

Global LCA database access (GLAD) network

Elementary flow (EF) mapping project

Deliverable 1.3 – Methodology and common issues report

Version 2.2

5/12/2022

Authors:

Carl Vadenbo (vadenbo@ecoinvent.org) and Thomas Sonderegger (sonderegger@ecoinvent.org),
ecoinvent Association

Koichi Shobatake (shobatake@tco2.com) and Karkour Selim (karkour@tco2.com), TCO2 Co. Ltd.

Simone Fazio and Antonio Valente (antonio.valente@ec.europa.eu), Joint Research Centre,
European Commission

Ashley Edelen (ashley.edelen@erg.com), Eastern Research Group

Acknowledgements

This project was made possible by the combination of funding provided to the ecoinvent association and TCO2 Ltd. (representing the IDEA database of Japan) by the European Commission (through the REAL project, implemented by UNEP) and the Life Cycle Initiative, and substantial in-kind contributions from the four participating database providers. Ashley Edelen's participation on behalf of the U.S. Federal LCA Commons was made possible through the U.S. Environmental Protection Agency's contract EP-C-16-015, Task Order 68HERC19F0292 with Eastern Research Group. Wesley Ingwersen of the U.S. EPA and Gregor Wernet at the ecoinvent Association provided technical input in the early stages of this work. The authors would also like to acknowledge Edward Diaconu at EC-JRC for the development of the *GLAD Mapper* tool, as well as the members of the GLAD Working Group on Nomenclature and Martin Baitz and the team at Sphera for the constructive input during the project work.

Executive summary

Background to project

In the context of Life Cycle Assessment (LCA), nomenclature lists define interactions with the natural environment or between economic activities. These interaction referred to as elementary flows, EFs, and technosphere flows or intermediate exchanges, respectively. Nomenclature lists are essentially dictionaries that establish the basis for a common ‘language’ to build life cycle inventory (LCI) models. This information is crucial for several reasons, including for linking unit processes (representing individual activities along the value chain) into product systems, or for the consistent application of Life Cycle Impact Assessment (LCIA) methods. This enables us to quantify the environmentally relevant physical flows over the entire product system, and to translate these flows into the potential contribution to environmental problems (or ‘impact categories’). Consistent mapping between different nomenclature systems facilitates the conversion of datasets from one data exchange format to another, a prerequisite for the interoperability of LCA data from multiple sources (e.g., LCI databases) for a common application.

Initiated within the GLAD Working Group on Nomenclature (WG1), the aim of this project was to develop a common system to map the nomenclature lists for elementary flows (EFs) between databases connected to GLAD. The databases connected to GLAD are also referred to as ‘nodes.’ The project deliverables provide the foundation to improve the data exchange format conversion tool provided in the GLAD online user interface or by external tool developers.

This project was commissioned by UNEP through a small-scale founding agreement (SSFA), and with financial support from the REAL initiative of the European Commission and internal resources from the Life Cycle Initiative. Theecoinvent association coordinated the implementation from October 2020 to September 2021 (the project duration was extended from July 2021). The activities were undertaken in close collaboration with representatives from the U.S. Federal LCA Commons, the Inventory Database for Environmental Analysis (IDEA) of Japan, and the International Reference Life Cycle Data System (ILCD)/Life Cycle Data Network (LCDN) developed by the Joint Research Centre of the European Commission.

Summary of approach taken

Different formats are currently being used within the field of LCA to structure and exchange data. In addition, parallel nomenclature systems – such as elementary flow (EF) lists defining how interactions between the technosphere and the natural environment are inventorized – exists. Taken together, the differences in data exchange formats and nomenclature poses a major hurdle for data interoperability. The various EF lists used in LCA not only represent different naming conventions or preferences for flow attributes (like reference flow units or defining properties), but to some extent also alternative ways of characterizing human interactions and interventions with the natural environment.

There are two basic ways to map nomenclature systems onto each other, with the key difference being whether (i) all possible combinations of source- and target-side flow lists are mapped directly or (ii) via a common central list. A central flowlist might enable more efficient mapping, but to avoid compromising its scope or completeness, all differences between the other list need to be bridged. While one common (reference) nomenclature system would mark a giant step towards harmonization of elementary flows in LCA, the necessary level of stakeholder consensus might not be attainable in the near-term. Direct pairwise mapping, on the other hand, instantly exposes these differences, and it does so in a (more) neutral way. And as such, it allows the EF list owners involved to develop a better understanding of alternative perspectives. The present project adopted this

direct translational approach, with an explorative component to uncover – but not to judge – similarities and differences between the participating lists. These insights lay the foundation for further harmonization efforts in the future.

The relevance of this work is not limited to LCI modelling and databases: the project outputs are also expected to provide a starting point for the recently initiated GLAD-GLAM dialogue, e.g., by facilitating the two sides to come together for concrete case studies and systematic evaluations, for better alignment between live cycle inventory analysis and impact assessment. It should be noted that – while highly desirable – ensuring LCIA score consistency during format conversion was considered beyond the scope of this project. Instead, the flow-level mapping consistency across the various list combinations was as far as possible prioritized, and sometimes resulting in relatively conservative matches (see deliverable 1.3 for details).

Main outputs and achievements

At the outset of the project in mid-2020, the project team departed from the common data format and structure for EF lists and EF mapped files, established by WG1 prior to the project inception. In addition, the team at the Joint Research Centre (JRC) of the European Commission provided, and continuously further developed, a dedicated software tool (the ‘GLAD Mapper’, also referred to here as the mapping script) as an in-kind contribution to support the project activities. Based on these initial mapping resources, the subsequent efforts of the project team related to three main aspects, namely: i. to establish and compile the information necessary to run the mapping script, while also expanding and fine-tuning its mapping algorithm, ii. to set up the procedures and tools for, as well as actually conducting, the review of the preliminary mapping results (i.e., the ‘raw’ outputs of the mapping script), and; iii. to document the approach and tools applied, as well as the common issues encountered, during the mapping activities. The last area also included developing a proposal for how the GLAD EF mapping system can be maintained beyond the present project.

The main inputs for the mapping script, besides the EF lists and the mapping algorithm itself, encompass manual mapping files (providing the relevant criteria and parameters for the mapping script algorithm) for all combinations and directions between the four participating database systems. The manual mapping input files were extracted from extensive collections of information, structured into (pairwise) context mapping files, a joint flowable library, and a conversion factor library. These mapping resources were not part of the required project deliverables but were set up by the team to simplify the handling and review of the information needed for the mapping script. In contrast, the output of the mapping script, referred to as preliminary EF mapped files, provided the starting point for cross-review and manual refinements to finalize mapped files between the following flowlists:

- European Commission’s Environmental Footprint version 3.0 (ILCD-EF3.0)
- ecoinvent version 3.6/3.7 of Switzerland
- IDEA versions 2.2 and 2.3 of Japan
- U.S. Federal LCA Commons version 1.0.3 (FEDEFLv1.0.3)

Finally, different aspects of the project work and recommendations for the future management of the GLAD EF mapping system were documented in two reports: a transparent account of the procedure, common issues and solutions implemented to establish the mapping files, as well as guidance for developers of data format converters, is provided in this report (towards deliverable D 1.3). This project document is complemented by deliverable 2.3 (‘Maintenance manual for the GLAD EF mapping system’), which contains a proposal for a governance structure and maintenance procedures for the GLAD EF mapping system beyond the present project.

Outlook

The project deliverables, mapping resources, and further relevant information will be uploaded and shared through UNEP/GLAD's GitHub repository upon final acceptance. This will enable end-users or tool developers to make improvement suggestions or report errors in the EF mapped files, while also ensuring that all changes to these knowledge resources are transparently recorded and provide version control of all files. The mapping information generated, together with the overview and analysis of common mapping issues presented in deliverable 1.3, will provide valuable inputs to the recently initiated GLAD-GLAM dialogue. Furthermore, the team at the EC-JRC has taken the on a scientific publication with the aim to summarize the background, goal and scope, approach, and results of the project, as well as the lessons learned in the process.

Table of Contents

Acknowledgements.....	2
Executive summary	3
Background to project	3
Summary of approach taken.....	3
Main outputs and achievements	4
Outlook	5
Glossary.....	8
Introduction and background to the project	9
General mapping procedure.....	10
Common list and mapping format	11
Context mapping.....	12
Mapping of flowables and flows	13
Mapping of flowables	13
Mapping of flows	13
Generation of preliminary mapped files.....	14
Conversion factors	18
Review of proposed matches and mapping results.....	19
Common mapping issues	20
Issues related to matching of contexts (flowable-independent).....	20
Compartments for long-term emissions.....	20
Matching of (partially) overlapping contexts.....	21
Proxy contexts for ‘unspecified’ and indoor air sub-compartments	23
Issues related to matching of flowable (context-independent)	24
Specific to less specific flowables.....	24
Mapping of closely ‘related’ chemical substances	25
Matching the main greenhouse gases	26
Mapping of elements vs. ions vs. compounds (unspecified)	30
Flowable approximations (proxies).....	31
Challenges related to matching of flows (flowable+context combinations)	32
Mapping of water flows	32
Other aspects: missing characterization factors.....	38
Guidance for developers of data format converters	39
Structure and content of GLAD’s elementary flow mapped files	39
Log of conversion	41
Mapping record in exchange comments	41

Version control/support	42
Proposals for corrections or further improvements.....	42
Appendix A – General mapping principles.....	43
Context mapping.....	43
Mapping of flows or flowables	43
Unit mapping	44
Appendix B – overview of project outputs	45

Glossary

Context	The attribute of a flow that defines its direction (to/from the environment, i.e., a resource/emission), as well as the source/receiving environmental compartment (e.g., air, water, or soil/ground) and one or more sub-compartments
Compartment	The main environmental compartments compass air, soil/ground, and water. Some elementary flow lists include additional compartment, e.g., for land use. Each compartment can be disaggregated further into sub-compartments to differentiate, for example, between emissions occurring in urban and rural areas, at different altitudes, or to different types of waterbodies.
Confidence rating	A rating system that allows the source- and target-side mappers/reviewers to indicate their respective perceived (i.e., subjective) levels of confidence in each pair of matched items.
Emission	Elementary flow to the natural environment caused by human activities (i.e., from the technosphere)
(Elementary) flow	The combination of a flowable, context, and reference unit that defines an exchange with the natural environment.
Flowable	The name (e.g., of a chemical substance or natural resource) and other attributes of a flow, but without the environmental compartment (context)
Match condition	The logical relationship between the scopes or boundaries for a pair of contexts, flowables, or flows to be mapped (i.e., the objective alignment of scope/boundaries between the source- and target-side, indicated by logical operator symbols, such as '=', '~', '<', '>', or '<>')
Resource	Elementary flow extracted from the natural environment and put into the technosphere by human activities
UUID	Universally Unique Identifier, a standardized system ¹ to label and uniquely identify information, such as individual flows defined in elementary flow list

¹ ISO/IEC 9834-8:2014 *Information technology — Procedures for the operation of object identifier registration authorities — Part 8: Generation of universally unique identifiers (UUIDs) and their use in object identifiers.*
<https://www.iso.org/standard/62795.html>

Introduction and background to the project

The *Global LCA Data Access (GLAD)*² network is a platform, established under the *Life Cycle Initiative* and hosted by United Nations Environmental Programme (UNEP), that allows users to find and access life cycle inventory (LCI) datasets from different data providers. On its open online user interface, GLAD allows anyone to search, compare, convert, and download datasets for life cycle assessment (LCA). Through GLAD, data are centralized to improve access and facilitate interoperability between common data exchange formats. This increased accessibility and functionality benefits data providers as well as data users, including LCA researchers and practitioners in business or industry and public policymakers, worldwide.

In the context of LCA, nomenclature lists define interactions with the natural environment or between economic activities (also referred to as elementary flows, EFs, and technosphere flows or intermediate exchanges, respectively). This information is essential for several reasons, including for linking unit processes into product systems or for the consistent application of Life Cycle Impact Assessment (LCIA) methods. Consistent mapping between different nomenclature systems facilitates the conversion of datasets from one data exchange format to another, a prerequisite for the interoperability of LCA data from multiple sources for a common application.

Initiated within the GLAD Working Group on Nomenclature (WG1), the aim of the present project was to develop a common system to map the nomenclature lists for elementary flows (EFs) between databases connected to GLAD. The databases connected to GLAD are also referred to as ‘nodes.’ The project deliverables provide the foundation to improve the data exchange format conversion tool provided in the GLAD online user interface. Commissioned by UNEP through a small-scale founding agreement (SSFA), and with financial support from the *REAL initiative*³ of the European Commission and internal resources from the Life Cycle Initiative, the project was implemented by theecoinvent association, the publisher of the ecoinvent LCI database, from October 2020 to July 2021. The activities were supported by representatives from the U.S. Federal LCA Commons, the Inventory Database for Environmental Analysis (IDEA) of Japan, and the International Reference Life Cycle Data System (ILCD)/Life Cycle Data Network (LCDN) developed by the Joint Research Centre of the European Commission.

More specifically, the main goal of this project was to create bidirectional EF mapped files between four widely used nomenclature lists, from four major LCA databases. These mapped were established using a common format and file structure and including transparent documentation of any assumptions made in the process. The documentation encompasses an account of ‘unmapped’ list items (indicated in the respective mapped files; project deliverable D1.2), any rationales for how imperfect matches or approximations are treated, and the identification of more general issues. The two latter aspects are addressed in this report (deliverable D1.3). Finally, the project team has outlined procedures for how the nomenclature mappings, established within the GLAD network, can be maintained, and expanded in the future (in deliverable D2.3).

² <https://www.globalcadataaccess.org/>

³ The *Resource Efficiency through Application of Life cycle thinking (REAL)* project <https://www.unep.org/explore-topics/resource-efficiency/what-we-do/life-cycle-initiative/real-project>

General mapping procedure

There are two basic ways to map nomenclature systems onto each other, with the key difference being whether (i) all possible combinations of source- and target-side flow lists are mapped directly or (ii) via a central list as the common reference system. The latter approach has a clear advantage in that number of mappings that needs to be established is linearly proportional to the number of participating lists (n). In contrast, the more translational pairwise mapping (in i.) of a list onto all other lists means that the number of mapping files required to be established grows exponentially, as $n \cdot (n-1)$, with each new list added to the mapping system.

The various elemental flow lists in use in LCA not only represent different naming conventions or flow units, but to some extent also fundamentally different ways of defining or characterizing human interactions/interventions with natural environment. So, to determine a common central flowlist for more efficient mapping, but without compromising its scope/completeness, implies these differences need to be bridged. While this would mark a giant step towards harmonization of elementary flows for use in LCA, the necessary level of consensus might not be attainable near-term. The direct pairwise mapping, on the other hand, instantly exposes these differences, and in a (more) neutral way. And as such, it allows the list owners to develop a better understanding of alternative perspectives. The present project adopted the translational n -to- $(n-1)$ approach, with an explorative component to uncover – but not to judge – similarities and differences between the participating flowlists. These insights provide a foundation for further harmonization efforts in the future.

The procedure established and followed in this project entails the following main steps, which are illustrated in Figure 1 and described in more detail in the sections below.

0. Transfer of the native elementary flow (EF) lists to a **common GLAD EF mapping format**
1. **Context mapping**, to determine the correspondence (as 'default' matches and acceptable proxies, the latter hierarchically ordered by proximity) between environmental compartments and sub-compartments across the EF lists.
2. **Mapping of 'flowables'**, by considering the flow name and the rest of the flow attributes without the environmental compartments. This is mainly applied to flowables associated with some type of ambiguity that might prevent the mapping script from finding an appropriate match based only on the main attributes (flow name, CAS, synonyms, context).
3. **Mapping of individual flows**, but also splitting or blocking/omitting specific flows (applied restrictively to resolve special situations only).
4. **Generation of preliminary mapped files**, using the GLAD Mapper mapping script, developed by JRC, that is fed with the information established in the preceding steps. This software tool establishes unidirectionally mapped files by applying a hierarchical set of rules to match the flows between the source- and target-side EF lists in question.
5. **Review of preliminary mapped files**, by representatives of both the source- and target-side EF lists. This step also includes **manual modifications** as well as the introduction of quantitative **conversion factors** (e.g., for matched flows with different units, such as energy resources in kg vs. MJ).
6. **Further fine-tuning** of context mapping and flowables not captured (or which mapping can be improved) in the preliminary mappings, steps 1-5 can be repeated iteratively, with progressive refining until a final mapping with a reasonable level of coverage is achieved.

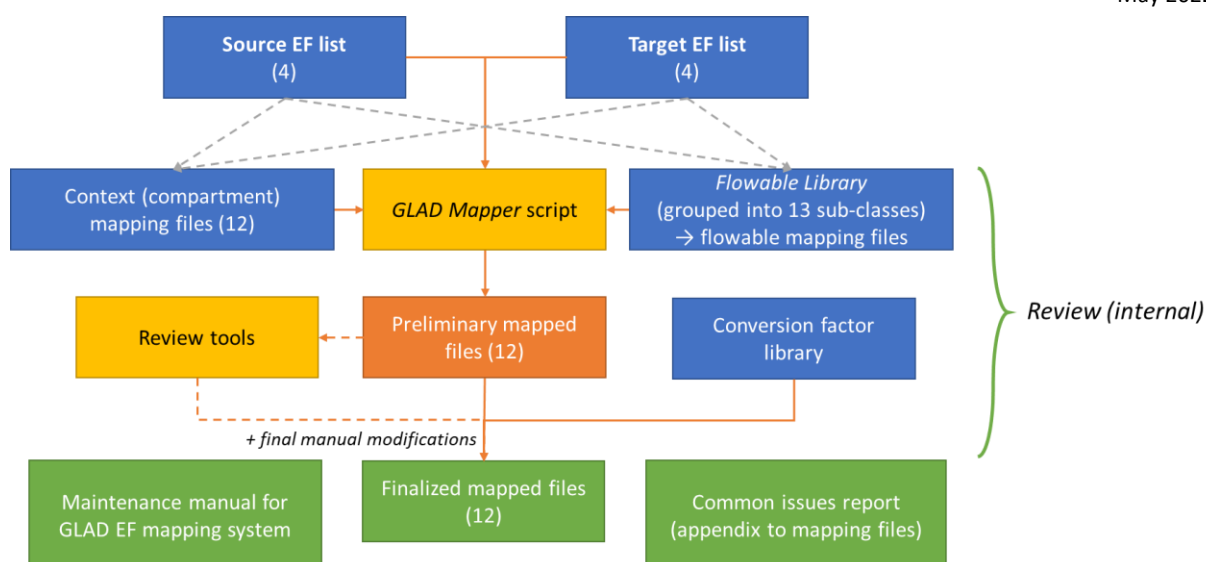


Figure 1 – overview of the procedure and tools applied to establish the EF mapping resources in the present project.

A summary of general mapping principles applied is provided in Appendix A at the end of the document. The rest of this section is dedicated to describing the approach and procedures developed within this project to establish the EF mapping resources. It is structured into six sub-sections, largely following the work/information flow outlined in Figure 1: i. Common list and mapping format, ii. Context mapping, iii. Mapping of flowables and flows, iv. Generation of preliminary mapped files, v. Conversion factors, and vi. Review of proposed matches and mapping results.

Common list and mapping format

The common file structures for the elementary flow (EF) lists and the resulting EF mapped files were defined ahead of the present project, setting the standard for the project deliverables. An overview of the information to be contained in the GLAD EF mapped files, ordered as columns in EF mapping file (as .xlsx- or .csv) for any pairwise combination of source- and target-side EF lists, is provided in Table 3 in the section ‘Guidance for developers of data format converters’ at the end of this document.

The common format for EF lists within GLAD covers the following flow attributes:

- Name and version of the **source flow list** [required]
- **Flowable** (or flow name), e.g., sulfur dioxide [required]
- **CAS Registry Number**: (CAS: Chemical Abstracts Service), e.g., 7446-09-5 for sulfur dioxide. It is important to notice, however, that a chemical substance might be associated with more than one CAS number. For example, the PubChem database⁴ also lists 12143-17-8 and 12624-32-7 for sulfur dioxide. [optional]
- (Molecular) **formula**, e.g., SO₂ or O₂S [optional]
- **Synonyms**, including alternative chemical naming conventions, trade names, etc. [optional]
- **Unit** of flow, e.g., kg, m³, MJ, kBq, etc. Although the units used in the respective EF lists could also have been mapped as part of establishing the mapping files, we opted for using a

⁴ NIH (2021) *PubChem compound summary - sulfur dioxide*. National Center for Biotechnology Information, U.S. National Library of Medicine, National Institutes of Health (NIH), U.S. Department of Health and Human Services. Online: <https://pubchem.ncbi.nlm.nih.gov/compound/1119> [assessed 2021-07-26]

common set of units from the start. For this, the set of units found in the OpenLCA software was chosen. *[required]*

- **Class**, according to classification of elementary flows proposed by Edelen et al. (2018)⁵ *[required]*
- **External reference**, e.g., a URL link *[optional]*
- **Preferred**: attribute to indicate the preference or priority in case of multiple corresponding or overlapping flows (e.g., sharing a common CAS number) *[optional]*
- **Context**: an aggregation of the direction and the environmental compartments for the elementary flows, e.g., ‘emission/air/urban air close to ground’ or ‘resource/ground/non-renewable/material’ *[required]*
- **Flow UUID** (UUID: universally unique identifier) *[required]*
- **Alternative unit**, e.g., ‘MJ’ for energy resources with ‘kg’ as the main flow unit *[optional]*
- **Alternative unit conversion factor**, e.g., based on the calorific value of energy resources *[optional]*

While the EF lists used within this project refer to specific published list/database versions, the project team allowed itself to add, correct, or otherwise modify attributes in the EF lists files established specifically for the GLAD EF mapping system. The direct correspondence to the original lists was maintained through unchanged UUIDs, with one exception (for which the same UUIDs had been used for multiple flows). The main objective here was to improve the mapping efficiency and/or success rate in subsequent iterations. This will also enable the EF list owners to correct or ‘enrich’ the information provided in future releases of their own flowlist.

Context mapping

The context mapping contains pairwise matches of flow attributes that indicate flow directions, environmental compartments, and one or more sub-compartment. The number of individual contexts found in the EF lists considered for this project ranged from 20 to 113 (see Table 3). The longest EF list (U.S. Federal LCA Commons) differs from the other three in that it provides contexts in a hierarchical structure, spanning up to four environmental compartment levels, e.g., ranging from ‘emission/air’ to ‘emission/air/troposphere/urban/ground-level’. The contexts in the other three lists are typically defined for a main environmental compartment (air, ground/soil, water) and one sub-compartment level (in addition to level “zero”, typically emissions, resources, or land use). In these lists, the ‘unspecified’ sub-compartment may span all and any of the more specific sub-compartment, or represents the rest (i.e., not captured by the more specific sub-compartment) depending on the EF list in question.

The contexts were mapped for each pairwise list combination in both mapping directions. The resulting context mapping files are structured into two sections, with the first listing what the project team considered the most suitable (‘default’) matches. The second section contains any further combinations of source- and target-side contexts that are also considered acceptable matches (‘proxies’), ordered from the most to the least preferred target-side option for each context in the source-side EF list. For example, the context ‘emission/air/urban/ground level’ might be assigned ‘emission/air/urban/close to ground’ as the default match in the target-side EF list, with second priority (as proxy) give to ‘emission/air/urban/low’ followed by ‘emission/air/urban/high’ or just ‘emission/air/urban’, and so on. Any match of contexts not defined among the allowable ‘default’ or ‘proxy’ combinations will not be considered by the mapping script, unless defined

⁵ Edelen, A., Ingwersen, W.W., Rodríguez, C. *et al.* Critical review of elementary flows in LCA data. *Int J Life Cycle Assess* **23**, 1261–1273 (2018). <https://doi.org/10.1007/s11367-017-1354-3>

explicitly for the mapping of individual flows (as described in the next sub-section on *Mapping of flowables and flows* below).

Mapping of flowables and flows

The length of the participating EF lists varies greatly, from a few hundred or thousand flows (in IDEA and ecoinvent) and into the hundred-thousands (Federal LCA Commons). Two factors greatly influencing the total length of the lists are the number of flowables (i.e., single entries of chemical substances, land use and water use types, resources, etc.) included, and the number of contexts (i.e., the environmental compartments and sub-compartments considered separately) for which these flowables are defined. That is, a flow is in this context the unique combination of one flowable and one context (see Table 1) and it is further defined by the unit assigned to the flow (e.g., kg, Nm³, MJ).

With the aim to minimize the manual efforts for identifying the ‘best’ matches between the flows of any combination of source- and target-side EF lists, a three-tiered approach is used: contexts are first mapped separately (as described in the section above), followed by groups of flowables that are either lacking attributes for unambiguous matching by the mapping script, or considered of particular importance. The manual mapping of flows (i.e., as unique combinations of flowables and contexts) is only applied in cases where the general mapping criteria applied cannot be applied in a systemic way, or would otherwise fail (e.g., for flows related to land use, due to inconsistent naming cross the lists and no further attributes, like CAS or synonyms).

	Flowables	Contexts	Flows
ILCD-EF 3.0	7'741	37	93'993
FEDEFL v1.0.3	5'933	113	278'602
IDEA v2.3	605	20	903
Ecoinvent v3.7	1'404	23	4'310

Table 1 - Count of flowables, contexts, and flows in the four native EF lists used for the mapping generation

Mapping of flowables

In this step, the flowables are grouped and manually matched to each other to establish the most appropriate mapping outcomes. This allows, for example, for pairing flowables for the main greenhouse gases CO₂ and CH₄ according to carbon source, such as fossil biogenic/non-fossil, or from soil or biomass stock, for the EF lists where this differentiation is made. The flowable mapping is also applied to flowables considered largely equivalent but associated with different flow names, multiple CAS numbers, chemical forms (enantiomers, isomers, etc.), or to assign overlapping or proxies flowables like ‘hydrocarbons, chlorinated’ or ‘herbicides, unspecified’ to specific substances.

These matches, as well as the evaluation thereof, are compiled in the ‘flowable library’, from which the pairwise flowable mapping is extracted as an input for the mapping script. Duplicate entries of source-side flowables are allowed, since the ideal match might not be defined for all relevant contexts in the target list, and hence require a suitable proxy to be provided in parallel. It is important to ensure that the proxy does not take priority over the preferred match in the contexts in which both are available in the target list. This can be resolved either by structuring the flowable library to adhere to the mapping sequence of the mapping script (ordering prioritized flowables ahead of proxies), or by defining (some of) these matches as flows (i.e., set flowable-context combinations) instead. This is described further in the next section.

Mapping of flows

In this part of the procedure, specific combinations of flowables and contexts are either matched, split, or blocked from being mapped. This step is intended to address flows that do not match or fit

with the criteria subsequently applied in the mapping script for the generation of preliminary mapping files. The main flow groups that fall into this category are related to land use, which typically carry different names across the EF lists and are neither associated with CAS numbers nor synonyms, and exchanges of water with the natural environment.

Another case that requires manual flow-to-flow mapping is natural resources that have been defined as belonging to different main environmental compartments across the flowlists. Natural gas, for example, is considered a resource in air (e.g., subterranean) in the Federal LCA Commons flowlist, whereas the other lists consider natural gas a resource in ground. These context combinations are not defined, and hence not allowed, according to the main context mapping (as described above). But specific exceptions to these rules can be created through the flow-to-flow mapping.

Flows representing land transformation in IDEA's EF list include both the "from" (or "start") and "to" (or "end") state in one single flow, while in the two other lists where land transformations are considered (ILCD-EF and ecoinvent), the "from" and "to" states are defined as separate flows. To cover at least the direction with IDEA's list on the source-side, a special mapping rule enables these land transformations to be mapped against two target-side flows, to capture both the start- and end-state of the land. This is possible since both target flows match the single source-flow with a 1:1 relationship, which does not alter the land 'balance' of a given inventory. The opposite direction (i.e., merging "from" and "to" land transformations in ILCD-EF or ecoinvent into single flows found in IDEA's flow list) is not mapped, as this could potentially distort the inventories during conversion. Finally, included here are also specific source-side flows that should not be mapped onto a given target-list for any particular reason.

Generation of preliminary mapped files

The team at the Joint Research Centre (JRC) of the European Commission developed and provided a dedicated software tool (named *GLAD Mapper*⁶, also referred to here as the 'mapping script') as in-kind contribution from the EC to support the mapping activities in this project. The tool follows the GLAD .xlsx-file format for EF mapping. Based on user-specified inputs, such as default matches and acceptable proxies for mapping contexts, it generates mapping files between the EF lists provided by different nomenclature systems or GLAD nodes, as well as between different versions of the same nomenclature system. The mapping script utilizes one source list, one target list, and the relevant context mapping inputs, to match the single flows in a 1:1 comparison. To establish a bidirectional mapping between two EF lists, the tool needs to be executed in both directions, with the source list and target list exchanged, resulting in two EF mapped files.

⁶ https://eplca.jrc.ec.europa.eu/permalink/GLAD_Mapper/GLAD_Mapper.zip

The tool procedure to match flows and establish these initial mapping files can be summarized by the information requirements and the algorithm outlined below. The tool is fed by three input files:

- **Manual mapping file**, which includes the relevant input information for the tool, besides the pair of elementary flow lists to be mapped. This information is structured into the following spreadsheets:
 - **DEFAULT_CONTEXT_MATCH**: definition of the default or primary proxy matches of compartments/sub-compartments (flow contexts) for a given combination of source and target EF lists. This table establishes the exact match or the best proxy available for each entry from the source to the target list
 - **PROXY_CONTEXT_MATCH**: specification of secondary context proxies, where applicable. Going from the source to the target list allows for more than one suitable or relevant context in the target list to be considered by the mapping script. If this is the case, then the secondary context proxies are ordered hierarchically according to proximity to the source list. For example, for the source-side context 'emission/air/urban/ground level' → default or best proxy: 'emission/air/urban/close to ground'; second-priority proxy: 'emission/air/urban/low'; third proxy: 'emission/air/unspecified', and so on. Proxies too different or far from the original compartment shall be avoided. In the example above, e.g., 'emission/air/non-urban/low' shall be avoided, usually the "unspecified" proxy should close the list of proxies unless the "unspecified" itself is the source context).
 - **NO_FLOW_MATCH_MANUAL**: a list of flows that should not be mapped at all, to prevent these from being considered by the tool and potentially leading to incorrect mapping entries. Both flow name and context are required, in case a specific flow should not be matched for specific compartments but can find matches in others.
 - **ONEZONE_FLOW_MANUAL**: includes groups of flows that do not match with the criteria considered by the mapping tool (e.g., land use flows that carry different names, have no CAS and no synonyms) or for which the criteria can lead to mismatches (e.g., carbon dioxide that carries the same CAS irrespective of list-specific differentiations based on the source of the carbon, i.e., biogenic/non-fossil, fossil, land use change, etc.). This table includes names and contexts for the specific entries both for source and target lists, target flows can be repeated, while the source list shall include each flow only once.
 - **FLOWNAME_MANUAL**: similar to **ONEZONE_FLOW_MANUAL**, this list includes groups of flow items that are not to be matched with the logical iterations of the mapping script. But while **ONEZONE_FLOW_MANUAL** operated on the level of individual flows (i.e., unique combinations of flow names and contexts), only flow names (flowables) in the source and target flow lists are considered here. The context (that can be multiple for each flow) is resolved by the mapping script at the level of **DEFAULT_CONTEXT_MATCH** and **PROXY_CONTEXT_MATCH** tables, as above the source side of the list shall include each flow only once, while the target list can contain multiple entries for the same flow
 - **FLOWNAME_MANUAL_PROXY**: same rules and function as **FLOWNAME_MANUAL**, but it is only applied after the other checks and linking steps in the mapping script and is typically meant for matches that have a lower quality rank, e.g., specific substances to groups of substances (for example a specific chemical to "pesticides unspecified" or a specific hydrocarbon to "hydrocarbons unspecified"). The rationale

of this second level for FLOWNAME_MANUAL is that if a better proxy is captured by other iterations, e.g., a CAS match, then this should take priority over the “low-rank” matches in FLOWNAME_MANUAL_PROXY.

- ONE2MANY_FLOW_MANUAL: specific for IDEA land transformation flows which includes “from” and “to” in one single flow, while in other systems the “from” and “to” are listed as separate flows
- MULTIMATCH (PROCESS ONLY): not meant for the mapping script, essentially is available only in context mapping from all other systems to IDEA, reverting the ONE2MANY_FLOW_MANUAL ratio, but given that the “from” and “to” in the process dataset may include different land transformation paths, the information can be used at the level data conversion assigning the proper (equal) quantities in the “from” and “to” entries coming from IDEA process datasets.
- CONVERSION: includes the list of conversion factors where applicable, the keys for assignment (info needed by the mapping tool) are the flow name and unit in source and on the target side, plus the conversion factor.
- **Source and target EF lists** (two input files), formatted according to the common GLAD EF format for mapping purposes, including the following flow attributes: flow name (i.e., flowable), CAS [optional], synonyms [optional], flow unit, unique identifier (UUID) in the native lists, context (i.e., the flow compartment and sub-compartments merged into one string, e.g., “emission/air/urban/close to ground”). Beyond the information provided in the native lists, an additional field is provided, containing any (secondary) CAS registry numbers according to those available in Pubchem⁸. The additional CAS numbers are obtained through the URL-based API “PUG REST”⁹ released by PubChem: an ad-hoc excel function was developed to query the API via VBA code. The function accepts in input the name of a substance as the only parameter¹⁰. The PubChem PUG RES API provides CAS numbers as a synonym of a substance or compound. The output of the VBA function is a list of CAS numbers (as a string separated by semicolons), which is used to fill the secondary CAS field for each substance in the GLAD flow lists.

Based on the abovementioned files, the tool performs the mapping according to a series of sequential logical steps, and for the matches found reports the string is taken from the specific flow from the source with all the info, from the target flow list. The steps are in a logical and hierarchical sequence, and once a match is identified on the source side, the flow is excluded from any further checks (i.e., the source flows are reported only once in the mapping file). The only exception to that rule is for land transformation flows where IDEA is the source, according to the framework defined in ONE2MANY_FLOW_MANUAL above the flows have indeed to be split (so mentioned twice in the source), in order to be matched with both the “from” and the “to” in the target list.

The logical checks performed by the tool are in hierarchical order, with each step generally performed in two iterations (exceptions are mentioned below): the first one checks DEFAULT_CONTEXT_MATCH for primary context, while the second one checks PROXY_CONTEXT_MATCH contexts, and therein according to the hierarchical order provided. The

⁸ <https://pubchem.ncbi.nlm.nih.gov/>

⁹ <https://pubchemdocs.ncbi.nlm.nih.gov/pug-rest>

¹⁰ This parameter is used to query the PUG REST API using the string “[https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/name/\[SUBSTANCE_NAME\]/synonyms/txt](https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/name/[SUBSTANCE_NAME]/synonyms/txt)” via HTTP request, the PubChem server returns the list of synonyms of “SUBSTANCE_NAME” in text-plain format. The list of synonyms is then screened by the VBA function to identify the CAS-like items.

tool maps flows and assigns a specific mapping type label to the entries that are matched, according to logical steps:

- 1) Manual checks based on the input tables created by users (ONE2ONE_FLOW_MANUAL, NO_FLOW_MATCH_MANUAL, FLOWNAME_MANUAL and ONE2MANY_FLOW_MANUAL), in this first block the iterations for name-name are performed against the primary context mapping before, then secondary context (PROXY_CONTEXT_MATCH)
 - a. NO_FLOW_MATCH_MANUAL (i.e., source-side EFs not to be mapped): excludes the listed flows from the subsequent steps of the mapping script
 - b. FLOWNAME_MANUAL checks: identification of matches based on flow names (flowables) only, for the specific source-target list-combination.
 - c. ONE2MANY_FLOW_MANUAL checks: generation of matches for flows related to land transformation in IDEA to the two pre-selected flows in the target list. This step does not search for a context match in DEFAULT_CONTEXT_MATCH and PROXY_CONTEXT_MATCH since it is established *a priori*.
 - d. ONE2ONE_FLOW_MANUAL checks: assignment of matches manually defined for specific flows (i.e., for unique combinations of flow names and contexts) for the specific source-target list combination. This step does not search for a contexts match in DEFAULT_CONTEXT_MATCH and PROXY_CONTEXT_MATCH since it is established *a priori*.
- 2) Automated checks: the iterations are performed against the primary (default) context mapping for CAS and name, then steps C and D are performed, then, a second iteration in the same order for CAS and NAME is performed for the secondary context (PROXY_CONTEXT_MATCH).
 - a. By CAS: the script searches the source and target EF lists for flows with identical CAS numbers. Some formats include CAS with a fixed number of characters, filling with a set of "0" prefix the gaps to reach the fixed value, this is ignored by the mapping script (e.g., 000110-63-4 is considered identical to 110-63-4).
 - b. By flow name: the name is identical in the source and target list (non case-sensitive).
 - c. By flow name to synonyms: the name of the source is identical to one of the synonyms of the target.
 - d. By synonyms to name: a synonym of the source matches one of the names in the target.
- 3) Additional checks: the iterations are performed against the primary context mapping for secondary CAS, then for the secondary context (PROXY_CONTEXT_MATCH), then step b is performed.
 - a. By secondary CAS: either the main (already in the native flow list) or one of the secondary CAS (provided in PubChem) matches with either the primary or one of the secondary CAS in the target.
 - b. By FLOWNAME_MANUAL_PROXY final checks: performs the last check on remaining unmapped, according to the same rules as point 1.a (but using table FLOWNAME_MANUAL_PROXY), the entries in this last check are those with a lower match ranking and therefore are checked as the last iteration so that if one of the previous steps capture a better proxy, the match is excluded from this last step.
- 4) All the remaining flows on the source list that don't find a match are considered unmapped.

- 5) Where applicable, a conversion factor is assigned. **If the conversion field is empty, the factor to be assigned is 1, if it is reported as N/A the mapping shall be excluded** (see further explanations in chapter “Guidance for developers of data format converters”)

The following labels are assigned by the tool to characterize the “map type”, according to the abovementioned iterations¹¹, in the logical order of the mapping script (the cases indicted below following the levels of the list above):

- **NO_FLOW_MATCH_MANUAL**: Case 1.a.
- **FLOWNAME_MANUAL**: Case 1.b. primary context mapped
- **FLOWNAME_MANUAL (PROXY)**: Case 1.b. proxy context mapped
- **ONE2MANY_FLOW_MANUAL**: Case 1.c.
- **ONE2ONE_FLOW_MANUAL**: Case 1.d.
- **CAS**: Case 2.a. primary context mapped
- **NAME**: Case 2.b. primary context mapped
- **SYNONYM_TO_NAME**: Case 2.c.
- **NAME_TO_SYNONYM**: Case 2.d.
- **CAS (PROXY)**: Case 2.a. secondary context mapped
- **NAME (PROXY)**: Case 2.b. secondary context mapped
- **SECOND_CAS**: Case 3.a. primary context mapped
- **SECOND_CAS (PROXY)**: Case 3.a. secondary context mapped
- **FLOWNAME_MANUAL_PROXY**: Case 3.b. primary context mapped
- **FLOWNAME_MANUAL_PROXY (PROXY)**: Case 3.b. secondary context mapped
- **NO_MAPPING**: Case 4.

Conversion factors

Conversion factors are needed in two cases, namely: i. if the units of the source and the target flows differ, and ii. if the units are the same but the LCIA-relevant properties or the contents of the source and the target flows differ. Examples for the latter case are energy carriers expressed by mass (in kilogram, kg). One kg of coal with an energy content 15 MJ/kg is not the same as one kg of coal with an energy content of 20 MJ/kg. In order to convert from kg of the former coal type to kg of the latter, both energy contents need to be considered. In this case, the common LCIA-relevant “currency” unit is MJ. The following rules are applied for calculation of conversion factors (also displayed in Figure 2):

- If the source unit differs from the target unit and...
 - ...if the target unit equals the currency unit, then use the source property for calculation of the conversion factor.
 - ...if the source unit equals the currency unit, then use the target property for calculation of the conversion factor.
- If the source unit and the target unit both differ from the currency unit, use source and target properties for calculation of the conversion factor.
- If needed properties are missing on either source or target side,
 - use the properties on the other side if available,
 - otherwise, do not map.

¹¹ The string “(PROXY)” indicates that the context of an elementary flow in the source list matched a proxy sub-compartment in the target list (i.e., the context is mapped by PROXY_CONTEXT_MATCH).

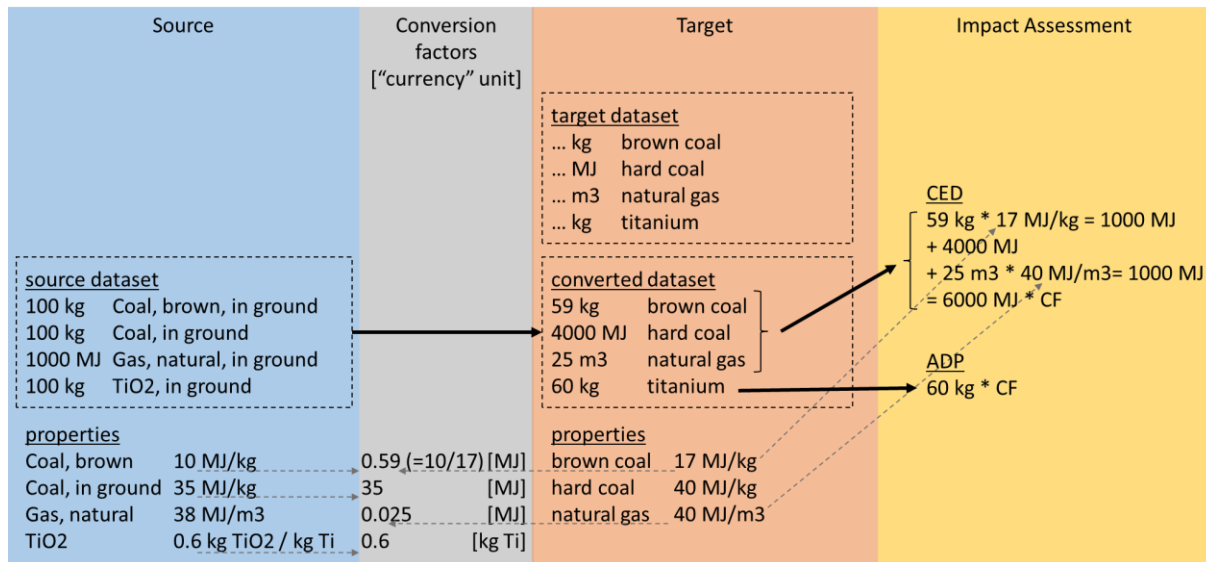


Figure 2: Example for calculation of conversion factors; CED = Cumulative Energy Demand, ADP = Abiotic Depletion Potential

Review of proposed matches and mapping results

Throughout the project activities, both manually defined and autogenerated matches between contexts, flowables, or flows were evaluated using a common scheme. The scheme consists of two indicators/criteria, intended to separate the logical match condition (i.e., the objective alignment of scope/boundaries between the source- and target-side) and mappers' perceived confidence in the match. Symbols of logical operators, such as =, ~, <, >, and <>, are used to indicate the match condition, and a three-level (A; B; C) rating system was defined to enable the source- and target-side to indicate their respective degree of confidence in each match:

- The mapping is considered (near) ideal, and hence fully acceptable.
- The mapping is reasonable, and judged acceptable, but not optimal (for example, slightly different contexts or less-than perfectly matched flowables).
- The mapping is not acceptable. Further action(s) is/are needed.

The scheme for rating the matches found in the GLAD EF mapped files, as well as the procedure to determine any further actions, is described in Appendix C of the Maintenance Manual (project deliverable D 2.3).

Common mapping issues

In this part of the report, we collected common issues encountered, as well as the solutions implemented, during the mapping activities in the project. The first section is dedicated to challenges related to mapping of flow contexts (i.e., only environmental compartments). The next section addresses common mapping issues related primarily to flowables without the context information, i.e., disregarding the environmental compartment of the elementary flow. The third and final section concerns flows, i.e., the combination of flowable and context. The headings of the sub-sections are listed here for a better overview and easy access:

- Issues related to matching of contexts (flowable-independent)
 - Compartments for long-term emissions
 - Matching of (partially) overlapping contexts
 - Proxy contexts for ‘unspecified’ and indoor air sub-compartments
- Issues related to matching of flowable (context-independent)
 - Specific to less specific flowables
 - Mapping of closely ‘related’ chemical substances
 - Matching the main greenhouse gases
 - Mapping of elements vs. ions vs. compounds (unspecified)
 - Flowable approximations (proxies)
- Challenges related to matching of flows (flowable+context combinations)
 - Mapping of water flows
 - Other aspects: missing characterization factors

Issues related to matching of contexts (flowable-independent)

Compartments for long-term emissions

Some LCI databases and LCIA methods differentiate between emissions released immediately or in the near-term and those that (are expected) to occur in a more distant future. The border for this temporal differentiation is typically set to around 100 years after the reference year (or start date) of an activity.

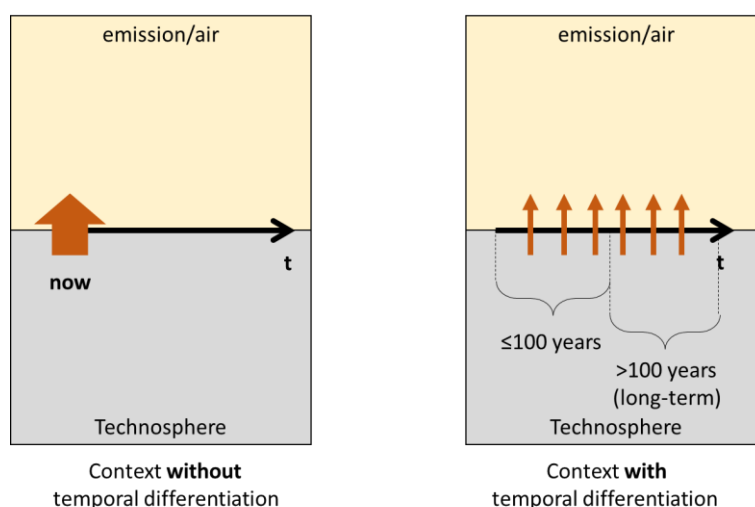


Figure 3 – elementary flow lists without (left) and with (right) temporal differentiation.

The elementary flow lists for the ecoinvent database and the ILCD-EF provide dedicated environmental sub-compartments for long-term emission, whereas Federal LCA Commons and IDEA do not distinguish these emissions separately. The project team agreed that emissions assigned to

long-term sub-compartments should still be mapped to ‘time-indifferent’ flow lists. In contrast, the long-term sub-compartments should be omitted from the list of proxy contexts to avoid paths or temporal discounting not intended for the original dataset.

<i>The challenge</i>	<i>Solution implemented</i>	<i>Potential implications in format conversion</i>
How to map ‘long-term’ contexts [on the source-side] to EF lists without dedicated sub-compartments for this?	The most unspecified sub-compartment on the target-side is set as default match, and further sub-compartments can be listed as acceptable proxies.	The temporal differentiation in the original datasets will be lost during format conversion.
Should ‘long-term’ contexts [on the target-side] be considered among the proxies for (time-indifferent) sub-compartments?	No, not to be considered as proxies to non-long-term contexts.	None expected, if the target-side EF list contains further sub-compartments (besides ‘long-term’) for the flowable in question.

Examples of the context mapping for long-term emissions (on the source-side) onto EF lists without temporal differentiation are provided in the table below.

<i>SourceFlow list</i>	<i>SourceContext</i>	<i>TargetFlow list</i>	<i>TargetContext</i>
ecoinvent EF v3.7	air/low population density, long-term	IDEA v2.3	Emissions/air/unspecified
ecoinvent EF v3.7	water/ground-, long-term	IDEA v2.3	Emissions/water/unspecified
ILCD-EF 3.0	air/low population density, long-term	FEDEF v1.0.3	emission/air
ILCD-EF 3.0	water/ground-, long-term	FEDEF v1.0.3	emission/water

Matching of (partially) overlapping contexts

The scope of the sub-compartments used to describe/define the intersection between the technosphere and the environment for elementary flows differs between the participating EF lists. In some cases, a higher level of disaggregation in one list still fits perfectly within the more aggregated contexts available in another list.

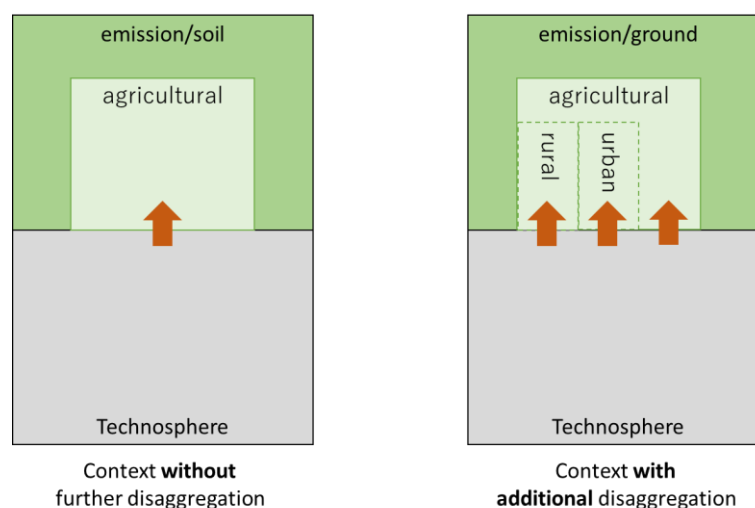


Figure 4 – emissions to agricultural soil (ground) as example of different levels of disaggregation with perfect ‘fit’.

For example, all three contexts for emissions to agricultural soil in Federal LCA Commons’ list align well to a single corresponding context in the ILCD-EF list, see Table 3. The tilde symbol (~) indicates

that, while not strictly equal or identical, the mapping between ‘emission/ground/human-dominated/agricultural/rural’ and ‘emission/soil/agricultural’ can be considered equivalent and is rated as an ideal match (i.e., assigned a confidence rating “A”, despite not being strictly identical in scope).

Table 2 – comparison of contexts for emissions to agricultural soil available in Federal LCA Commons and the Environmental Footprint EF lists, respectively.

Source-side context Federal LCA Commons	Match condition	Target-side context ILCD-EF	Match priority
emission/ground/human-dominated/agricultural	=	emission/soil/agricultural	default
emission/ground/human-dominated/agricultural/rural	~	emission/soil/agricultural	default
emission/ground/human-dominated/agricultural/urban	<	emission/soil/agricultural	default
Source-side context ILCD-EF	Match condition	Target-side context Federal LCA Commons	Match priority
emission/soil/agricultural	=	emission/ground/human-dominated/agricultural	default
emission/soil/agricultural	~	emission/ground/human-dominated/agricultural/rural	proxy (1 st priority)
emission/soil/agricultural	>	emission/ground/human-dominated/agricultural/urban	proxy, (2 nd priority)
emission/soil/agricultural	<	emission/ground/human-dominated	proxy (3 rd priority)
emission/soil/agricultural	<	emission/ground	proxy (4 th priority)

Partially overlapping contexts pose a bigger challenge to map consistently. For example, both the ILCD-EF and ecoinvent’s EF lists contain the sub-compartment ‘lower stratosphere and upper troposphere’ for emissions to air at high altitudes (e.g., from aviation at cruising altitude). This context is deemed most closely related in Federal LCA Commons is ‘emission/air/stratosphere’. The latter is not strictly falling within the scope of the former, but it is here considered equivalent and more suitable than the closest alternatives, such as the FEDEF context ‘emission/air/troposphere/very high’ (intended to be used for stacks and chimneys higher than 150 meters)¹² or just ‘emission/air’.

¹² Edelen, A., T. Hottle, S. Cashman, W. Ingwersen (2019) *Federal LCA Commons Elementary Flow List: Background, Approach, Description and Recommendations for Use*. U.S. Environmental Protection Agency (U.S. EPA).

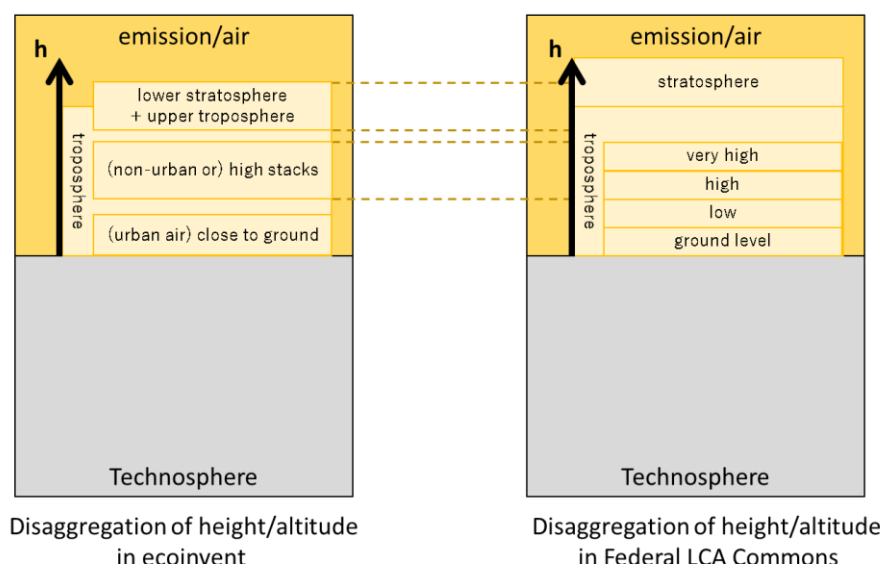


Figure 5 – Schematic comparison of the contexts for emissions to air in the elementary flow lists of ecoinvent and Federal LCA Commons, with sub-compartments based on the height/altitude of the emission source.

A related issue concerns whether target-side contexts that do not overlap at all with the source-side context to be mapped should be considered among the acceptable proxies. A key question here is whether a non-overlapping sub-compartment can be considered a better approximation to the source context than the corresponding ‘unspecified’ context. This might, for example, be the case for air emission at ground level vs. at low elevation, whereas industrial, forest, or non-agricultural soil might not be considered preferable over the ‘unspecified’ sub-compartments for mapping emissions to agricultural soil (on the source-side).

<i>The challenge</i>	<i>Solution implemented</i>	<i>Potential implications in format conversion</i>
How to choose the most appropriate target-side match among multiple contexts (sub-compartments) that fit within the source-side context?	We strive to identify the target-side context most aligned (in terms of physical scope) with the source-side context. Consistency of LCIA characterization factors might be considered on a case-by-case basis.	Inconsistent LCIA scores may result if characterization factors differ among more disaggregated sub-compartments
Should contexts representing non-overlapping sub-compartments be considered acceptable proxies?	Yes, but to be decided on a case-by-case basis. Priority should be given to proxy contexts expected to be assigned the same or as similar characterization factors as possible.	Mapping decision might affect LCIA score consistency

Proxy contexts for ‘unspecified’ and indoor air sub-compartments

Whether or not to use more specific contexts as proxies for ‘unspecified’ sub-compartments would be irrelevant if all flowables were – as a minimum – defined for ‘unspecified’, besides any further contexts, across all EF lists. This is not always the case in practice. At the same time, it is not uncommon that one or more specific contexts are assigned the same characterization factors as the ‘unspecified’ (and vice versa) during the implementation of LCIA methods, and hence treated as equivalent (for the lack of better options or more information). In this project, we therefore agree to list further contexts as acceptable proxies to ‘unspecified’ and to prioritize less specific over more

specific sub-compartments. The sub-compartment ‘non-urban air or from high stacks’ in the flow lists ofecoinvent and Environmental Footprint is consequently prioritized over ‘urban air close to ground’ as a proxy to unspecified air emissions.

The contexts for emissions to indoor air are at the other end of the spectrum, compared to the ‘unspecified’ sub-compartment. With the assumption that fate and effect factors for human health impacts from indoor air emissions are distinctly different from other sub-compartments, we propose to only map this context either to the equivalent target-side context (i.e., indoor to indoor) or ‘unspecified’ (as the only proxy, if maintaining mass/energy balances over the datasets is prioritized).

<i>The challenge</i>	<i>Solution implemented</i>	<i>Potential implications in format conversion</i>
Should ‘unspecified’ contexts on the source side be mapped to contexts representing more specific (i.e., smaller) sub-compartments?	Yes, but to be decided on a case-by-case basis. Priority should be given to proxy contexts expected to be assigned the same or as similar characterization factors as ‘unspecified’.	Mapping decision might affect LCIA score consistency
How to map contexts related to indoor air emissions?	Source-side contexts for indoor air emissions are only mapped to contexts on the target-side representing either indoor (default) or completely unspecified (as proxy). IDEA on the target side constitutes an exception, onto which indoor air emissions are not mapped	Tradeoff between completeness of mapping (e.g., to maintain the mass balance of inventory) and LCIA score consistency (expected to decrease for LCIA methods related to human health impacts).

Issues related to matching of flowable (context-independent)

Despite the title of this section, we do in some cases need to consider the main compartments (i.e., air, ground/soil, and water) when trying to establish the most suitable match between flowables.

Specific to less specific flowables

Flowables representing groups or “totals” of, for example, chemicals can provide suitable matches to a wide range of more specific flowables, e.g., the individual chemical substances. This is a form of aggregation that results in a loss of information, as the mapping cannot be readily reversed (without a separate log/record of the original flows). The characterization of the more unspecified flowables in impact assessment is generally associated with higher uncertainty. Equivalents, such as the (human) *toxicity equivalency factor (TEF)* for dioxins, based on a reference substance (2,3,7,8-tetrachlorodibenzo-*p*-dioxin, TCDD),¹³ work well for elementary flows contributing to single or a few distinct environmental problems.

¹³ Van den Berg M, Birnbaum LS, Denison M, De Vito M, Farland W, Feeley M, et al. (2006). The 2005 World Health Organization reevaluation of human and Mammalian toxic equivalency factors for dioxins and dioxin-like compounds. *Toxicological Sciences*. **93** (2): 223–241. DOI: [10.1093/toxsci/kfl055](https://doi.org/10.1093/toxsci/kfl055)

<i>The challenge</i>	<i>Solution implemented</i>	<i>Potential implications in format conversion</i>
How to map overlapping flowables (of different scope)?	These matches are considered (near) ideal, even though information may be lost in the aggregation of flowables, constituting logical sub-sets of the target-side match.	Conversion factors, e.g., based on clearly defined reference substances (for equivalency), might be required for LCIA score consistency
Should unspecified flowables, including those representing groups or totals, be mapped to more specific flowables (like individual substances)?	We generally advise against this, unless a well-accepted and widely applied reference substance is available, e.g., 2,3,7,8-TCDD for dioxins	Unspecified flowables are dropped during dataset format conversion. The relevance should be assessed on a case-by-case basis, and justified accordingly

The table below contains several concrete examples illustrating how flowables of different scope or degree of specificity were mapped in the present project.

<i>Source flow list</i>	<i>Source flowable</i>	<i>Target flow list</i>	<i>Target flowable</i>
ILCD-EF 3.0	Particles (PM2.5)	FEDEFL v1.0.3	particulate matter, ≤ 2.5µm
	Particles (PM0.2 - PM2.5)		
	Particles (PM0.2)		
FEDEFL v1.0.3	particulate matter, ≤ 2.5µm	ILCD-EF 3.0	Particles (PM2.5)
FEDEFL v1.0.3	particulate matter	ILCD-EF 3.0	N/A
IDEA v2.3	copper salts(water-soluble, except complex salts)	FEDEFL v1.0.3	Copper compounds
IDEA v2.3	N-butoxymethyl-2-chloro-2',6'-diethylacetanilide	ecoinvent EF v3.7	Herbicides, unspecified
IDEA v2.3	N-butoxymethyl-2-chloro-2',6'-diethylacetanilide	FEDEFL v1.0.3	Butachlor
ecoinvent EF v3.7	Herbicides, unspecified	IDEA v2.3	N/A
ecoinvent EF v3.7	Herbicides, unspecified	FEDEFL v1.0.3	Herbicides

Mapping of closely 'related' chemical substances

As mentioned above, chemical substances might be associated with one or multiple identifiers under the same system, most notably the CAS Registry Number. This is further complicated by the fact that some chemicals sharing molecular formula might have the functional group at different locations (isomers) within the molecules or representing mirrored – and distinguishable as such – but otherwise identical versions (enantiomers) of the same molecule.

<i>The challenge</i>	<i>Solution implemented</i>	<i>Potential implications in format conversion</i>
Whether to map isomers or enantiomers?	It is generally assumed justified to match these closely related molecules.	The actual equivalence, e.g., in terms of toxicity, should be investigated based on relevance to the results. It is also recommended to check whether the LCIA method in question provides different characterization factors for the substances in question.

Some examples of mapping between isomers and enantiomers are provided in the table below:

Source flow list	Source flowable	Source CAS	Target flow list	Target flowable	Target CAS
ecoinvent EF v3.7	Butanol	71-36-3	FEDEFL v1.0.3	1-Butanol	71-36-3
FEDEFL v1.0.3	1-Butanol	71-36-3	ecoinvent EF v3.7	Butanol	71-36-3
FEDEFL v1.0.3	2-Butanol	78-92-2	ecoinvent EF v3.7	Butanol	71-36-3
ILCD-EF 3.0	2-hydroxypropanoic acid	10326-41-7	ecoinvent EF v3.7	Lactic acid	50-21-5
	Lactic acid	50-21-5			
	L-(+)-lactic acid	79-33-4			
ecoinvent EF v3.7	Lactic acid	50-21-5	ILCD-EF 3.0	Lactic acid	50-21-5

Matching the main greenhouse gases

The main greenhouse gases (GHGs, i.e., carbon dioxide/monoxide, CO₂/CO;¹⁴ methane, CH₄; nitrous oxide N₂O; sulfur hexafluoride, SF₆; and hydro-/chlorofluorocarbons, HFC/CFCs) arguably belong to the most important elementary exchanges in LCA. For carbon dioxide/monoxide and methane, the characterization factors assigned (based on the gases' respective global warming potentials, GWPs) to these gases typically differs depending on whether the source of the carbon is considered of fossil (i.e., non-renewable) or biogenic (renewable) origin. Changes in the long-term carbon balance of soil or biomass stocks, e.g., because of altered land use/management, are in this context often considered separately and characterized as 'fossil'. (Recent developments in LCA have also introduced further temporal differentiation by characterization of carbon uptake and release in long-rotation biomass resources, such as wood from boreal forests).

The elementary flowables – both as resources and emissions – for carbon dioxide, methane, and carbon monoxide available in EF lists of the four participating databases are listed next. The mapping applied to these flowables is summarized in the tables below, and the potential consequences or limitations of this mapping are discussed at the end of this section.

- Resource flowables (for flows from the environment)
 - ecoinvent EF v3.6/3.7
 - Carbon dioxide: 'in air' – *carbon content considered 100% non-fossil, and CF=0 for GWP (in own LCIA method implementation)*
 - EF3.0
 - Carbon dioxide: biogenic; biogenic-100yr; land use change; fossil
 - FEDEFLv1.0.3
 - Carbon dioxide: [unspecified]
 - Methane: [unspecified]
 - IDEAv2.3
 - Carbon dioxide: [unspecified]
- Emission flowables (for flows to the environment)
 - ecoinvent EF v3.6/3.7
 - Carbon dioxide: fossil; non-fossil; from soil or biomass stock; to soil or biomass stock
 - Methane: fossil; non-fossil; from soil or biomass stock; [unspecified]
 - Carbon monoxide: fossil; non-fossil; from soil or biomass stock

¹⁴ Carbon monoxide (CO) is technically not a greenhouse gas. But emissions of CO to the atmosphere may subsequently oxidize to CO₂, and hence commonly considered to contribute to global warming in LCA.

- EF3.0
 - Carbon dioxide: fossil; biogenic; land use change; [unspecified]
 - Methane: fossil; biogenic; land use change; [unspecified]
 - Carbon monoxide: fossil; biogenic; land use change; [unspecified]
- FEDEFLv1.0.3
 - Carbon dioxide: [unspecified]
 - Methane: [unspecified]
 - Carbon monoxide: [unspecified]
- IDEAv2.3
 - Carbon dioxide: fossil; biogenic; [unspecified]
 - Methane: fossil; biogenic; [unspecified]
 - Carbon monoxide: [unspecified]

Mapping of carbon dioxide, CO₂, as resource (for flows from the environment)

[CO ₂ as resource]	Target list	ei EF v3.7	ILCD-EF 3.0				FEDEFL v1.0.3	IDEA v2.3
Source list	Carbon source	[biogenic]	fossil	bio	bio-100yr	land use change	[unspec.]	[unspec.]
ei EF v3.7	'in air' [biogenic]		X	O	X	X	X'	X'
ILCD-EF 3.0	fossil biogenic biogenic-100yr land use change	X O O X					O X' X' O	X' X' X' X'
FEDEFL v1.0.3	[unspecified]	X'	O	X'	X'	O		X'
IDEA v2.3	[unspecified]	X'	X'	X'	X'	X'	X'	

O: mapped; X: not mapped; X': considered equivalent, but not mapped to avoid potential carbon imbalances across life cycles

Mapping of emissions of carbon dioxide, CO₂ (for flows to the environment)

[CO ₂]	Target	eiEF v3.7				ILCD-EF 3.0				FEDEFLv1.0.3	IDEA v2.3		
Source	Carbon source	fossil	bio	from soil	to soil	fossil	bio	land use	[unspec.]	[unspec.]	fossil	bio	[unspec.]
ei EF v3.7	fossil non-fossil (biogenic) from soil or biomass stock (land use change) to soil or biomass stock (land use change)					O X X X	X O X X	X X X X	X X O O	O X' O O	O X X X	X X' X X	X X O X
ILCD-EF 3.0	fossil biogenic land use change [unspecified]*	O X X X	X O X X	X X O X	(o) (o) X (o)					O X' O O	O X X X	X X' X X	X X O X
FEDEFL v1.0.3	[unspecified]	O	X'	X	(o)	O	X'	X	O		X	X	O
IDEA v2.3	fossil biogenic [unspecified]	O X O	X X' X	X X X	X X X	O X O	X X' X	X X X	X X X	O X' O			

*O: mapped; X: not mapped; X': considered equivalent, but not mapped to avoid potential carbon imbalances across life cycles; *: only for emissions to water*

Mapping of emissions of methane, CH₄ (for flows to the environment)

[CH ₄]	Target	ei EF v3.7				ILCD-EF 3.0				FEDEFL v1.0.3		IDEA v2.3		
Source	Carbon source	fossil	bio	from soil	[unspec.]	fossil	bio	land use	[unspec.]	[unspec.]	[unspec.]	fossil	bio	[unspec.]
ei EF v3.7	fossil non-fossil (biogenic) from soil or biomass stock (land use change) [unspecified]					O X X O	X O X X	X X O X	X X X	O O O		O X X X	X O X X	X X O O
ILCD-EF 3.0	fossil biogenic land use change [unspecified]	O X X O	X O X X	X X O X	X X X X					O O O O		O X X X	X O X X	X X O O
FEDEFL v1.0.3	[unspecified]	O	X	X	X	O	X	X	X			X	X	O
IDEA v2.3	fossil biogenic [unspecified]	O X O	X O X	X X X	X X X	O X O	X O X	X X X	X X X	O O O				

O: mapped; X: not mapped

Mapping of emissions of carbon monoxide, CO (for flows to the environment)

[CO]	Target	eiEF v3.7				ILCD-EF 3.0				FEDEFL v1.0.3		IDEAv2.3	
Source	Carbon source	fossil	bio	from soil	[unspec.]	fossil	bio	land use	[unspec.]	[unspec.]	[unspec.]	[unspec.]	[unspec.]
eiEF v3.7	fossil non-fossil (biogenic) from soil or biomass stock (land use change)					O X X	X O X	X X O	X X X	O O O		O O O	O O O
ILCD-EF 3.0	fossil biogenic land use change [unspecified]	O X X O	X O X X	X X O X	X X X X					O O O O		O O O O	O O O O
FEDEFL v1.0.3	[unspecified]	O	X	X	X	O	X	X	X				O
IDEAv2.3	[unspecified]	O	X	X	X	O	X	X	X	O			

O: mapped; X: not mapped

Establishing a consistent mapping of GHG is not only challenging due to the different approaches taken in the EF lists of the participating databases, but also further complicated by how LCIA methods might subsequently be applied to these flows. Alternative ways of representing and characterizing flows for uptake and release of GHGs can still produce the same net result. The tables above should be considered an attempt at bringing clarity to what the different EF lists contain and how these items align across the lists. The expected consequences of the proposed mapping, in terms of LCIA consistency during format conversion, need to be assessed more carefully in the future, e.g., as part of the GLAD-GLAM dialogue, to evaluate whether the mapping and/or LCIA implementation can be improved.

Two cases, proving especially challenging during the project, warrants special consideration: firstly, the ILCD-EF flowlist defines four flowables for the CO₂ as a resource in air, whereas the other list

contains only one CO₂ resource flow each. The fossil CO₂ resource in air in ILCD-EF is mainly used when the origin of the CO₂ input is known, e.g., for CO₂-enriched in greenhouses.

The second aspect, in contrast, concerns the different approaches used to account for the uptake of carbon dioxide from the atmosphere (i.e., the resource flow) in combination with the release of “short-rotation” carbon dioxide back to the atmosphere (i.e., emissions of “biogenic” or “non-fossil” carbon dioxide). Two of the participating database systems (ecoinvent and ILCD-EF) strive to balance biogenic carbon explicitly across the datasets and product life cycles.¹⁵ IDEA’s flow list also contains a flow specifically for biogenic CO₂ emissions to air, but uptake from the atmosphere is not accounted for systematically in the database. This approach rests on the assumption that biogenic CO₂ flows are not characterized in impact assessment, i.e., that the characterization factors applied for both uptake and release of biogenic CO₂ are zero (0).

Finally, the flow list of the Federal LCA Commons does not differentiate between carbon origins for the various GHG emissions. Treating CO₂ emissions of fossil and biogenic origins as “equal”, through one common flow, does not alter the impact on global warming during format conversion, as long as the uptake of CO₂ from the atmosphere (i.e., for the share of carbon otherwise considered biogenic) is also assigned a characterization factor of -1 in subsequent impact assessment. But an inconsistency would be introduced if both uptake and release of biogenic/non-fossil CO₂ were to be considered ‘climate-neutral’ and consequently assigned a characterization factor of 0 for global warming potential, or when carbon uptake is not included in the LCI.

Rather than trying to anticipate all possible scenarios for the feasible combinations of source- and target-side flow lists, as well the LCIA implementation (i.e., GWP characterization factors of [0; 0] vs. [-1; 1]), the project team ultimately opted for a conservative approach: this meant that neither (unspecified) uptake CO₂ from the atmosphere nor emissions of biogenic CO₂ to air were mapped, despite potentially yielding consistent results (if biogenic carbon is balanced across product life cycles). One main exception was made for mapping between ecoinvent and ILCD-EF, as the risk of distortions during format conversion in this case was considered rather low:

- Uptake: ‘carbon dioxide, in air’ ↔ ‘carbon dioxide (biogenic)’
- Release: ‘carbon dioxide, non-fossil’ ↔ ‘carbon dioxide (biogenic)’

¹⁵ Carbon balances across product life cycles can still get distorted through database linking, e.g., due to co-/by-product allocation or cut-offs without carbon corrections.

Mapping of elements vs. ions vs. compounds (unspecified)

Emissions of elements especially various metals to the environment are also known to have significant toxicity impacts. However, these elementary flows are implemented in various forms and can be roughly categorized as “ions with values” (specific oxidation states), “ion without values” (unspecified oxidation states), “element”, “element compounds”, and “element and element compounds” (see table below).

Notation	Meaning	Examples
<i>ion(x)</i>	<i>Element ions with oxidation state</i>	<i>copper(iii)</i>
<i>ion</i>	<i>Element ions without oxidation state</i>	<i>Copper, ion</i>
<i>element</i>	<i>Element (name) only</i>	<i>Copper</i>
<i>compounds</i>	<i>Unspecified compounds of an element</i>	<i>Copper compounds</i>
<i>elements and compounds</i>	<i>Element and element compounds</i>	<i>Cobalt and its compounds</i>

The participating flow lists do not define these flowables consistently across the lists, and as a result, a wide variety of mapping ratings were obtained in the first iteration of the flowable mapping exercise. For the sake of improving mapping consistency, and to align within the project team on how to rate the ‘quality’ of these mappings, there was the need to define a uniform *ConfidenceLevel* and *MatchCondition* when different states/forms of a particular element were mapped. We therefore adopted the following rules and applied them uniformly when mapping flowables related to elements, ions, and/or compounds.

ConfidenceLevel						
		target				
	Mapping	ion(x)	ion	element	compounds	element and compounds
source	ion(x)	A	B	B	B	B
	ion	B	A	B	B	B
	element	B	B	A	B	B
	compounds	B	B	B	A	A
	element and compounds	B	B	B	A	A
MatchCondition						
		target				
	Mapping	ion(x)	ion	element	compounds	element and compounds
source	ion(x)	=	<	<	<	<
	ion	>	=	~	<	<
	element	>	~	=	<	<
	compounds	>	>	>	=	~
	element and compounds	>	>	>	~	=

It is important to note that we agreed within the project team, NOT to map any chemical substances that could possibly be included as one of the element compounds (e.g., “copper(ii) oxide” to “Copper compounds”). Mapping all substances with the principal element in its molecule to a corresponding unspecified flowable, like “copper compounds”, might have increased mapping coverage further. But the project team opted against this step, motivated by the great uncertainties (and information loss) involved, as well as the unresolved question about whether and how to map substances containing more than one such principle element.

Flowable approximations (proxies)

It is difficult to determine strict criteria for the logical alignment, ‘overlap’, or similarity for matches to be considered reasonable approximations (or ‘proxies’) to the source-side flow. The subjective confidence in this type of mapping, an indicator of the readiness to accept less-than-ideal matches, also tends to involve large variations between the mappers and reviewers. This ‘grey zone’ basically reflects the more general tradeoff between prioritizing the completeness vs. the accuracy/precision the of mapping. That is, proxies enable a larger portion of the source- and target-side flows to be mapped against each other, but that increase of mapping coverage comes at the expense of more uncertain or less precise matches being considered acceptable. This tradeoff is illustrated in Figure 6.

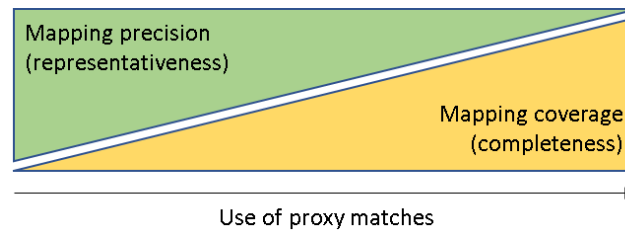


Figure 6 – tradeoff between the precision (representativeness) and coverage (completeness) in EF mapping as a function of the (degree of) use of proxy matches.

In addition to the examples provided in the previous sections, and without being exhaustive, the table below presents a summary of further challenges related to the use and representativeness of proxies in the EF mapped files.

<i>The challenge</i>	<i>Solution implemented</i>	<i>Potential implications in format conversion</i>
How to rate the confidence level when mapping to more generic/proxy flows?	It is generally assumed justified to map to flowables of a larger scope, and that such a match may even be considered (near-) ‘ideal’. This is especially when the target-side flow list has been established using more aggregated flows.	The actual representativeness or equivalence, e.g., in terms of toxicity, should be investigated based on relevance to the overall results. Information will be lost during format conversion.
Whether/how to map unconventional oil (petroleum) resources, e.g., ‘oil sand’ and ‘oil shale’?	The project team agreed to not consider these unconventional resources equivalent to crude petroleum oil. These items are consequently not mapped unless available as such in the target flow list.	LCIA scores for impact categories related to fossil/abiotic resource depletion or energy/exergy demand will not be consistent over format conversion.
Whether/how to map coal bed methane and mine off-gas to ‘conventional’ natural gas (as a resource)?	These resource flows are considered largely equivalent to natural gas, and hence mapped accordingly.	Results for fossil resource depletion and related impact categories will be consistent over format conversion.
Whether/how to map generic metal ores, e.g., “aluminium ore”?	These generic ores are mainly found in the list of Federal LCA Commons. They are not considered preferable and hence not recommended to be used for inventory modelling. It was decided not to map to/from these flows.	No major implications are expected, as long as data providers avoid using these generic metal ores.

<i>The challenge</i>	<i>Solution implemented</i>	<i>Potential implications in format conversion</i>
How to prioritize between multiple potential proxies?	As a general rule, we recommend that the proxy that is most closely related or otherwise with the narrowest (overlapping) scope to be prioritized. Unspecified pesticides, insecticides, etc. were consequently preferred over flowables representing groups of hydrocarbons (e.g., aliphatic, aromatic, or unspecified).	We here strive to minimize the loss of information resulting from the mapping step. But it should be noted that the choice among more-or-less unspecified or generic flowables may carry little practical relevance, if none of the options is assigned a characterization factor in LCIA.

Challenges related to matching of flows (flowable+context combinations)

Several aspects and challenges related to the mapping of flows (as unique combinations of flowables and contexts) have already been addressed in the mapping methodology part (e.g., in sections *Mapping of flows* and *Conversion factors*). This included mapping of land transformation and use. This section is primarily dedicated to explaining the list features and approaches taken in this project to map water flows across the participating flow lists. The section finishes, though, with a brief account of *Other aspects: missing characterization factors*.

Mapping of water flows

Water resources are an important impact area and therefore show a wide variety of elementary flows to express the conceptual water model that each EF system adopts. Shown in Figure 7 is the common basic model of water to roughly structure elementary flows related to water. Please note that the objective of this diagram is to illustrate the water-related elementary flows only, and all intermediate flows (exchanged between activities in within the technosphere), such as industrial water, irrigation water, sewage/wastewater as well as water embedded into products, are not included into this diagram.

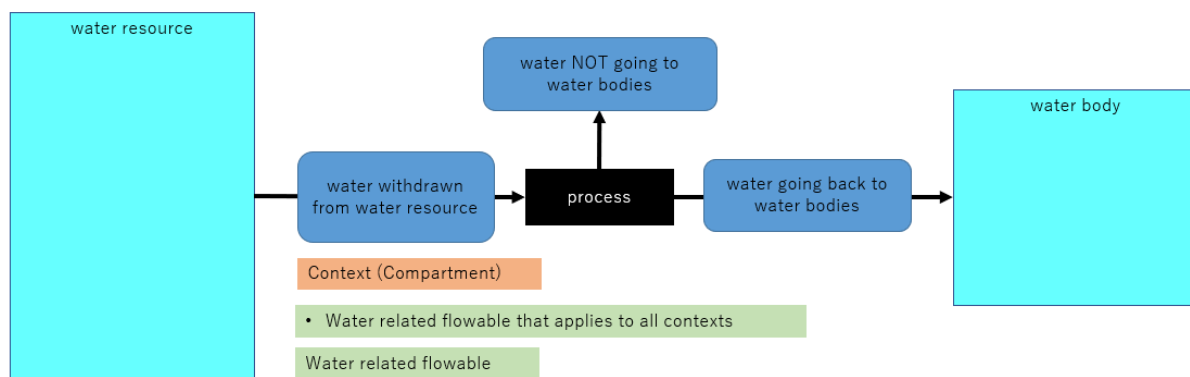


Figure 7 – The basic water model to roughly group water related elementary flows

As shown in Figure 7, we group water flows from the biosphere to a technosphere process into “water withdrawn from water resource”, and these flows express water extracted from a natural “water resource” such as rivers, lakes, underground water or precipitation. Water flows from an economic activity or process back to the biosphere can roughly be separated into two groups: “water going back to water bodies” and “water NOT going to water bodies”. An example of the former would be water used in a factory and emitted to water bodies such as rivers and lakes. In contrast, water that evaporates in the process (i.e., released to the atmosphere) or is emitted to the ground (soil) and will not be considered returned to a water body.

In addition to the basic water model, each EF system has additional granularity from certain aspects that provides further details of the water flow. The main aspects considered are summarized in the following table and illustrated in Figure 8 below.

Aspect	Explanation
surface situation	Distinguish the water resource and bodies with keywords such as “river”, “lake” or group these as “surface”. Some ground emissions are separated to describe the characteristic to describe in detail where the water was released with keywords such as “urban”, “rural”, “terrestrial” etc.
vertical height	Distinguish the vertical height of where the water is emitted as is done with other chemical substances emitted into the air.
rainwater	Distinguish withdrawal and emission of water flows that are of “rainwater” origin.
ground water	Distinguish water from the ground with keywords such as “well”, “ground water”, “subterranean”. In some cases, moisture in the ground is distinguished from water that is from underground water bodies.
origin	Distinguish water flows with the origin of the water with tags such as “natural”, “surface”, “ground”, “rain”, “green”.
usage	Distinguish water flows with the usage of the water such as “turbine”, “cooling”.
time	Distinguish water flows with time-related aspects such as “long-term”, “fossil”.
water salinity	Distinguish water flows by the salinity of the water that is under consideration.
water body salinity	Distinguish water flows by the salinity of the water body in relation (emission and withdrawal) to the water that is under consideration.

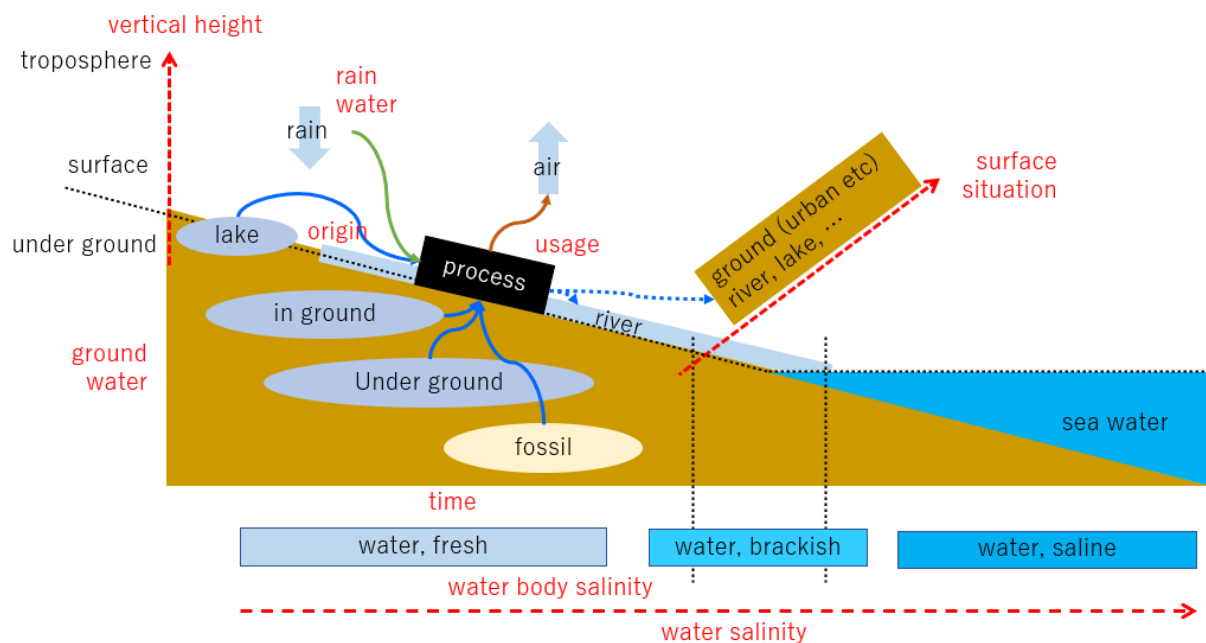


Figure 8 – illustration of aspects and attributes of elementary flows for water considered during mapping.

We next show how each EF list implements the elementary flows to model water starting withecoinvent.

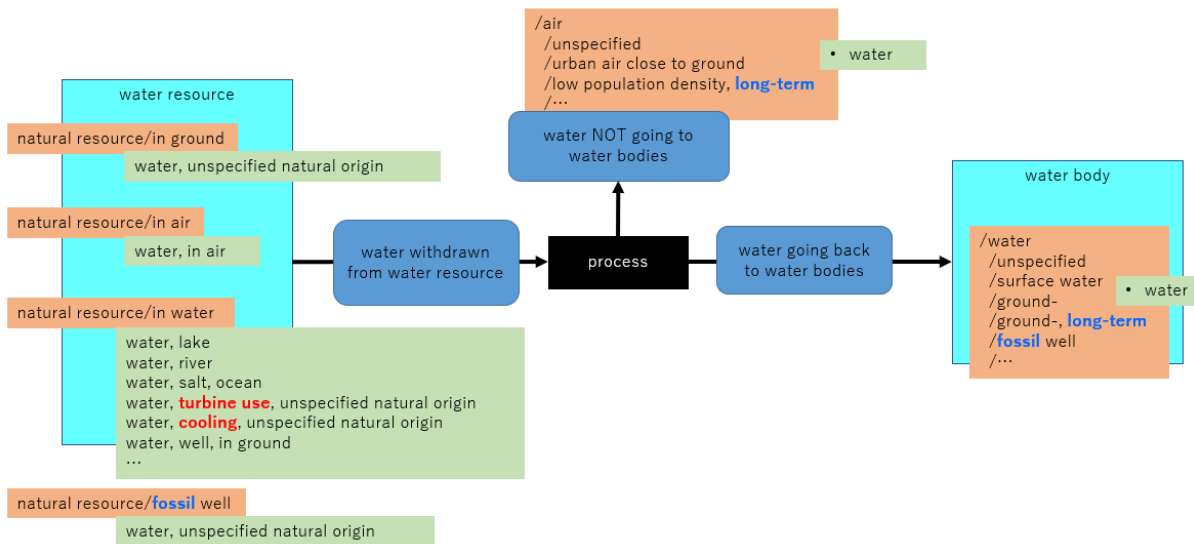


Figure 9 – The water-related elementary flows in ecoinvent's flow list projected to the basic water model

The flow list from ecoinvent is rather straightforward in term of how it cover water-related elementary flows. Rainwater is not explicitly mentioned, and withdrawal of rainwater is expressed as “water, in air”. Special features of this EF list are that it includes usages of water such as “turbine”, “cooling” in its flowable names. Temporal aspects such as “long-term” and “fossil” are distinguished at the context level.

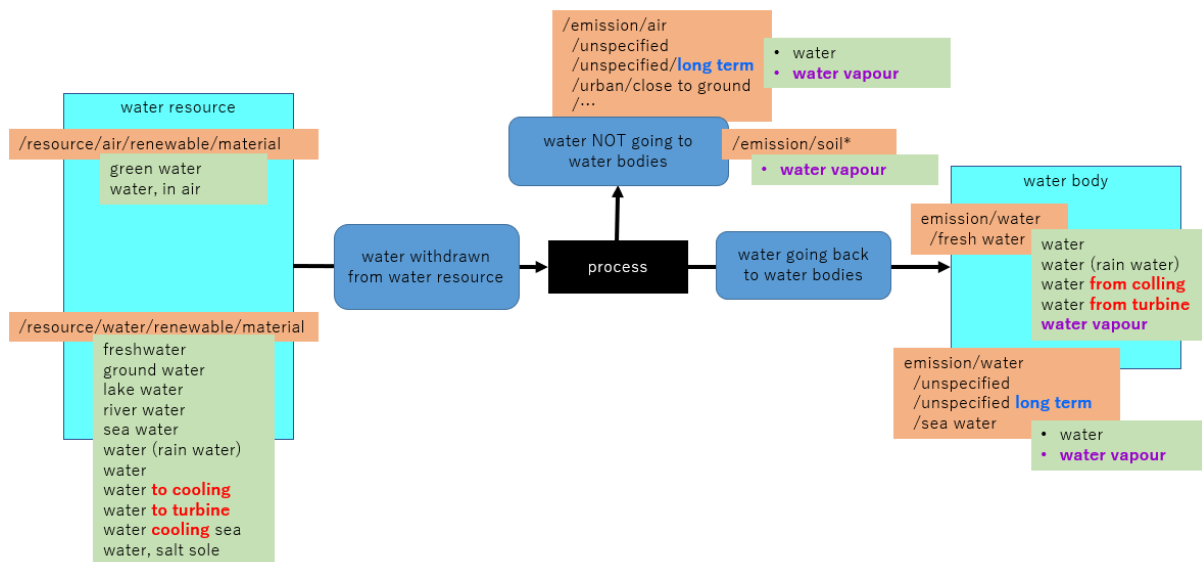


Figure 10 – ILCD-EF 3.0's water-related elementary flows projected to the basic water model

ILCD-EF 3.0 is similar to ecoinvent, but rainwater is distinguished explicitly in the flowable name “(rain water)”, both as a resource and an emission to water. Special features are that it also includes usages of water such as “cooling” and “turbine”. For time-related aspects, “long term” emissions to air and water are distinguished at the context level. One unique feature is that ILCD-EF 3.0 also has the flowable “water vapour” defined for all of its emission contexts.

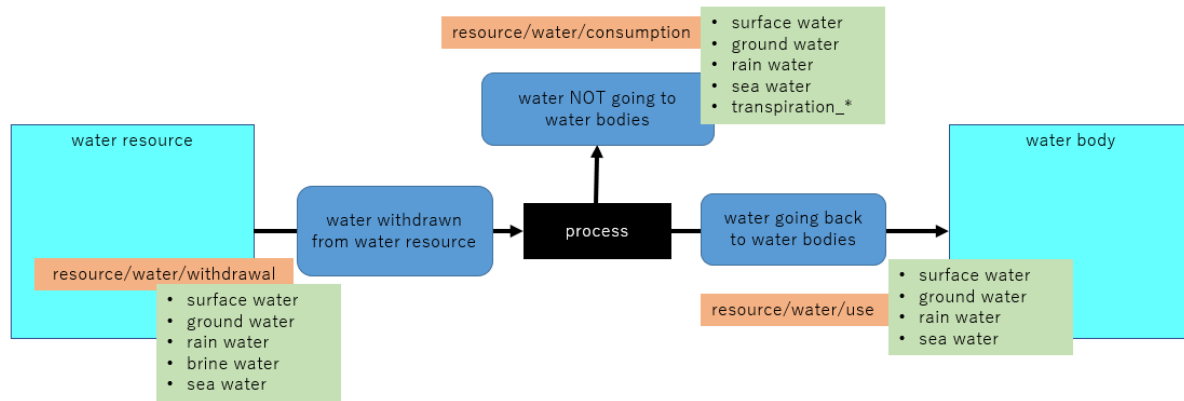


Figure 11 – IDEA's water-related elementary flows projected to the basic water model

IDEA's approach to water flows is simpler than the other EF lists, but it explicitly distinguishes rainwater. There are some points that are unique with IDEA's water elementary flow naming conventions: one thing to note is that for the water output contexts, IDEA does not use "emission/air", "emission/water" but expresses those contexts as "resource/water/consumption" and "resource/water/use". Further, the flowables "surface water", "ground water", "rain water", "sea water" that are associated with these output contexts point to what the original withdrawal water source was, and do not point to where the water is going. For example, the elementary flow "ground water" with context "resource/water/use" would mean that, this water was originally withdrawn as "ground water" and being emitted to a water body that can be considered as blue water.

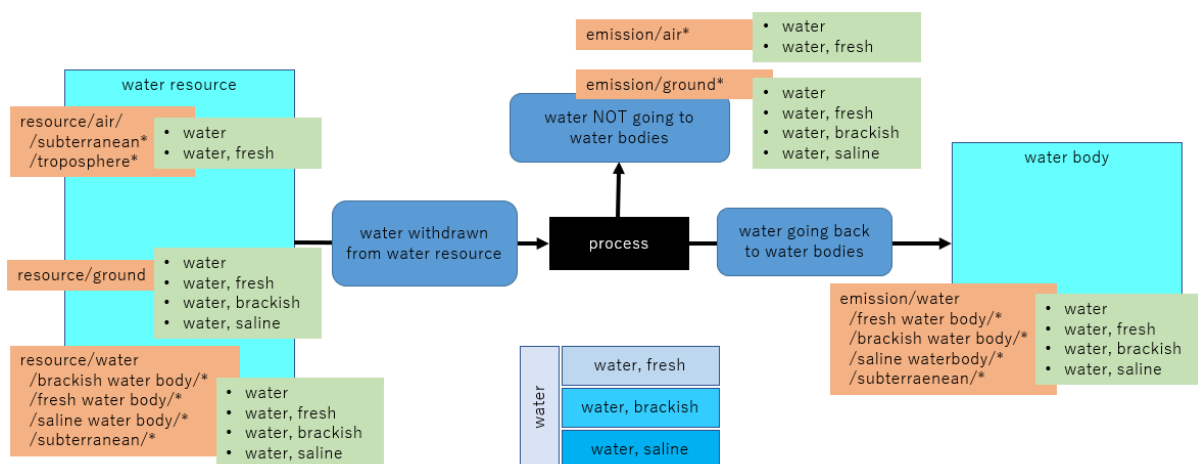


Figure 12 – The water-related elementary flows in Federal LCA Commons' flow list projected to the basic water model

Federal LCA Commons has by far the largest elementary flow list out of the four participating database systems, which is also reflected in how water flows are defined. One of the unique features of this list is that the salinity of the water can be specified in detail. Salinity is divided into four types/levels ("water", i.e., unspecified, "water, fresh", "water, brackish", "water, saline") where the salinity of the water itself is specified by the flowable part, and the salinity of the "water body" is specified by specifying the context. For example, "emission/water/brackish water body/lake/water, fresh" means that "fresh water" is emitted to a brackish water body lake and this is distinguished from "emission/water/fresh water body/lake/water, fresh" where "fresh water" is emitted to a fresh waterbody lake. Federal LCA Commons also has a long list of contexts to express where the flow is occurring. For example, "emission/air" is separated in detail by vertical height (stratosphere,

troposphere/ground-level, low, high, very high) as well as the surface situation (e.g., urban, rural). Regarding fossil water, there is a distinction between “confined aquifer” and “unconfined aquifer” where “confined aquifer” can be considered close to “fossil water”. Ground water is expressed in the context with the keyword “subterranean” meaning underground.

The next table shows the water related elementary flow features of each EF list.

Category	Feature	ecoinvent	ILCD-EF	IDEA	FEDEF
water withdrawn from water resource	usage of water included (e.g., turbine, cooling)	✓	✓		
	surface situation specified (e.g., lake, river)	✓	✓		✓
	explicit ground water (not “under ground” water)	✓			✓
	fossil well water	✓			✓
	explicit rainwater		✓	✓	
	water in air	✓	✓		✓
	explicit water body salinity of resource specified				✓
	explicit salinity of withdrawn water specified				✓
water NOT going back to water bodies	vertical height specified (e.g., close to ground etc.)	✓	✓		✓
	surface situation specified (e.g., non-urban etc.)	✓	✓		✓
	long-term	✓	✓		
	water to ground (not to under ground water bodies)				✓
	water vapour to ground (soil)		✓		
	keep origin of water withdrawal			✓	
water going back to water bodies	surface situation specified (e.g., urban)				✓
	fossil well	✓			✓
	long-term	✓	✓		
	explicit rainwater		✓	✓	
	water vapour to water bodies		✓		
	explicit salinity of emitted water specified				✓
	explicit water body salinity of resource specified		(✓)		✓
	keep origin of water withdrawal		(✓)	✓	

We decided on a rough rule when treating these advanced features and applied it to the mappings:

Category	Feature	When on target side the source side should...	When on source side the target side can ...
water withdrawn from water resource	usage of water included (e.g., turbine, cooling)	not map if not the same usage	map to an EF without usage is OK
	surface situation specified (e.g., lake, river)	not map unless the surface situation is the same	map to a wider surface situation EF
	explicit ground water (not water “under ground”)	map to “under ground” as a proxy.	map to under ground water is OK
	fossil well water	not map unless source is also fossil	map to plain under ground water is OK
	explicit rainwater		map to anything “water, in air” is OK
	water in air		treat as same as rain water if no other option
	explicit water body salinity of resource specified	try to map to the “fresh water body” when applicable.	discard the “saline” and “brackish” water bodies if applicable.
	explicit salinity of withdrawn water specified	select “water” flowable as the proxy when applicable.	discard “saline” and “brackish” water when applicable.
water NOT going back to water bodies	vertical height specified (e.g. close to ground etc.)	map to the nearest vertical height	map to a wider surface situation EF except for stratosphere
	surface situation specified (e.g. non-urban etc.)	map to the nearest surface situation that has an overlap	map to a wider surface situation EF
	long-term	not map unless also it is long-term	map to EF without long-term is OK
	water to ground (not to under ground water bodies)	treat “water” and “water vapour” as the same but “water” takes priority as a proxy	
	water vapour to ground (soil)		
	keep origin of water withdrawal	Use IDEA “surface water” as proxy	not mind the origin and map

Category	Feature	When on target side the source side should...	When on source side the target side can ...
water going back to water bodies	surface situation specified (e.g., urban)	map to the nearest surface situation that has an overlap	Map to a wider surface situation EF
	fossil well	not map unless source is also fossil	Map to under ground water is OK. map to other surface water is OK to balance blue water.
	long-term	not map unless source is also long-term	Map to EF without long-term is OK
	explicit rainwater	not map unless the origin is rainwater	
	water vapour to water bodies	treat “water” and “water vapour” as the same but “water” takes priority as a proxy	
	explicit water body salinity of resource specified	try to map to the “freshwater body” when applicable.	discard the “saline” and “brackish” water bodies if applicable.
	explicit salinity of emitted water specified	select “water” flowable as the proxy when applicable.	discard “saline” and “brackish” water when applicable.
	keep origin of water withdrawal	Use IDEA “surface water” as proxy	not mind the origin and map

Special care must be taken when mapping rainwater (or green water), since these green water flows tend to have much larger values than blue water flows. The next figure describes the problems that could occur when green water flows are mapped to blue water flows.

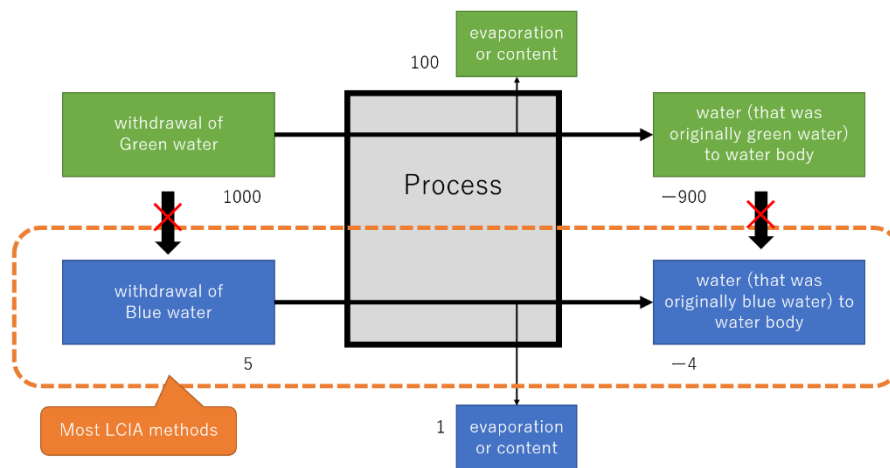


Figure 13 – Green water and Blue water problem and mapping

ILCD-EF and IDEA separate rainwater withdrawal and emission flows. Since most LCIA methods only consider blue water, these rainwater flows should not be mixed up with blue water flows during mapping and conversion activities, as they will cause bogus results at the level of LCIA.

Other aspects: missing characterization factors

Ensuring LCIA score consistency during format conversion was considered beyond the scope/requirements of this project. But the fact that ILCD-EF is also associated with an own LCIA method (based on a selection of ILCD-recommended methods) still allowed the project team to screen the characterization factors in cases where multiple mapping candidates could be found among the flows in the ILCD-EF' list. Several instances found point towards potential improvements for future updates of the flow list and/or method implementation. Examples include:

- Duplicates of flows in ILCD EF3.0 such as the “Peat, in ground” and “peat” example where only “peat” has a characterization factor.
- Flows that have a seemingly optimal but uncharacterized match, and a proxy match with a non-zero characterization factor, in LCIA. Among the examples found are “tin, ion” and “vanadium, ion” or “plutonium-alpha” for which there are seemingly perfect matches in ILCD-EF 3.0 (i.e., identically named as “tin, ion”, “vanadium, ion”, and “plutonium-alpha”).
- Characterization factors missing for specific sub-compartments only: although there is a seemingly ideal match, including the correct sub-compartment, the characterisation factor is missing for this sub-compartment. An example here is “Methane, tetrafluoro-, R-14” to ‘air/urban air close to ground’ that has been mapped to “fc-14” to ‘emission/air/urban/close to ground’. The target-side flow has no characterisation factor in ILCD-EF 3.0, but the emission of “fc-14” to air with sub-compartment ‘unspecified’ does.

Guidance for developers of data format converters

This part of the report is intended to support the implementation of the EF mapped files established under the Nomenclature Working Group of GLAD in data exchange format conversion tools.

Structure and content of GLAD's elementary flow mapped files

The common file structures for the elementary flow (EF) lists and the resulting EF mapped files were defined ahead of the present project, setting the standard for the project deliverables. An overview of the information to be contained in the GLAD EF mapped files, ordered as columns in EF mapping file (as .xlsx- or .csv) for any pairwise combination of source- and target-side EF lists, is provided in Table 3.

The outputs from this project encompass, besides the pair-wise EF mapped files as the main deliverables, several information resources that can help improve or interpret data exchange format conversion. This underlying/supplementary information might hence both support converter developers, and the end-users. These resources shall be made freely available (e.g., through UNEP/GLAD's GitHub file and code repository) to support anyone looking for more details or interested in providing suggestions for corrections or further improvements.

- The elementary flow lists of the participating database systems, formatted according to the GLAD EF mapping format and with consistent flow units (as found in the OpenLCA software) [.xlsx/.csv]
- Flowable library [Google-sheet/.xlsx/.csv]: the mapping script provided by the JRC can efficiently find matches for any combination of source- and target-side EF lists for flows (i.e., flowable+context combinations) with unambiguous attributes, e.g., by searching for identical matches of unique flow names, synonyms, or CAS Registry Numbers®. While this approach works well for a large proportion of flows in most EF list combinations, it is insufficient for several important sub-classes or groups of flowables, for which the flow attributes are either not unique (i.e., multiple flowables for carbon dioxide with CAS "124-38-9", to reflect differences in the origin of the carbon, e.g., fossil vs. non-fossil/biogenic) or too heterogeneous (e.g., energy resources).
- Twelve (4x3) pairwise and unidirectional input files for JRC's mapping script [.xlsx/.csv]. The spreadsheets contained in these files cover the following aspects:
 - Context-level: Default mapping (sheet DEFAULT_CONTEXT_MATCH) and acceptable proxies (PROXY_CONTEXT_MATCH) for contexts (i.e., the environmental compartments and sub-compartments defined for the source-and target-side EF lists, respectively).
 - Flowable-level (FLOWNAME_MANUAL): extracted from the flowable library, this list defines pairs of flowables to be mapped for any acceptable (i.e., default or proxy) combination of contexts (according to DEFAULT_CONTEXT_MATCH and PROXY_CONTEXT_MATCH).
 - Flow-level: Unique pairwise flow-matches that do not fit into the schemes listed above are defined as combinations of the flowable and specific context in sheet ONE2ONE_FLOW_MANUAL. In contrast, source-side flows that correspond to two (or more) target-side flows (e.g., land transformations in IDEA, with both start- and end-state defined in a single flow) are defined in ONE2MANY_FLOW_MANUAL. Finally, source-side flows that are not intended (i.e. specifically chosen as no-matches by source/target owners) to be mapped against the target-side EF list are entered in **NO_FLOW_MATCH_MANUAL**.

- **NO_MAPPING** indicates that the tool did not find a match, according to all the sub-iterations and the info contained in the Context Mapping file.
- Matches (any type) with **N/A as a conversion factor**: For those the tool actually finds a match but no conversion factor is provided (or not robust enough to avoid conversion errors).

The last three cases (**in bold** above) have to be excluded from conversion, and within the conversion log shall be listed, and differentiated by type (e.g., “not to be mapped” for **NO_FLOW_MATCH_MANUAL**; “not mappable” for **NO_MAPPING**; “unit not convertible” for **N/A** as conversion factors).

- MULTIMATCH is essentially working in the opposite way compared to ONE2MANY_FLOW_MANUAL, to resolve the distribution of “from” and “to” in Land use, and re-create the unique flow in IDEA system, inventory data are needed, to assign the proper quantities.
- Twelve (4x3) unidirectional and pairwise mapping files

Table 3 – overview of information provided in the GLAD EF mapped files (ordered as columns in the .xlsx/.csv file)

Flow information		Comment
SourceListName	TargetListName	Name and version of source-/target-side EF list
SourceFlowName	TargetFlowName	Name of source-/target-side EF (i.e., ‘flowable’)
SourceFlowUUID	TargetFlowUUID	The flow’s <i>Universally Unique Identifier (UUID)</i> in source-/target-side EF list
SourceFlowContext	TargetFlowContext	source-/target-side context, as concatenation of flow direction (i.e., resource or emission), main environmental compartment (e.g., air, ground, water), and any sub-compartments
SourceUnit	TargetUnit	Flow units (as found in the OpenLCA software)
Match criteria		Enumeration options
MatchCondition	=, ~, <, >, <>, [...]	Qualitative (logical) relationship between source- and target side EF, e.g., nitrogen dioxide (NO ₂) < nitrogen oxides (NO _x).
ConversionFactor		Quantitative relationship between source- and target-side EF (e.g., converting kg to MJ) <u>Note: the entry “N/A” here indicates that a quantitative relationship could not be established (with an acceptable degree of uncertainty) by the project team, and should be flagged in the log file of the format conversion.</u>
Meta-information		
MapType	CAS CAS (PROXY) NAME ONE2ONE_FLOW_MANUAL ONE2MANY_FLOW_MANUAL FLOWNAME_MANUAL FLOWNAME_MANUAL (PROXY) NAME (PROXY) NO_MAPPING SYNONYM_TO_NAME [...]	Record of how the mapping was established. Please note that further enumeration options might be added in the future. Also, the entry ‘ONE2ONE_FLOW_MANUAL’ applies both to matches established in ONE2ONE_FLOW_MANUAL (as an input for the mapping script), and items entered manually directly into the EF mapped files.
Mapper		person responsible for first establishing or validating (in the case of mapping script outputs) the EF match

Verifier	person responsible for verifying appropriateness of the EF match
LastUpdated	Date of last edit/modification to the specific EF mapping entry

Log of conversion

Conversions between the data exchange formats and nomenclature systems commonly used for LCA should strive to preserve and convey the complete information contained in the original dataset in its native format. “Perfect” conversion, without any loss of information, however, is currently rarely possible in practice. This is partially due to the differences in scope of, and conventions applied in, the different EF lists. As a result, a trade-off is necessary between (maximizing) completeness and ensuring a high degree of accuracy, or flow correspondence, for each item matched, when establishing EF mapped files.

Regardless of whether data exchange format conversion is applied to unit process datasets (UPRs) or (partially) aggregated LCI datasets, a ‘log’ shall be generated alongside the converted datasets. This log should, as a minimum, provide an account of:

- The data exchange formats considered (source and target);
- the name and version of the EF list used in the input dataset(s);
- the names and version of the EF lists used for the data exchange format conversion;
- the number of flows successfully converted (per dataset);
- each individual EF dropped (i.e., not converted into the target format) during the conversion, and;
- Each instance for which the conversion factor is not equal to 1 (i.e., not left blank) or could not be established (**indicated by “N/A” in the EF mapped files**) has been applied.

This information will allow the end-user to review the conversion results, without comparing the datasets in the input- and output- formats side-by-side, to assess whether any loss of information is acceptable for the intended use.

A more extensive conversion log may go beyond a compilation of mapped or dropped EFs. It might, for example, be helpful to screen the datasets and/or log against relevant impact assessment methods, to get an indication of LCIA score consistency. As outlined in the next sub-section, it might also be relevant to create a more detailed record, including an explicit account of any flows merged/split, unit conversions, or proxy contexts applied.

Mapping record in exchange comments

Several data exchange formats enable comments to be entered for the individual exchanges (i.e., both elementary and intermediate/technosphere flows) in the datasets. These fields typically carry meta-information about the data sources, model parameters, or assumptions used for the original inventory modelling. It is here proposed to use these exchange-specific comment field to also record aspects of the data exchange conversion for full transparency. This could take the form of a text string added to an existing exchange comment, with a clear notation indicating the original flow name, context, UUID, amount and unit. This could be applied consistently to all converted flows, or just those changes considered relevant enough. This includes when a conversion factor is used (that is, for mapped flow amounts other than 1:1 and/or of different flow units), merged or split flows, or when an ‘unspecified’ flowable or context on the target-side is used for a more specific flows on the source-side.

Version control/support

The EF mapped files established within the present project relate to specific versions of the EF lists of the participating database systems. Smaller discrepancies can be expected for EF lists updated in smaller increments (e.g., for ecoinvent EF v3.6 → v3.7) compared to more substantial revisions (Environmental Footprint 2.0 → 3.0, or IDEA EF v2.2 → v2.3). Other databases, e.g., among the other nodes connected to GLAD, can be set up using any EF list (or version thereof). As such, the precision/accuracy of conversion between two data exchange formats will also be affected by the degree of correspondence between the EF list used for the original dataset, and the EF mapped files available in the format converter.

Proposals for corrections or further improvements

While we foresee that a ‘mapping coordinator’ would be the contact point for general inquiries about the GLAD EF mapping system, the UNEP/GLAD repository on GitHub will enable end-users or tool developers to report errors in the EF mapped files or other more specific improvement suggestions. The affected EF list owners would be asked to review the requested changes before any modifications are merged into the master branch. The GitHub platform also ensures that all changes to the mapping resources are recorded and provides easy version control of all files.

Appendix A – General mapping principles

The general mapping principles listed in the tables below are largely aligned with (and in many cases adapted directly from) an input received in April 2021 from Martin Baitz with the logics for nomenclature mappings applied at Sphera (e.g., for the integration of external databases and LCIA methods in the GaBi software).

Context mapping

Case	Flow list 1	Flow list 2	Comment/option
1.1	“Standardized” scope of compartments and sub-compartments	“Standardized” scope of compartments and sub-compartments	Mapping 1 ↔ 2 between perfectly consistent contexts <i>without loss of information</i>
1.2	(Sub-) compartment existing	(Sub-) compartment not existing	Mapping to acceptable proxy or to ‘unspecified’ compartment <i>two-way information loss only avoidable if proxy mapping is unique</i> (i.e., not used for multiple sub-compartments in Flow list 1)
1.3	Sub-compartment aggregated via “or” logic (Comp A or B)	Sub-compartment disaggregated in Comp A and separate Comp B	Mapping 1 → 2 is random Mapping 2 → 1 aggregation <i>Information loss unavoidable</i>
1.4	Sub-compartment aggregated via “and” logic (Comp A and B)	Sub-compartment split into Comp A and Comp B separately	Mapping 1 → 2 is random Mapping 2 → 1 aggregation <i>Information loss unavoidable</i>
1.5	Compartment indirectly defined, e.g., as “non-XYZ”	Compartment directly defined “ABC” (being a valid non-XYZ)	Mapping 1 ↔ 2 unique, if “ABC” is the only “non-XYZ” <i>Information loss unavoidable, DEF is also “non-XYZ” or if further sub-compartments exist “ABC1”, “ABC2”, ...</i>
1.6	Sub-compartment ‘unspecified’	Sub-compartment ‘unspecified’	Mapping 1 ↔ 2 <i>without loss of information</i>

Mapping of flows or flowables

Case	Flow list 1	Flow list 2	Comment/option
2.1	Standardized flow list A	Standardized flow list B	Mapping 1 ↔ 2 between perfectly consistent flows <i>without loss of information</i>
2.2	Flow A existing	Flow A does not exist	Mapping to acceptable proxy or more generic flow in Flow list 2 <i>two-way information loss only avoidable if proxy mapping is unique</i> (i.e., not used for multiple flows in Flow list 1)
2.3	Flows for metal resource extraction defined as ores	Flows for metal resource extraction defined as target elements/minerals	Ores to be split into corresponding target elements/minerals and gangue, and <i>vice versa</i>
2.4	Emission flow defined as ion	Emission flow defined as element	Same LCIA values? Partly irrelevant differences, partly extreme different impacts

2.5	Flow with additional information (scope in information given in flow name or in docu field)	Standard physical flows (scope information given via software)	N to 1 mapping, Documentation/information loss due to aggregation
2.6	Sum flows (e.g., nitrogen oxides, NO _x ; sulfur oxides, SO _x ; dioxins)	Specific flows (NO, NO ₂ ; SO ₂ , SO ₃ ; 2,3,7,8-Tetrachlorodibenzodioxin, etc.)	Most representative or standard reference substance (specific flow) may be used to map sum/group flows with little to no information loss
2.7	Unspecified flows like 'phosphates'	Specific flows, e.g., sodium hydrogen phosphate	Mapping of unspecified → specific flows generally discouraged, unless the latter is regarded equivalent or well-representative of the unspecified flow

Unit mapping

Case	Flow list 1	Flow list 2	Comment/option
3.1	Standardized SI units, <i>identical to the flow property determining impact (characterization)</i>	Standardized SI units, <i>identical to the flow property determining impact (characterization)</i>	Direct mapping 1 ↔ 2 between identical units generally possible without loss of information
3.2	Standardized SI units, <i>different from the flow property determining impact (characterization)</i>	Standardized SI units, <i>different from the flow property determining impact (characterization)</i>	Mapping with conversion factor based on ratio of determining physical properties, as defined for each list (e.g., LHV ₁ /LHV ₂)
3.3	Flow A with SI units	Flow A with Imperial units	Mapping with conversion factor (e.g., lb vs kg, MMBTU vs MJ) related to same physical quantity (mass to mass, energy to energy)
3.4	Flow A defined by mass (e.g., reference flow in kg)	Flow A defined by energy content (reference flow in MJ based on upper or lower heating value)	Mapping with conversion factor based on physical properties (for mass to energy)

Appendix B – overview of project outputs

- Common data format/structure for elementary flow (EF) lists and EF mapped files within GLAD – *prepared by the GLAD Working Group on Nomenclature (WG1) prior to project*
- The ‘GLAD Mapper’, an EF mapping script for the semi-automatic generation of preliminary EF mapped files – *developed and provided in-kind by the Joint Research Centre of the European Commission in conjunction with the project implementation*
- Mapping information and resources for all combinations and directions between the four participating database systems
 - Context mapping files
 - Flowable library
 - Conversion factor library
- Cross-reviewed mapped files between the following EF list versions (*deliverable D 1.2*)
 - European Commission’s *Environmental Footprint version 3.0* (ILCD-EF 3.0)
 - *ecoinvent version 3.6/3.7* of Switzerland
 - *IDEA versions 2.2 and 2.3* (IDEA v2.2/v2.3) of Japan
 - U.S. *Federal LCA Commons version 1.0.3* (FEDEFL v1.0.3)
- Scripts for the systematic analysis and review of EF mapped files
- A proposal for governance and maintenance of the GLAD EF mapping system beyond the present project (*deliverable D 2.3 – Maintenance manual for the GLAD EF mapping system*)
- Documentation of the procedure, common issues and solutions implemented to establish the mapping files, as well as guidance for developers of data format converters (*deliverable D 1.3; this report*)
- A scientific publication summarizing the background, goal and scope, approach, and results of the project, as well as the lessons learned in the process (*in preparation*)

The total coverage of the mapped files (deliverable D1.2) is summarized in Table 4.

Table 4 – summary of mapping coverage for the pair-wise combinations of EF lists covered in the project.

Source Target	ILCD-EF 3.0	ecoinvent EF 3.7	FEDEFL v1.0.3	IDEA v2.3	IDEA v2.2
ILCD-EF 3.0	X	98.6%	89.4%	95.0%	92.9%
ecoinvent EF 3.7	23.5%	X	28.8%	68.0%	89.8%
FEDEFL v1.0.3	62.3%	94.5%	X	90.3%	84.9%
IDEA v2.3	17.3%	41.2%	20.7%	X	X
IDEA v2.2	15.1%	31.3%	16.5%	X	X