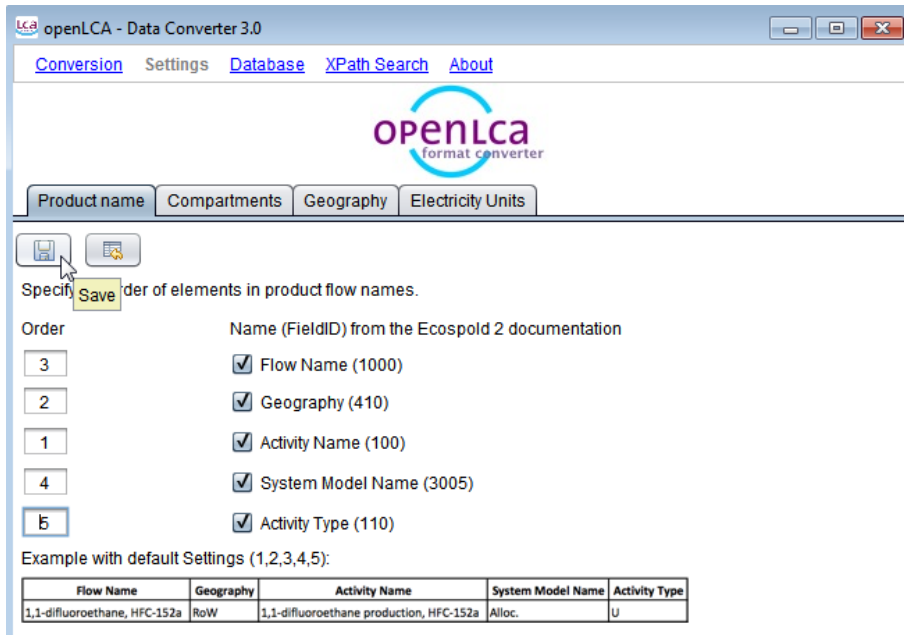
Figure 8: Settings for CSV format, product names, modification

Unchecking the checkbox under name will have the effect that the attribute is omitted in the product name. However, be aware that the product may not be unique any more if attributes are omitted.

3.3.3.3.2   Compartment and geography mapping icons

Compartment and geography mapping each contain following buttons:

	Scan the data in the source file
	Open the mapping file
	Save the mappings
	Add a new row to the list
	Remove the selected rows

3.3.3.3.3   Compartment Mapping

The mappings for compartments are used for the elementary flows. They map a compartment together with a subcompartment to a SimaPro elementary type. Open the provided mapping in the database (Figure 9) or scan the source files for compartments (Figure 10).
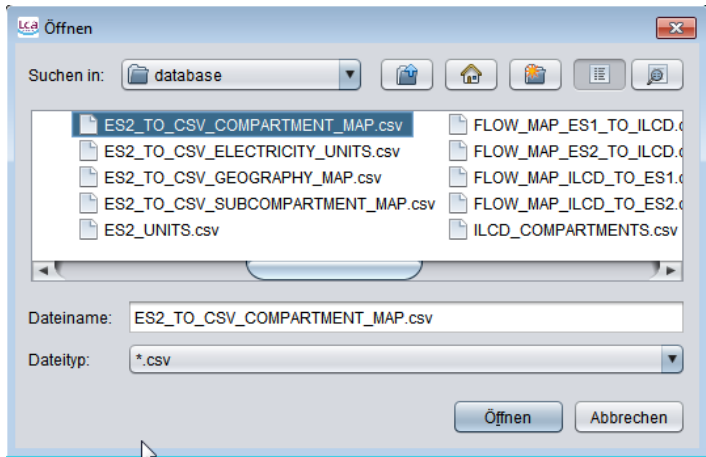
Figure 9: Opening the compartment mapping in the database
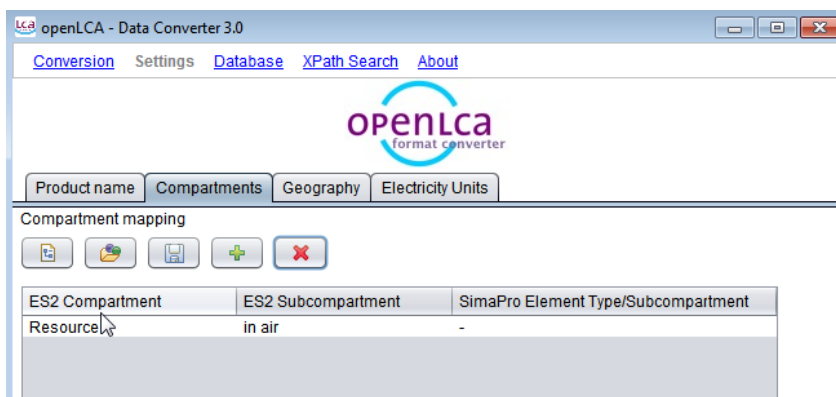
Figure 10: Result after scanning the compartments in the source data sets

You can then click on the "SimaPro Element" column on the right and assign the SimaPro compartments (Figure 11).

Figure 11: Assigning SimaPro Compartments to EcoSpold 2 compartments

When you are done, save.
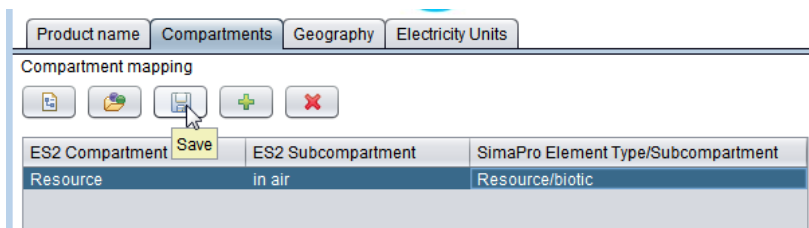
3.3.3.3.4   Geography mapping

Also for geography, SimaPro works with a fixed list of locations; values in the EcoSpold 2 field 'geography shortName (410) ' need to be assigned to this list. This is done in the geography mapping. The structure is very similar to the compartment mapping (Figure 12).
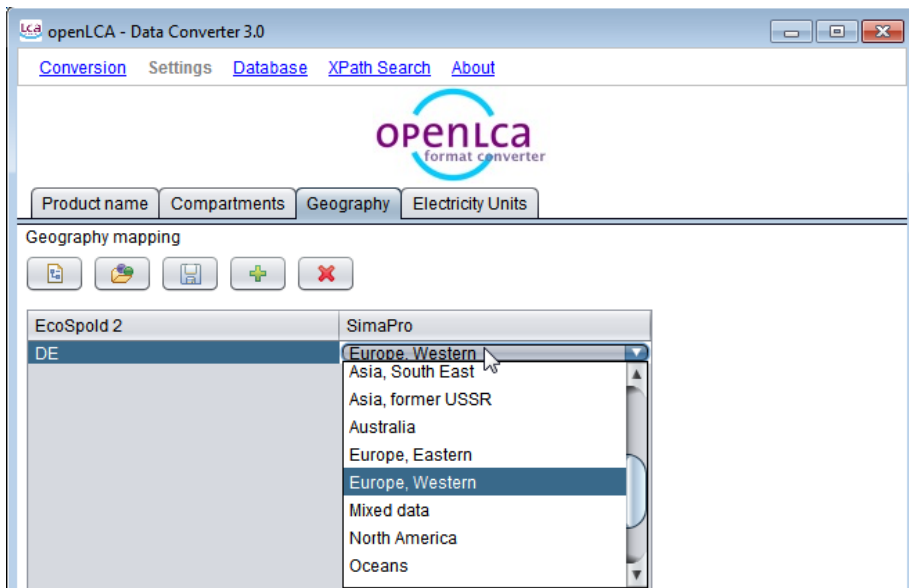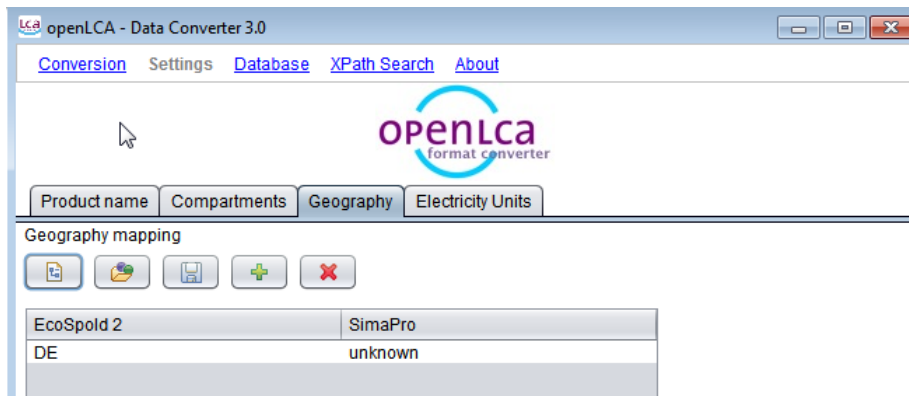
Figure 12: Assigning SimaPro locations to EcoSpold 2 locations

3.3.3.3.5   Electricity units

Electricity units are used for the conversion of intermediate exchanges. If in EcoSpold 2, the inputGroup (1500) is 5 and the EcoSpold activity is a process, then this exchange is normally written to the "materials/fuels" section in the SimaPro csv. However if the exchange has one of the units in the "electricity units" list, then it is written to the "electricity/heat" section.

To modify this list, simply click in the list on one line and then delete or add. Capital writing is ignored. If you have changed anything, don't forget to save.
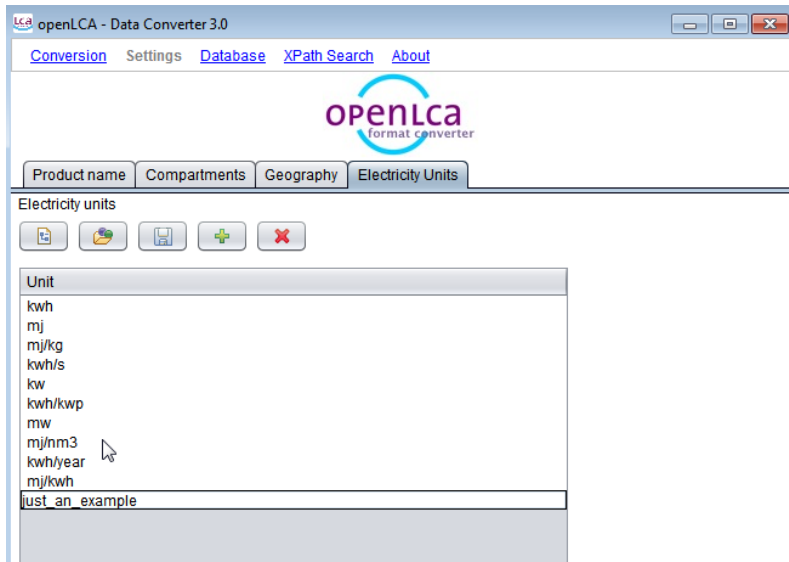
Figure 13: Changing the electricity units

Once you are done with the settings specification, you can (save and edits and) go back to the conversion, with a click on the conversion menu entry.

3.3.4   Conversion results

As soon as the conversion is finished, the converter shows the log file that contains links to the created files. For ILCD and EcoSpold 2 formats, more than one data set is created per input process data set, due to the additional flow data sets and so on. If errors occur during the conversion, a log file is created (Figure 14); a list of the created files is always shown (Figure 14, Figure 15).

Figure 14: Conversion overview, with errors.

Figure 15: Conversion overview, conversion without errors.

3.3.4.1   Inspecting conversion results

There are various ways to access the converted files.

If you click on one of the links in the conversion index, you can inspect the XML file that is created.
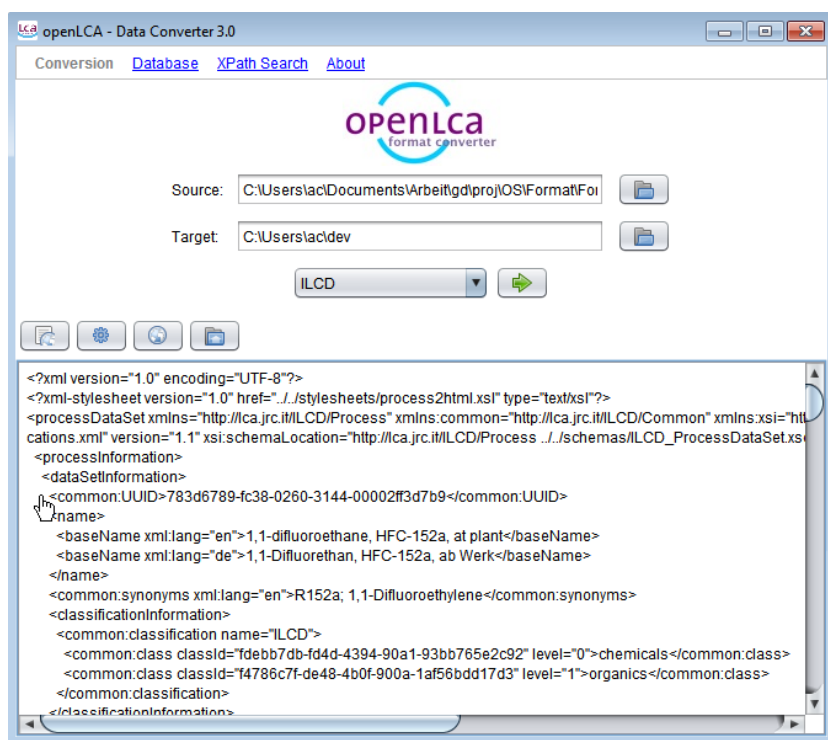
Figure 16: XML view of one created file

Clicking on the conversion index icon goes back to the conversion overview. As an alternative, you can open the conversion list also in a webbrowser by clicking on the 'open in browser' icon, or access it in the file explorer by clicking on the respective 'open in file explorer' icon.
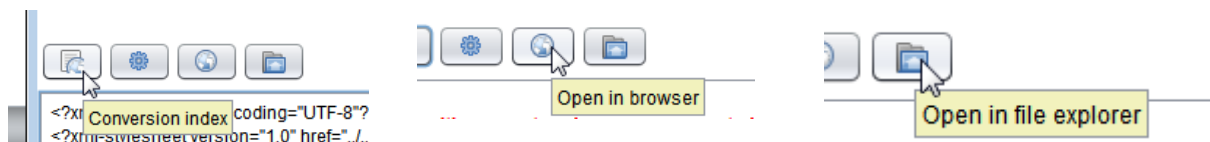
If the conversion index is open in a web browser, a click on one of the links in the index will open the respective file in the web browser. For those formats that provide stylesheets, the XML file is formatted according to the style sheet of the format. For example, for ILCD, you see the typical yellow / orange table structure (Figure 17, Figure 18). These formatted tables are usually much nicer to browse through, but they might not contain all information that is provided in the XML file.
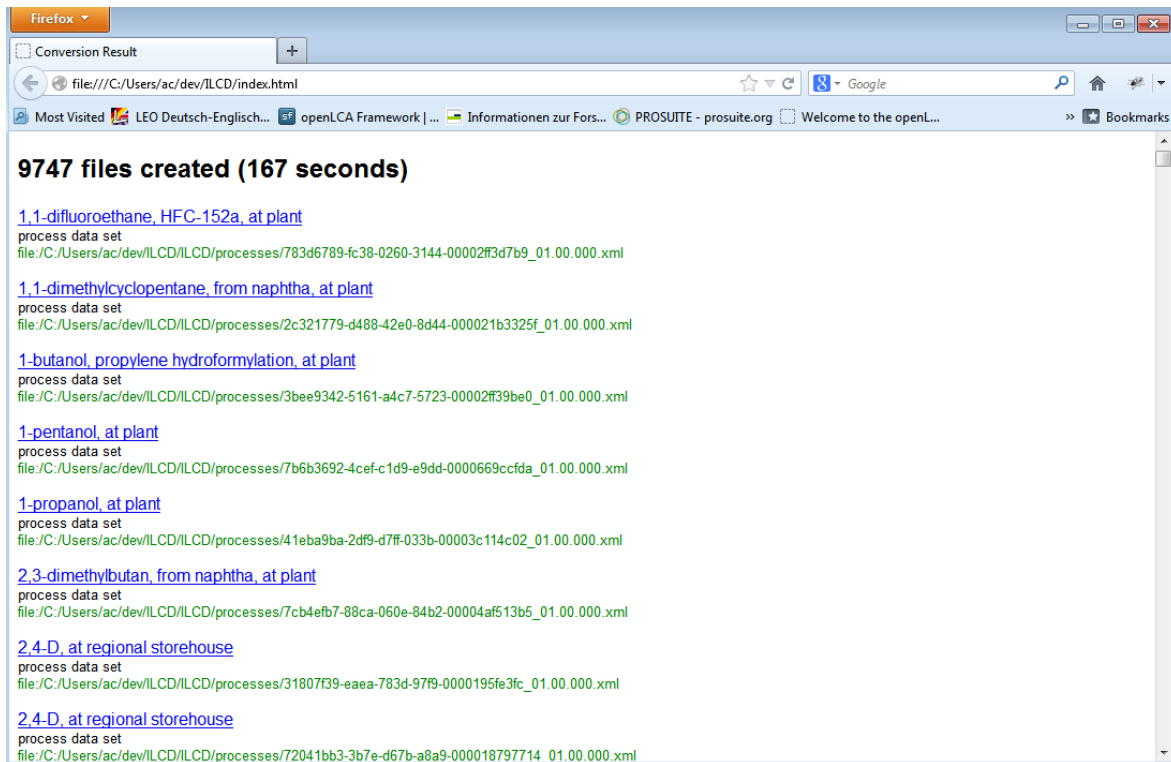
Figure 17: Conversion index in a web browser.
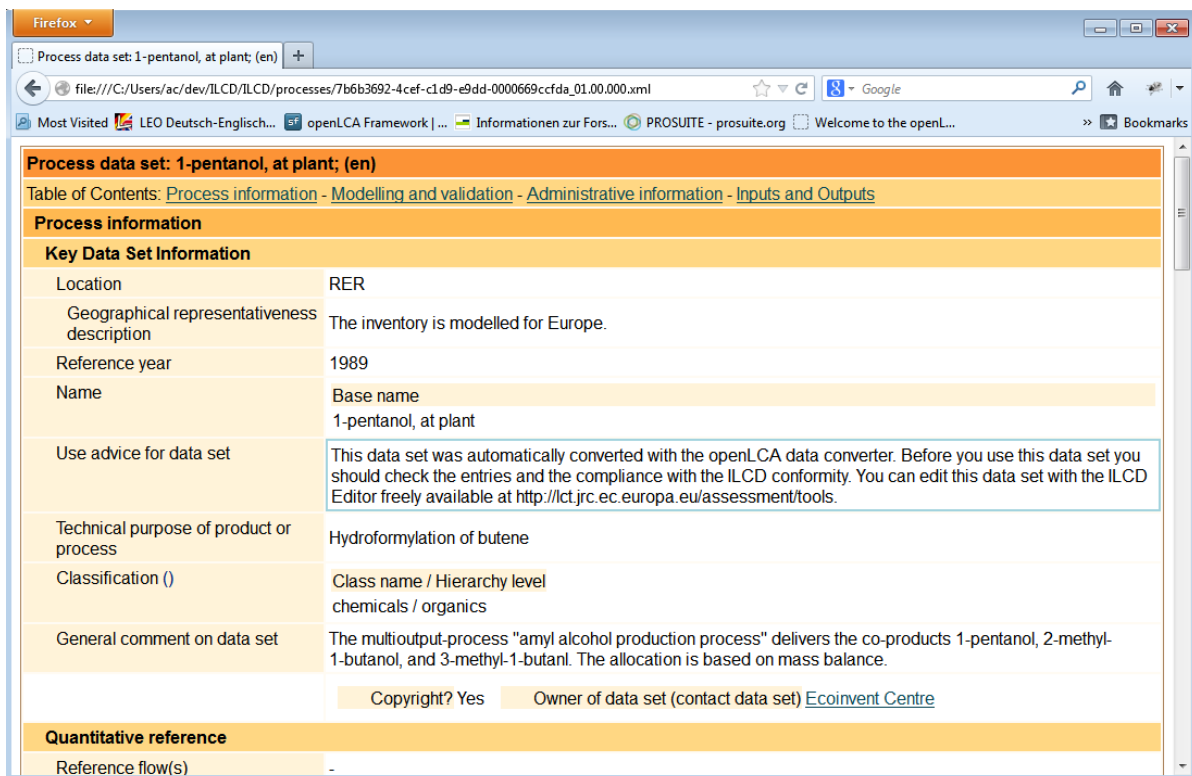
Figure 18: A data set converted to ILCD format, opened in a web browser.

Then, you can simply work with the files outside of the converter and outside of a web browser; in windows, 'open file in explorer' just opens the windows explorer.
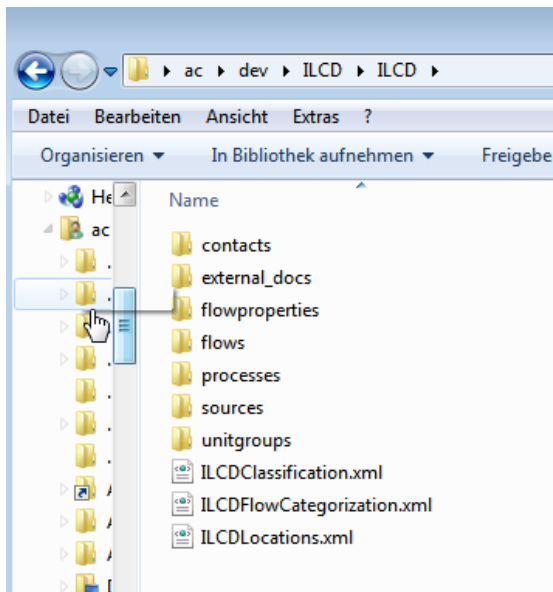
Figure 19: Converted files in windows explorer.

3.3.4.2   Validating conversion results

A click on the 'cog' icon will start a validation of the converted files.
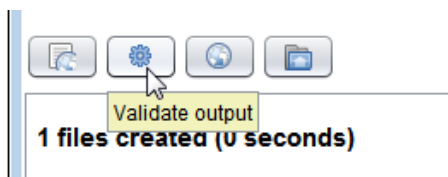
Depending on the amount of converted data, this might take a while.
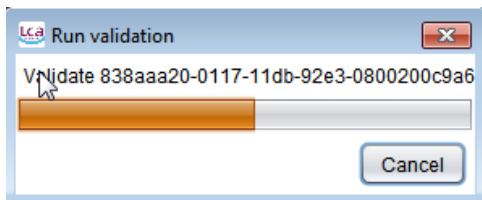
The validation usually does not pass without errors. However, often, the errors are minor, although every error is classified as "severe" in the validation output. Figure 20 shows an example. In this example, the CAS number (which is taken from the source file) is not correctly formatted, for several flows. In the next figure, a date format is not correctly provided (Figure 21).
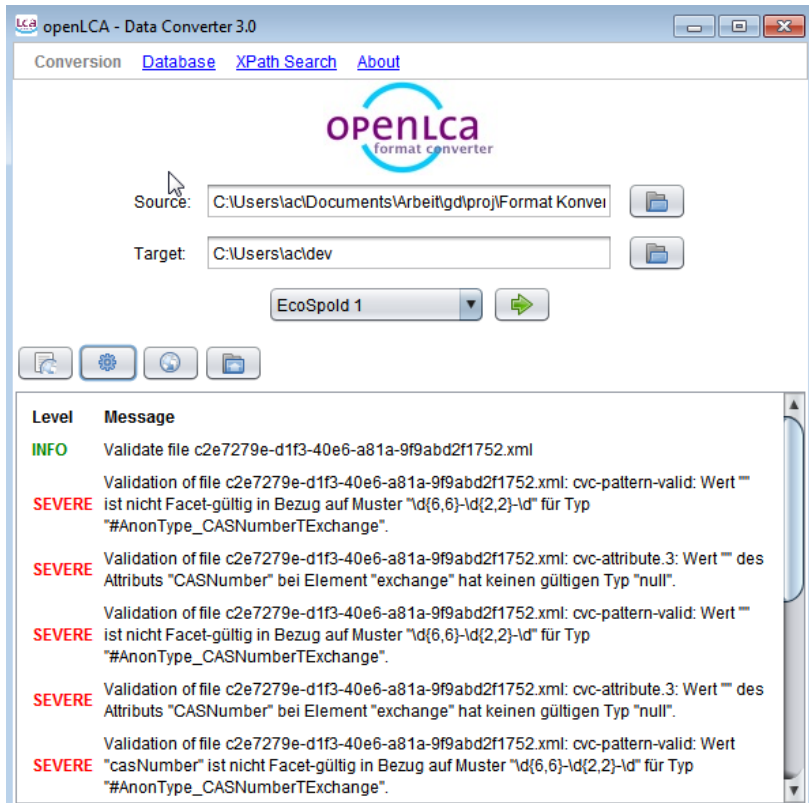
Figure 20: Validation output example: ill-formatted CAS number
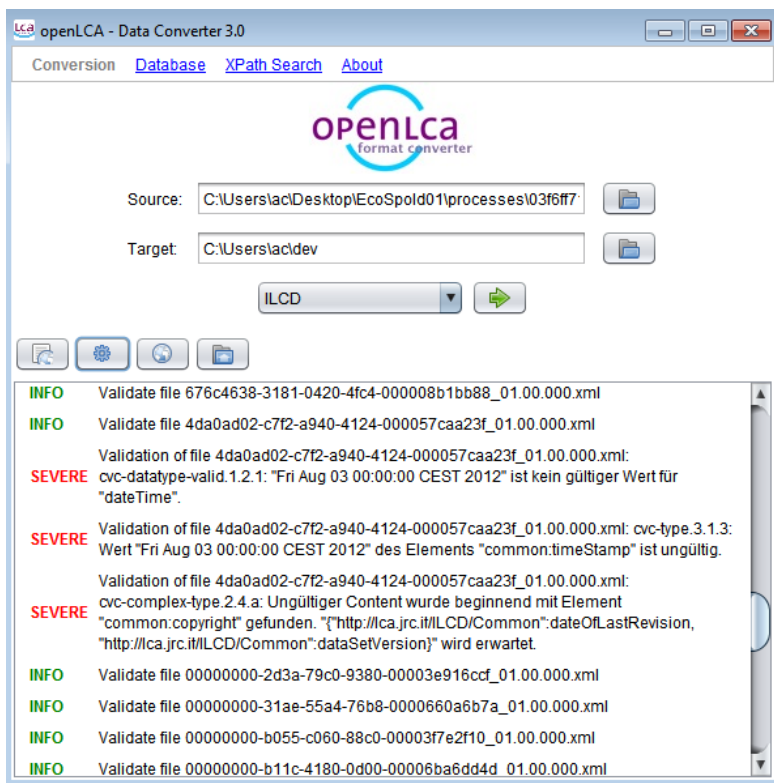
Figure 21: Validation output example: incorrect time format

Of course also more severe errors can happen. For example, the conversion from Figure 14 (conversion with errors), has the following error log (Figure 22).
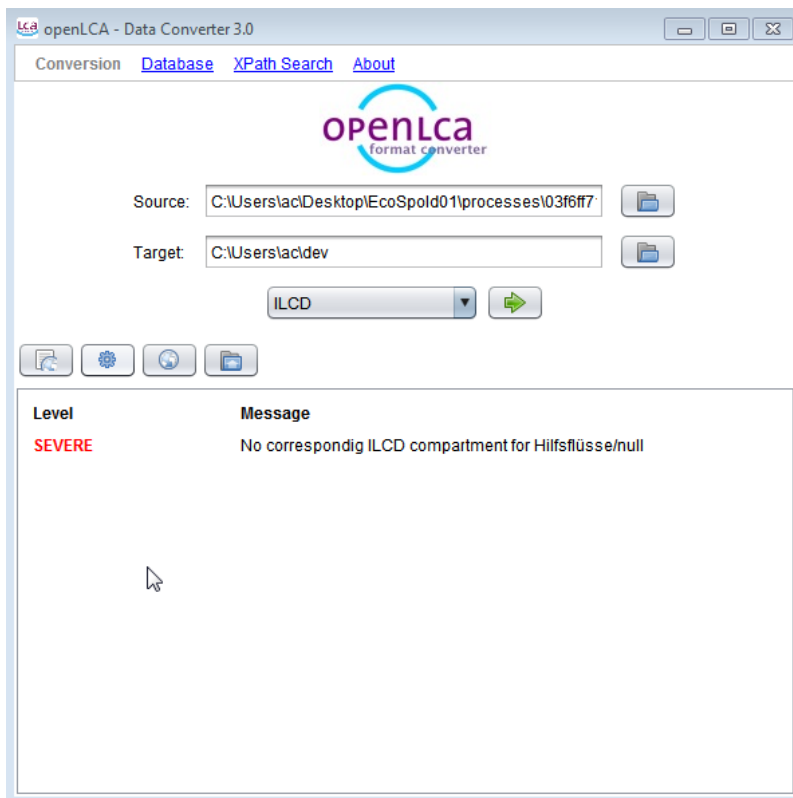
Figure 22: error log example

It shows that a subcompartment for elementary flows is missing (hence the entry: "/null"), and for the main compartment, no reference counterpart and no mapping is provided. This needs to be fixed before the conversion.

The validation errors can also be fixed in the source files, but are sometimes rather formalistic errors, caused by field entries that violate the format specification. Of course, usually, you can still work with a data set even if the CAS number is not following the official specification.

3.3.5   Other elements in the converter

Besides the conversion, the converter contains additional tools that can be useful "around" the conversion job. These tools are a query editor for the underlying database of the converter, and another editor for XPath expressions.

3.3.5.1   Database query editor

The converter contains a HyperSQL database which is used for managing the mapping files of the converter. HyperSQL is a free and open source Java database engine. A copy of the HyperSQL license is available at [7]. The database can be accessed via the "database" menu entry of the converter; physically,

it is located in a subfolder of the converter which is created at its first start. This subfolder contains the different tables of the database as csv files (Figure 23).

Name	Änderungsdatum	Typ	Größe
COMPARTMENT_MAP_ES1_TO_ES2.csv	21.01.2013 17:04	Microsoft Excel-C...	1 KB
COMPARTMENT_MAP_ES1_TO_ILCD.csv	21.01.2013 17:04	Microsoft Excel-C...	1 KB
COMPARTMENT_MAP_ES2_TO_ILCD.csv	21.01.2013 17:04	Microsoft Excel-C...	2 KB
COMPARTMENT_MAP_ILCD_TO_ES1.csv	21.01.2013 17:04	Microsoft Excel-C...	1 KB
COMPARTMENT_MAP_ILCD_TO_ES2.csv	21.01.2013 17:04	Microsoft Excel-C...	3 KB
database.properties	30.03.2013 14:51	PROPERTIES-Datei	1 KB
database.script	30.03.2013 14:51	SCRIPT-Datei	8 KB
ES1_COMPARTMENTS.csv	21.01.2013 17:04	Microsoft Excel-C...	1 KB
ES1_ELEM_FLOWS.csv	21.01.2013 17:04	Microsoft Excel-C...	225 KB
ES2_COMPARTMENTS.csv	21.01.2013 17:04	Microsoft Excel-C...	2 KB
ES2_ELEM_FLOWS.csv	21.01.2013 17:04	Microsoft Excel-C...	561 KB
ES2_GEOGRAPHIES.csv	21.01.2013 17:04	Microsoft Excel-C...	14 KB
ES2_TO_CSV_COMPARTMENT_MAP.csv	21.01.2013 17:04	Microsoft Excel-C...	1 KB
ES2_TO_CSV_ELECTRICITY_UNITS.csv	21.01.2013 17:04	Microsoft Excel-C...	1 KB
ES2_TO_CSV_GEOGRAPHY_MAP.csv	21.01.2013 17:04	Microsoft Excel-C...	1 KB
ES2_TO_CSV_SUBCOMPARTMENT_MAP.csv	21.01.2013 17:04	Microsoft Excel-C...	1 KB
ES2_UNITS.csv	21.01.2013 17:04	Microsoft Excel-C...	2 KB
FLOW_MAP_ES1_TO_ILCD.csv	21.01.2013 17:04	Microsoft Excel-C...	181 KB
FLOW_MAP_ES2_TO_ILCD.csv	21.01.2013 17:04	Microsoft Excel-C...	235 KB
FLOW_MAP_ILCD_TO_ES1.csv	21.01.2013 17:04	Microsoft Excel-C...	148 KB
FLOW_MAP_ILCD_TO_ES2.csv	21.01.2013 17:04	Microsoft Excel-C...	240 KB
ILCD_COMPARTMENTS.csv	21.01.2013 17:04	Microsoft Excel-C...	5 KB
ILCD_ELEM_FLOWS.csv	21.01.2013 17:04	Microsoft Excel-C...	2.460 KB
ILCD_FLOW_PROPERTIES.csv	21.01.2013 17:04	Microsoft Excel-C...	11 KB
ILCD_UNIT_GROUPS.csv	21.01.2013 17:04	Microsoft Excel-C...	1 KB
ProductNamePrefSave.ser	09.03.2013 19:01	SER-Datei	1 KB
UNIT_MAP_ES1_TO_ILCD.csv	21.01.2013 17:04	Microsoft Excel-C...	2 KB
UNIT_MAP_ES2_TO_ILCD.csv	21.01.2013 17:04	Microsoft Excel-C...	2 KB
UNIT_MAP_ILCD_TO_ES2.csv	21.01.2013 17:04	Microsoft Excel-C...	9 KB

Figure 23: Files in the 'database' subfolder of the converter.

You can browse the database using SQL syntax. Figure 24 shows a diagram of the different tables and their relations in the database (without the CSV additions).
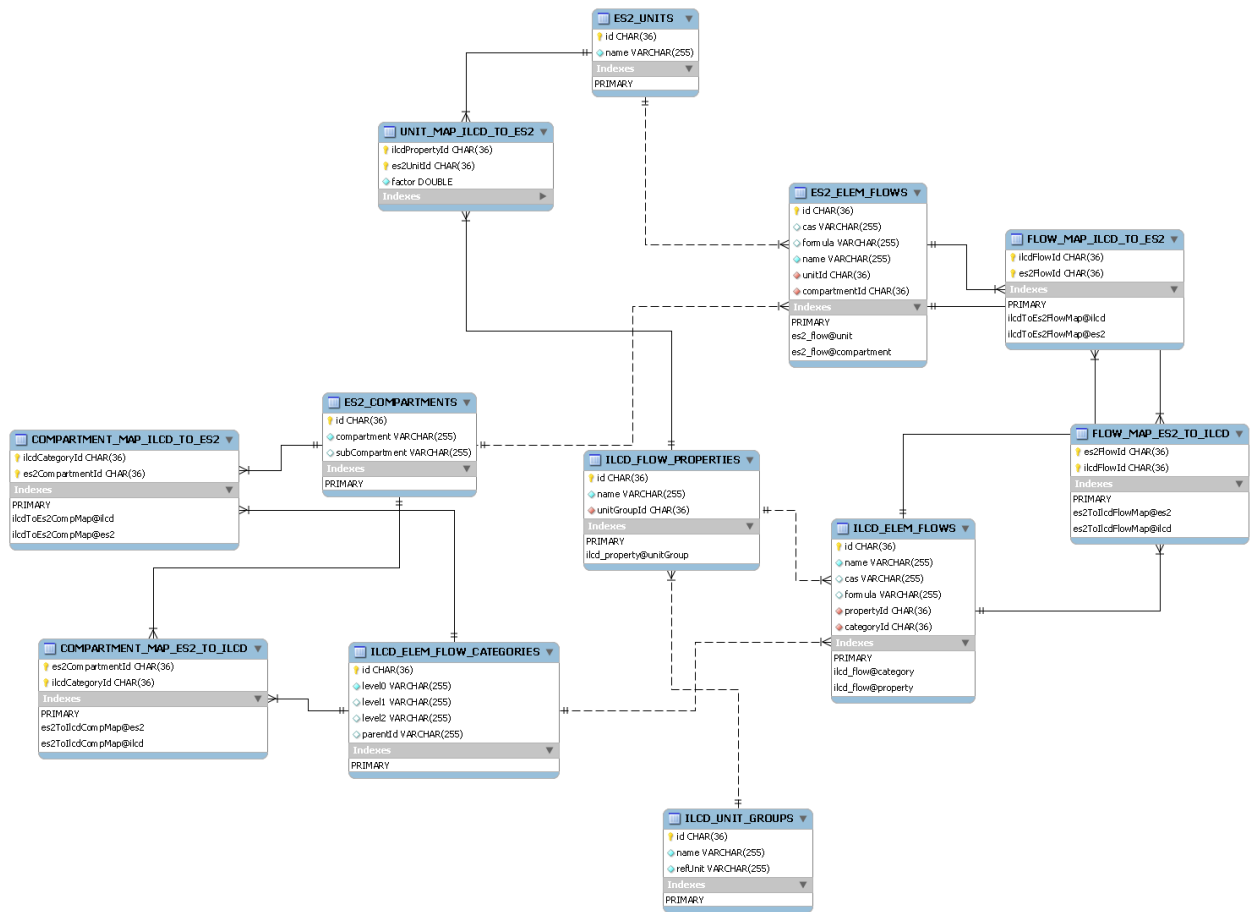
Figure 24: Entity-relationship diagram of the database in the format converter
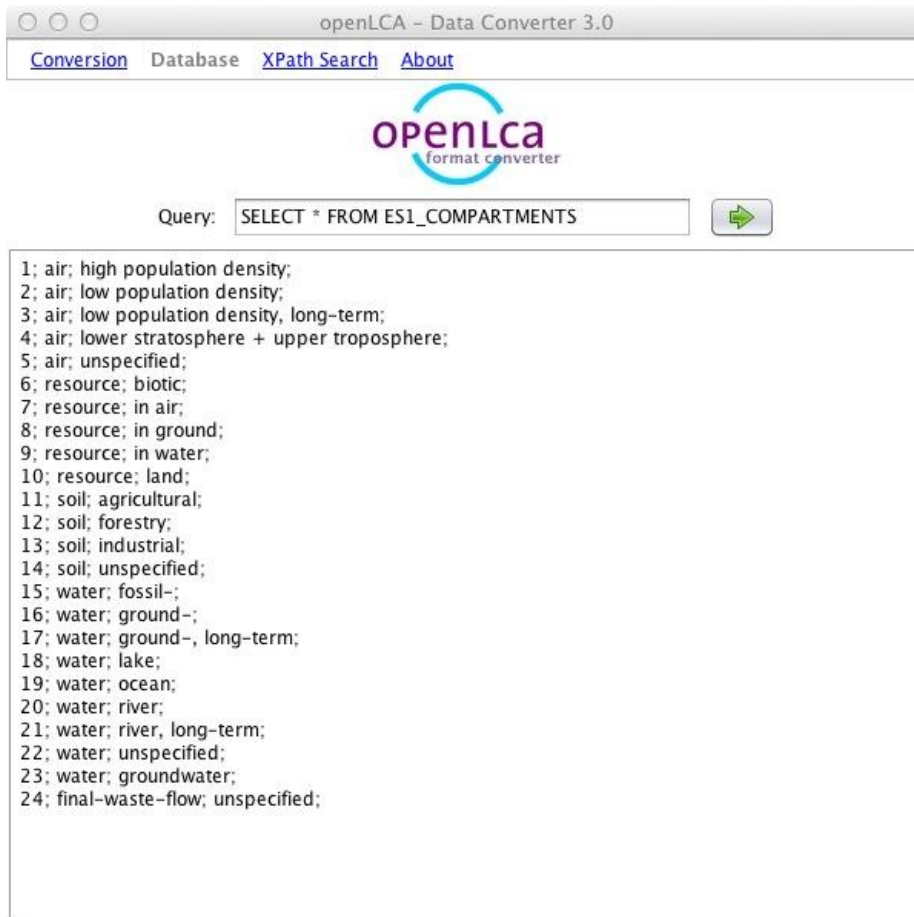
Figure 25: Database query editor with an example query.

3.3.5.2 XPath editor

The main function of the converter is to transfer data from one XML data format to another XML data format. To specify the conversion or analyse the conversion result it is helpful to know which kind of data are in the respective formats. As it is hard to read all the data from the XML files manually, the converter comes with a XML query tool – the XPath Search.

XPath is a query language for XML documents. You can find the official language specification here: www.w3.org/TR/xpath/, but you do not need to read this specification as the language is very simple and you can understand the key concepts with the query example below or the tutorial at www.w 3schools.com/xpath/.

XPath example: 'ecoSpold/dataset/flowData/exchange/@name' selects all flow names in the process data set; name is the attribute, and ecoSpold, dataset and so forth are nodes.
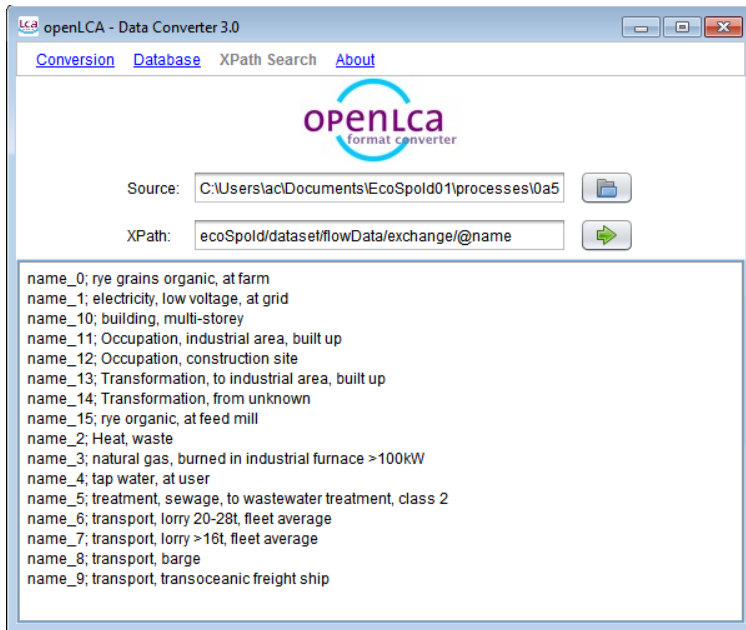
You can also filter, e.g. for those names that contain transport:

'ecoSpold/dataset/flowData/exchange[contains(@name, 'transport')]/@name' produces this:
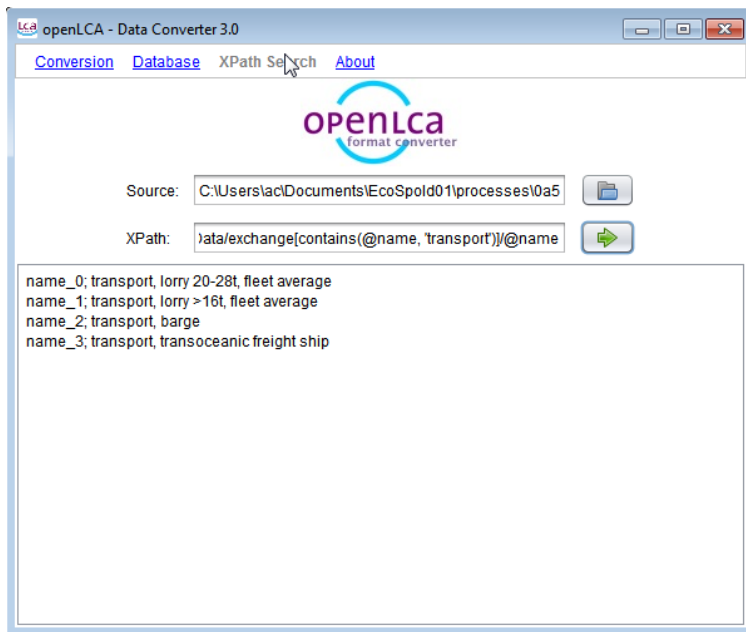
4 References and helpful links

[1]: Ciroth, A.: openLCA format converter user guide, Version 1.0 (2008)

[2]: www.openLCA.org

[3]: www.openLCA.org/downloads

[4]: www.ecoinvent.org/fileadmin/documents/en/EcoSpoldSchema_v1.0.zip

[5]: www.ecoinvent.org/fileadmin/documents/en/EcoSpold02.v1.0.2.zip

[6]: http://lca.jrc.ec.europa.eu/lcainfohub/developerDownload.vm

[7]: http://hsqldb.org/web/hsqlLicense.html

5 Contact & thanks

The openLCA format converter is created and maintained since 2006 by GreenDelta (before October 15 2012: GreenDeltaTC) in Berlin, and released as free, open source software.

In the course of the development, we received support and funding from various institutions, especially:

- PRé Consultants, Amersfort, NL (www.pre-sustainability.com)

- The UNEP/SETAC life cycle intiative (www.lifecycleinitiative.org)

- The ecoinvent centre, Zurich, CH (www.ecoinvent.org)

- PE International, Stuttgart, DE (www.pe-international.com)

- Forschungzentrum Karlsruhe, Institue for Applied Informatics, DE (www.iai.fzk.de)

Many, many thanks. It is not exagerated that without the sponsors, the format converter and the whole openLCA project would not have gone far.

But likewise, input and feedback from users is very important for us and for the development of the converter.

→ If you have ANY comments concerning the converter, be it

- Feature requests

- Possible bugs that you may have detected

- Improvement proposals

- …

then please do not hesitate to contact us:

Dr. Andreas Ciroth, GreenDelta GmbH, Berlin, Germany. www.greendelta.com

✉ gd@greendelta.com, ☎ +493048496030   THANK YOU.

Annex

6 Ecospold 2 to CSV (SimaPro) conversion assumptions and requirements

6.1 Conversion assumptions

1. If the intermediate exchange with OutputGroup 0 has a positive amount then the activity is a Process.

2. If an intermediate exchange with OutputGroup 0 has a negative amount then the activity is a Waste treatment.

3. If an activity has more than one intermediate exchange with OutputGroup 0 then the exchanges are written:

   - For Processes to CoProducts
   - For Waste treatments to Avoided Products.

6.2 Conversion requirements

Each activity requires an intermediate exchange with OutputGroup 0 (reference flow).

The following fields are necessary to convert the data set (in brackets the EcoSpold 2 ID number of the field):

   - Activity Id (102)
   - Activity Name (100)
   - Exchange Id (1005)
   - Exchange Name (1000)
   - Exchange Unit Name (1035)
   - Exchange Amount (1020)
   - Elementary Compartment (1660)