Structured Publication of Life Cycle Assessment Models and Results

Brandon Kuczenski
Institute for Social, Behavioral, and Economic Research
University of California, Santa Barbara

2017-03-31

Abstract

Many of the challenges facing knowledge synthesis from life cycle assessment (LCA) studies stem from the inability of study authors and readers to formally agree on the structure and content of life cycle inventory models. This article presents a framework for LCA study authors to publish inventory models in a structured format based on a mathematical formulation of a *foreground study*. Using linked semantic data as a foundation, the publication framework provides for both machine readability and unambiguous human interpretation of published studies. To implement the framework, the LCA community should find ways to enable study authors and data users to make common reference to shared data sources, including background databases, elementary flows, and impact assessment methods, that are widely used in many studies. The framework contributes to ongoing efforts within industrial ecology to improve the reproducibility and verifiability of scholarly works, and if implemented, plots a course toward distributed, platform-independent computation and validation of LCA results.

1 Introduction

Life cycle assessment (LCA) is a standard methodology to estimate the potential environmental impacts of products or services by modeling the network of industrial processes that must occur to deliver them to a consumer. The technique is well-established, widely practiced, and draws on an extensive body of standardization and scholarship. Increasingly, the results of LCA studies are appearing in environmental declarations, corporate sustainability reports, and product marketing information. LCA is also gaining prominence as a tool for developing and evaluating environmental policy. At the core of an LCA study is a model of a "product system," which represents a collection of "processes . . . performing one or more defined functions", that describes the life cycle of a product (ISO 14044, 2006). Product system models encode information about how products are manufactured, distributed, used by the consumer, and what happens to them after they are disposed. A product system can be divided into a foreground, which denotes the portions of the life cycle whose operations are directly modeled by the study, and a background, which represents the global industrial system (Clift *et al.*, 1998). While the foreground is often modeled from direct observation or other primary data, background processes are typically drawn from a life cycle inventory (LCI) database prepared by a third party.

Although there has been considerable effort to normalize and harmonize LCI database design methodology and data interchange (UNEP/SETAC, 2011; Recchioni *et al.*, 2013; Mila i Canals *et al.*, 2015; Ingwersen *et al.*, 2015), less attention has been paid to harmonizing product system models. These models are often highly complex and include countless modeling decisions, approximations, and assumptions made by a study author (Lloyd & Ries, 2007; Reap *et al.*, 2008). Consequently, while comparative evaluations can be made regarding alternative cases or scenarios within a single study, comparisons across studies are much more challenging (Heath & Mann, 2012; Henriksson *et al.*, 2014). Even when relatively simple and homogeneous systems are considered, wide variations can be found in the results of studies from different authors (van der Harst & Potting, 2013; Turconi *et al.*, 2013). A lack of transparency in reporting, particularly regarding definitions of study scope and system boundary, is a crucial challenge to the interpretation of results in comparative analysis (Cleary, 2009; Laurent *et al.*, 2014). One consequence is the high cost and uneven rigor of critical review, in which the complexities of LCA come face to face with the limitations of current practice (Curran & Young, 2014).

As LCA gains prominence, particularly in the policy realm, these problems become more acute. If LCA is to be used in policy, it will necessarily have the effect of recommending certain product systems, technologies, or approaches at the expense of others; yet in that event it is critically important that there be a consensus among stakeholders that the quantitative results are credible and well-supported (Rainville *et al.*, 2015; McManus *et al.*, 2015). As noted above, this level of consensus is hard to realize, in part because of the conflicting implications inherent in different analytic modes. Other factors such as allocation strategy (Pelletier *et al.*, 2014) can also render study findings unreliable or highly contingent. Plevin et al. famously observed in the strongest terms that careless reporting of attributional LCA study results can distort the significance of findings in a policy context (Plevin *et al.*, 2013), a shortcoming that can be found in other modes as well (Brandão *et al.*, 2014). In the scope of Environmental Product Declarations (EPDs), studies within an industry are supposed to be rendered comparable by adhering to a common product category rule (PCR) (Fet & Skaar, 2006). However, this prescription is insufficient to ensure the comparability of different declarations, even if they use the same PCR (Modahl *et al.*, 2012). Moreover, PCRs are often themselves not unique and can include wide variations in scope and boundary requirements (Subramanian *et al.*, 2012), leading to ambiguity.

A common theme in all these controversies is the inability for authors and their readers to formally agree on the structure and content of the product system model. In the scope of academic research, the transparency of techniques and reproducibility of results are paramount concerns (Mesirov, 2010). But reproducibility of a computation requires a distinction between the computing algorithm and the input data (Buckheit & Donoho, 1995; Fomel & Claerbout, 2009); contemporary LCA lacks this distinction. While the ostensible input to an LCIA computation is a product system model, this input cannot be separated from the computational environment used to author it. Instead, models are distilled into written documents describing LCA results, be they academic papers, ISO reports, PCRs or EPDs. These are uniformly *unstructured*, meaning their interpretation must be secured by a human reading text and graphical images such as system boundary diagrams, but cannot be expressed to a machine.

The objective of *structured* publication, on the other hand, is to enable the reader of a publication to automatically interpret its contents and, ideally, reproduce its results. Such a publication is both machine-readable and easily human-interpretable. In this paper, I propose a framework for the structured publication of product system models in LCA that achieves these dual requirements.

2 Requirements for Structured Publication of Study Models

An LCA study result is ultimately an assertion that, for some input data known as the "model," the delivery of a particuar reference flow is associated with a certain amount of environmental impact or impact potential. Structured publication of a model by itself is simply documentation of the model's design. On the other hand, structured publication of the model together with the result constitutes an assertion that the model and the result correspond. The fundamental objective of structured publication is to enable a data user to validate that assertion.

As discussed in the introduction, a structured publication must achieve the dual requirements of human interpretability and machine readability. This is often accomplished using linked semantic data (Bizer *et al.*, 2009), in which an object to be interpreted is signified by a link to a resource on the World Wide Web, typically referred to as a Uniform Resource Identifier (URI) or a hyperlink. The resource at the end of the link functions as a point of agreement: both study author and data user can follow it from anywhere on the Internet to obtain the same information. The content pointed to by the hyperlink, along with context provided by *other* references to the same resource, gives meaning to the information and allows it to be curated by the study authors and other members of the community (Khan *et al.*, 2011). Semantic data are structured by reference to knowledge models called ontologies, which describe the relationships among different types of entities (Madin *et al.*, 2008). Some preliminary descriptions of the entity types involved in LCA computation, and their relationships to one another, have recently been developed (Ciroth & Srocka, 2014; Janowicz *et al.*, 2015; Kuczenski *et al.*, 2016).

However, the existence of semantic data resources is not enough by itself to establish a reproducible result: it must also be clear how the information is used to generate the result. A *research object* (Bechhofer *et al.*, 2013) is an aggregation of linked references, combined with a structure or system for converting the information into the scientific model for reproducing the result. In order to develop a research object that can encapsulate the publication of an LCA study, it is necessary to identify the computation to be performed, and then supply the data necessary to perform it. The following objectives will be used to guide the this development:

1. *Linked Data Foundation*. Both human and machine interpretation of the publication should rely on the use of references to shared resources identified by URIs or hyperlinks.

2. *Computation*. The role of the publication in an LCA computation should be clear. If the publication includes a result, it should be clear how the result was obtained.

3. *Attribution*. The publication should distinguish information attributable to the author from information attributable to other sources.

4. *Completeness*. The publication should contain all the parts of the model necessary to accomplish the goal.

5. *Conciseness*. The publication should includes the smallest amount of information necessary to accomplish the goal.

6. *Flexibility*. The publication form should accommodate a variety of levels of disclosure and transparency.

2.1 Mathematical Formulation of an LCA study

Collections of LCA datasets can be partially ordered based on their dependency on one another (Kuczenski, 2015). This observation allows for a formal distinction between foreground systems and background systems. A foreground system has a dependency on an LCI database, hybrid or input-output database, or software system, whereas the background system does not have a reciprocal dependency on the foreground. Because of this, it is possible to describe foreground systems precisely without including the background in the description. This section develops a precise formulation of an LCA *foreground study*, which is used as the computational basis for the research object.

We begin with the classic formulation of the LCA problem (Heijungs & Suh, 2002):

$$s = \mathbf{e} \times B \times A^{-1} \times \mathbf{y} \tag{1}$$

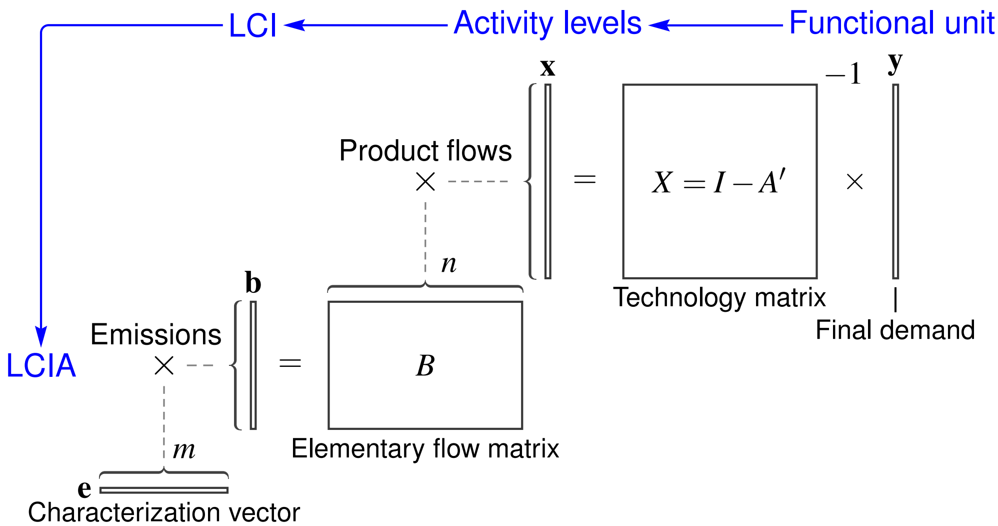
where A is the technology matrix, B is the environmental intervention or emission matrix, $\mathbf{e}$ is a vector of characterization factors, $\mathbf{y}$ is the externally specified final demand, and s is the numerical impact score or category indicator. For simplicity, we will consider the computation of a single impact result. However, it is straightforward to imagine the more typical case, in which E is a matrix of characterization vectors, and $\mathbf{s}$ is a vector of results.

It can be shown that, under conditions typical for LCA databases, the technology matrix can be represented as an input-output database, which is made up of separate supply and use tables, each of which has been normalized to a unit of output (Suh *et al.*, 2010; Pauliuk *et al.*, 2015):

$$s = \mathbf{e} \times B \times \left(I - A'\right)^{-1} \times \mathbf{y} \tag{2}$$

Here A' represents the *direct requirements matrix*, where each column of values reports the necessary inputs (positive values) and generated non-reference outputs (negative values) per unit of a process's reference output. The conditions under which this system is solvable, particularly with respect to different strategies for handling co-production, have been discussed extensively (Majeau-Bettez *et al.*, 2014). Eq. 2 is visualized in Figure 1.

The direct requirements matrix also describes a directed graph, in which each column is a node, and nonzero entries of A' define edges between nodes. This graph can be partially ordered based on the directions of the edges, so the foreground can be identified as the set of processes earliest in the ordering, on which the background does not depend. An appropriately ordered A' matrix can



$$s = \mathbf{e} \times \mathbf{b} = \mathbf{e} \times B \times \mathbf{x} = \mathbf{e} \times B \times (I - A')^{-1} \times \mathbf{y}$$

Figure 1: Traditional matrix structure of an LCA computation.

be written in block-triangular form:

$$A' = \begin{bmatrix} A_f & 0 \\ A_d & A* \end{bmatrix}; \quad B' = \begin{bmatrix} B_f & B* \end{bmatrix} \tag{3}$$

In this case, the submatrix A_f represents the foreground; $A*$ represents the background, and the rectangular matrix A_d represents the dependency of the foreground on the background. The ordered B matrix is similarly partitioned into B_f, which includes foreground emissions, and $B*$ which includes background emissions. The constitutive characteristic of this formulation is that the background system does not depend on the foreground, allowing for the submatrix in the top right corner of Eq. 3 to be zero. As long as this is the case, all computations regarding the background are invariant with respect to any foreground, and they can be computed in advance. The following formulation is equivalent to Eq. 3:

$$A' = \begin{bmatrix} A_f & 0 \\ A_d & 0 \end{bmatrix}; \quad B = \begin{bmatrix} B_f & B_x \end{bmatrix} \tag{4}$$

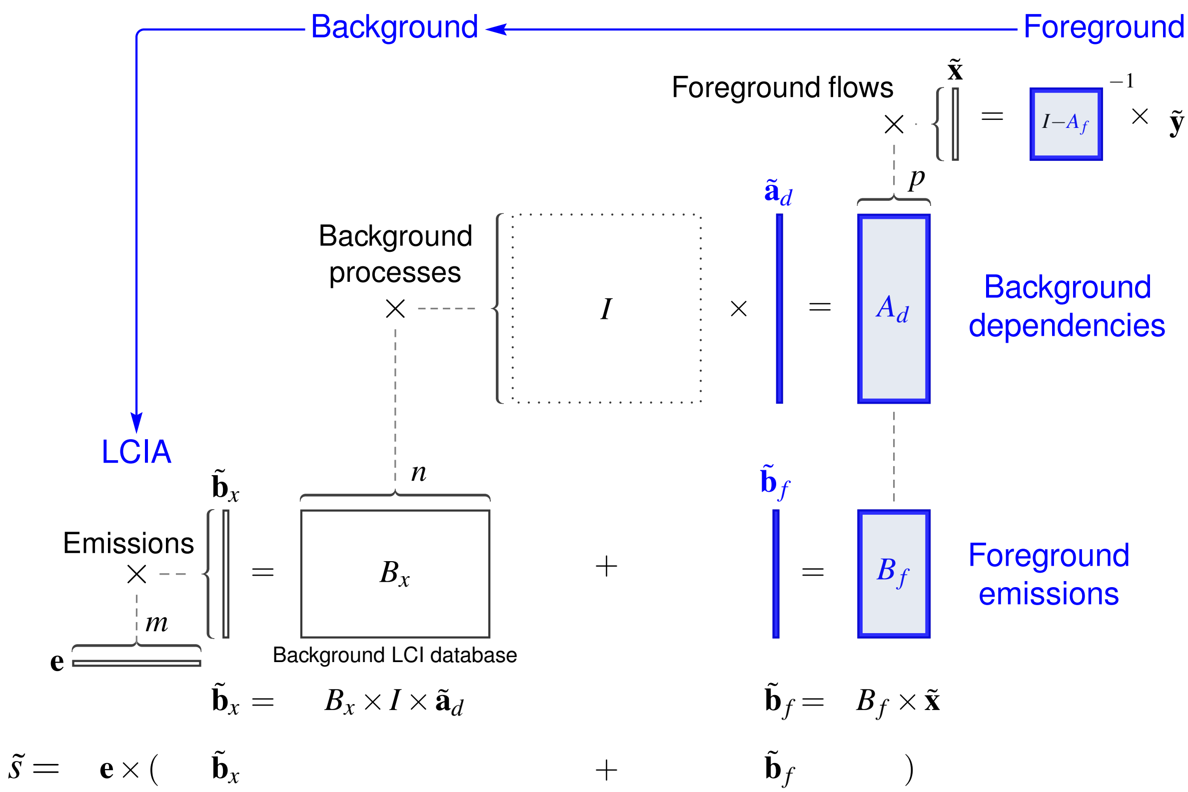


Figure 2: Matrix structure of an LCA foreground study. The background technology matrix X has been replaced with a background LCI database B_x. The foreground input matrices and aggregation results are highlighted.

where $B_x = B* \times (I - A*)^{-1}$ is a background LCI database. It is currently common practice for B_x to be prepared in advance by a background database maintainer. In this case, Eq. 2 can be written so that the computationally costly foreground and background inversions can be performed separately:

$$s = \mathbf{e} \times (B_f + B_x \times A_d) \times (I - A_f)^{-1} \times \mathbf{y}_f \tag{5}$$

where $\mathbf{y_f}$ represents final demand of the foreground processes only. Eq. 5 is visualized in Figure 2. This formulation shows that given access to the background database B_x and the characterization vector $\mathbf{e}$, the study results can be fully reproduced using only the submatrices A_f, A_d, and B_f.

The foreground graph A_f can be distilled into a vector of activity levels for the foreground nodes, which determine the magnitude of dependencies and emissions for a particular output. With the functional unit for one study formulation given as $\tilde{\mathbf{y}}$, the foreground activity levels can be given as:

$$\tilde{\mathbf{x}} = (I - A_f)^{-1} \times \tilde{\mathbf{y}} \tag{6}$$

Although the variable $\tilde{\mathbf{x}}$ can be computed through matrix inversion as expressed in Eq. 6, often matrix inversion is not required. If the foreground does not include any feedback loops, then the activity levels of foreground can be determined simply by traversing the graph (Bapat & Ghorbani, 2014), which can be performed in linear computational time with respect to the number of foreground links.

For a typical foreground study, a "canonical" $\tilde{\mathbf{y}}$ may be suggested in which the first foreground node produces the functional unit as output:

$$\tilde{\mathbf{y}} = [1, 0, 0, \ldots, 0]^T \tag{7}$$

In the canonical case the vector $\tilde{\mathbf{x}}$ is the first column of the matrix $(I - A_f)^{-1}$ and can be computed by traversal of the foreground tree beginning on the first node.

The node weights $\tilde{\mathbf{x}}$ can be used to describe an *aggregated foreground*, which generates the same results as the fully expanded foreground:

$$\begin{aligned} \tilde{\mathbf{b}}_f &= B_f \times \tilde{\mathbf{x}} \\ \tilde{\mathbf{a}}_d &= A_d \times \tilde{\mathbf{x}} \\ \tilde{\mathbf{b}}_x &= B_x \times \tilde{\mathbf{a}}_d \end{aligned} \tag{8}$$

These vectors summarize the contents of the foreground without disclosing its detailed structure. The aggregated dependency vector $\tilde{\mathbf{a}}_d$ has the same dimension as the background database, and the aggregated emission vector $\tilde{\mathbf{b}}_f$has the same dimension as the elementary flow matrix. The dependency vector can be transformed into an elementary flow vector by multiplying by B_x. The result of the study $\tilde{s}$ is the sum of foreground and background impact scores:

$$\begin{aligned} \tilde{\mathbf{b}} &= \tilde{\mathbf{b}}_f + \tilde{\mathbf{b}}_x \\ \tilde{s} &= \mathbf{e} \times \tilde{\mathbf{b}} \\ &= \mathbf{e} \times \tilde{\mathbf{b}}_f + \mathbf{e} \times \tilde{\mathbf{b}}_x \\ &= \tilde{s}_f + \tilde{s}_x \end{aligned} \tag{9}$$

2.2 Components of a Study

The formulation shown in Figure 2 shows four distinct data types in an LCA foreground study:

- **Foreground flows**: (Rows and columns of A_f; rows of $\tilde{\mathbf{y}}$; columns of A_d or B_f) The foreground flows show how the model is constructed. The foreground itself is made up of a collection of nodes. Each node represents both a process and that process's reference product (flow). The node weight vector $\tilde{\mathbf{x}}$ indicates the activity levels of all foreground nodes within the system boundary resulting from a particular (or canonical) final demand. It also expresses the magnitudes of product flows emanating from the foreground processes themselves.

- **Background processes**: (rows of A_d or $\tilde{\mathbf{a}}_d$; columns of B_x) The background processes show how the dependencies of the foreground on the industrial system are resolved to specific inventory data.
- **Elementary flows**: (columns of $\mathbf{e}$, rows of B_f or $\tilde{\mathbf{b}}_f$; rows of B_x or $\tilde{\mathbf{b}}_x$) Elementary flows cross the nature-technosphere boundary and generate environmental or social impacts.
- **Characterization quantities**: (rows of E or identity of $\mathbf{e}$) These indicate the impact assessment measurements (or impact category indicators) reported in the study.

The foreground is enclosed by an implicit system boundary which includes all foreground nodes and their reference flows. Therefore, the list of foreground nodes and reference flows can act as a de facto definition of the effective scope of the model.

The foreground matrices A_f, A_d, and B_f together include all the study-specific information. Assuming a data user had access to B_x and $\mathbf{e}$, the three foreground matrices would allow a reader to reproduce the result. If the foreground data were omitted, the result could be reproduced from the intermediate aggregation results: $\tilde{\mathbf{a}}_d$ and $\tilde{\mathbf{b}}_f$.

2.3 The Foreground

Generally, any directed graph can be represented in A_f, but some fundamental designs are common. Some simple foreground models are shown in Figure 3. The first (a) is a sequential model, in which each node requires one foreground input and generates one output. This model is equivalent to a "gate to gate" model. Here the weights k_i indicate the amount of the preceding reference flow that is required by the subsequent node. Figure 3(b) shows an additive model, in which the outputs of several foreground nodes are added together, equivalent to a "mixer" or a horizontal average. In this arrangement the weights should add up to a unit output of the reference node.

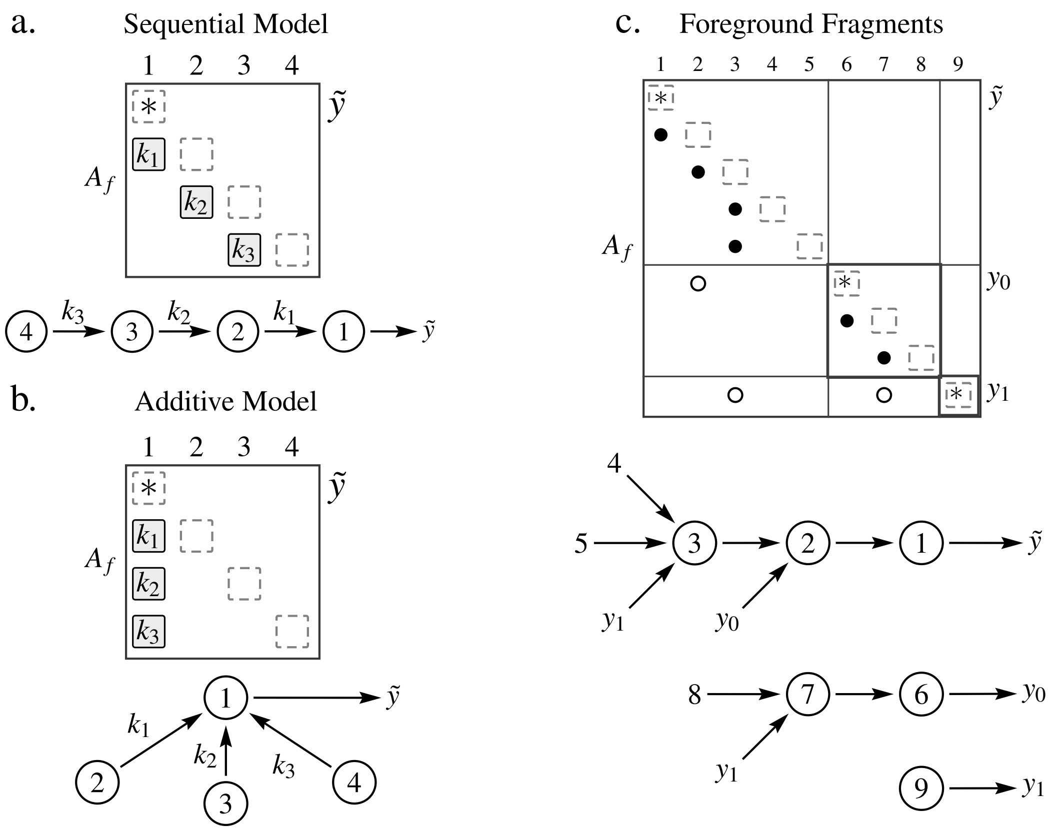


Figure 3: Equivalent matrix representations and graphs for different foregrounds.

2.3.1 Model fragments and foreground inputs and outputs

Conventionally, the foreground flows are entirely inside the system boundary: they are produced by one foreground node and consumed by one or more others. In this case, the only product flow which crosses the implicit system boundary is the reference flow, which provides the functional unit of the study given in $\tilde{\mathbf{y}}$. By implication, the difference $\tilde{\mathbf{x}} - \tilde{\mathbf{y}}$ represents the magnitudes of interior flows within the foreground.

However, not all these flows are necessarily interior. Most complex life cycle models involve nested model designs, in which low-level model components are encapsulated within higher-level models. A foreground matrix that describes an incomplete life cycle model could be called a model "fragment." In model fragments, some foreground flows represent product flows consumed within the fragment but not produced there, or vice versa. These flows represent inputs and outputs to the foreground model that may be produced or consumed by other foreground systems. Input/output flows can be detected by identifying foreground flows that show up as row entries but have no column entries in the A_f, A_d, or B_f matrices. If there is no other foreground system that connects to an input/output flow, it may be considered to be a "cutoff flow" that is excluded from the model scope.

An example of foreground composed of fragments is illustrated in Figure 3(c). Here, the nodes labeled 1–5 represent one fragment, which generates the foreground's canonical reference flow $\tilde{y}$. This fragment requires two interior flows from separate fragments (y_0 and y_1), and has two cutoff flows (4 and 5). The reference y_0 is supplied by a second fragment, made up of nodes 6-8. The reference y_1 is supplied by another fragment made of only one node (9). The reference flow y_1 is consumed in two different places by the other fragments.

2.3.2 Foreground Dependencies and Foreground Emissions

While the A_f matrix establishes exchanges of foreground flows, these exhanges have no environmental impacts and only serve to compute the activity levels of the foreground nodes themselves. The quantifiable impacts from these nodes arise either from their dependency on background processes or their direct environmental emissions.

In this context, background processes include any activities that are not modeled explicitly by the model author, but are drawn from an external data source. These generally include flows such as the consumption of grid electricity or fuels, the procurement of raw materials, the use of transportation or waste disposal services, and other points of reliance on the industrial ecosystem. The contents of this matrix are often kept secret because they can reveal sensitive information about the foreground processes being modeled. While the model depicted in Figure 2 shows reliance on a single background LCI database, foreground nodes could easily draw on processes that belong to multiple data sources.

Foreground emissions include direct emissions from foreground processes that are not already included in background models. Examples include process emissions from vents or into waterways or soil, explicitly modeled foreground combustion processes, or use phase emissions. It is tempting to think of the B_f matrix as including all direct foreground emissions. This is inaccurate because many direct-emitting processes, such as fuel combustion, are in fact modeled in the background and only referenced by the foreground, via A_d. Many studies could have a zero B_f matrix.

2.4 Publishing the Foreground Model

A structured publication could include any of these elements, bundled into a research object that meets the objectives laid out above. This research object must include:

- *tabular data* which enables the data user to construct the matrices and perform the computation;
- an *entity map*, which allows for precise interpretation of the elements of the study.

2.4.1 Tabular Data

The objective for reproducibility of computation can be achieved with reference to the formulation in Equations 5–9. The publication contents can include any of the intermediate results, or all of them from end to end. These components are all vectors or matrices, and it is easy to produce a machine-readable rendition of a vector or matrix. Of innumerable possible formats, two broad subtypes deserve explicit mention: "full" representation and "sparse" representation. A full representation of a matrix is a literal grid of numbers, such as a spreadsheet. The spreadsheet can be equivalently represented as a delimited file, where each row is a single line of the file (e.g. separated by newlines), and the cells within the row are separated by a delimiter such as a comma.

The sparse representation assumes the matrix is largely zeros, and only reports the non-zero elements. A sparse representation could also take the form of a delimited file, except each line would contain exactly three entries: the row index, column index, and value of the non-zero matrix entry.

Given that the matrices being represented are largely made up of zeros, the sparse approach is preferable to the full one.

The data tabulation described here is, ultimately, optional: for reasons of privacy, an author may wish to conceal some or all data points, or include the data points but exclude the actual data values. This is allowable if it is consistent with the author's goal, and it will accordingly limit the ability of a data user to reproduce the model.

2.4.2 The Entity Map

The matrix formulation also provides a clear statement of the components that make up the LCA study. To be specific, the rows and columns of each vector or matrix correspond to the entities in the model. The publication should include a "mapping" in which every row and column that contains a nonzero matrix entry is linked to an externally published reference that describes what the row or column represents. This mapping from mathematical element to semantic meaning forms the "linked data foundation" of the publication.

The entity map must provide some minimal information about the entities in each of the four groups discussed above:

- **Foreground flows**: Because the foreground is specific to the model, the publication itself *is* the basis for the semantic meaning of the foreground flows. Authors should simpy provide enough detail that the significance of each flow is clear to a data user. Each foreground flow must specify its *reference unit*, which is the amount of the flow that corresponds to a unit activity of the foreground node. The data values reported elsewhere in the column must be normalized to this unit.
- **Background processes**: A properly identified background data set should correspond to a distinct aggregated unit process dataset from a background data provider. The reference units of the background processes should already be part of the background data publication

and do not need to be reproduced in the foreground publication except for clarity.

- **Cutoff flows**: Cutoff flows are explicit exclusions from the system boundary. They may represent inputs or outputs from the foreground that are produced by another foreground system, or they may represent dependencies for which no background process is available. Thus, they are semantically similar to foreground flows. They may or may not make reference to externally defined entities. However, complex foreground models can be published in parts by making shared reference to a set of cutoff flows.
- **Elementary flows**: The best practices for identifying elementary flows are a matter of ongoing discussion and research. At minimum, each elementary flow should be identified in terms of (1) its chemical composition, using a reference nomenclature, CAS number, chemical formula, or other concrete identifier; and (2) the environmental compartment with which it is being exchanged. Geographic locale may also be a part of an elementary flow specification. However, if the flow originates in the foreground, the locale may be inferred from the locale of the originating foreground node.
- **Characterization quantities**: The LCIA characterization method is already required to be unambiguously reported in current LCA practice. If a structured publication includes results, it should also include a link to an authoritative reference for the method(s) computed.

The entity map should include all of the foreground flows, background processes, cutoff flows, elementary flows, and characterization quantities referred to in the publication. Taken together, the entities in the map provide a working formal definition of the scope of the model. Even in the absence of the numerical data that make up the input matrices or aggregation results, this *functional scope* can be used as a basis to describe the contents of the model and formally state the system boundary of the foreground.

The entity map also provides a template for empirically-generated LCA studies: once the flows are enumerated, an observer can simply report the flow magnitudes observed over a reference period, and use it to construct an LCA study.

3 Examples

LCI databases often contain product systems that can be modeled as foreground studies because they describe products that are not required by the background. In this section, two product systems selected from LCI databases are used to illustrate the concept of structured publication. The systems selected because are complex enough to illustrate the premise but simple enough to review easily.

Each system is illustrated as a table that shows the foreground model, cutoff flows, background dependencies and foreground emissions included in the system. Aggregation results $\tilde{\mathbf{x}}$, $\tilde{\mathbf{a}}_d$, and $\tilde{\mathbf{b}}_f$ are also reported. The table omits numeric data from most of the table for clarity. Instead, a black square indicates the presence of a nonzero value. The systems are also reproduced as structured publications in the supplementary materials, formatted as Excel spreadsheets.

3.1 US LCI: Secondary Aluminum

The US LCI database contains a small background system of 39 processes, as well as 395 foreground product systems that range in size from one to 83 foreground flows.

The example product system reports the production of secondary aluminum from automotive scrap (Figure 4). The aluminum production process has direct requirements for two other foreground systems, including transportation services and quicklime production. The quicklime production itself depends on limestone extraction, which is also part of the foreground. The aluminum production requires two background systems, electricity and natural gas combustion, and reports 10 cutoff flows. The main input to the process, "Aluminum scrap, automotive," appears as a cutoff flow. Overall, the four foreground processes require the background for only combustion (five fuels), transport (three modes) and grid electricity. The same natural gas combustion and electricity models were used in all three non-transport foreground nodes.

The system also reports several direct emissions, mainly from the aluminum process. Only one

Aluminum, secondary, ingot, from automotive scrap, at plant [RNA]

(node) Foreground flow	0	1	2	3	
(0) Aluminum, secondary, ingot, from automotive scrap, at plant (kg) [RNA]	⬚				
(1) Aluminum recovery, transport, to plant (kg) [RNA]	1.03	⬚			
(2) Quicklime, at plant (kg) [RNA]	2.35e-05		⬚		
(3) Limestone, at mine (kg) [RNA]			1.87	⬚	
Foreground Node Weights $\tilde{x}$	1	1.03	2.35e-05	4.39e-05	
Input: CUTOFF Disposal, solid waste, unspecified, to sanitary landfill [CUTOFF Flows]	■		■		0.0842
Input: CUTOFF Filter media, at plant [CUTOFF Flows]	■				4.79e-05
Input: CUTOFF Lube oil, at plant [CUTOFF Flows]	■				8.75e-07
Input: CUTOFF Treatment gases, unspecified, at plant [CUTOFF Flows]	■				0.003
Input: CUTOFF Alloying additives, at plant [CUTOFF Flows]	■				0.021
Input: CUTOFF Chemicals, unspecified, used for wastewater treatment [CUTOFF Flows]	■				0.008
Input: CUTOFF Grain refiners, at plant [CUTOFF Flows]	■				0.0003
Input: CUTOFF Treatment salts, unspecified, at plant [CUTOFF Flows]	■				0.009
Input: CUTOFF Packaging, unspecified, at plant [CUTOFF Flows]	■				1.67e-05
Input: CUTOFF Aluminum scrap, automotive [CUTOFF Flows]	■				1.03

Background Dependencies	0	1	2	3	$\tilde{a_d}$
Diesel, combusted in industrial boiler (l) [RNA]			■	■	4.79e-08
Transport, train, diesel powered (t*km) [RNA]		■	■		0.0415
Electricity, at grid, US, 2000 (kWh) [RNA]	■		■	■	0.668
Transport, combination truck, diesel powered (t*km) [RNA]		■	■		0.374
Natural gas, combusted in industrial boiler (m3) [RNA]	■		■	■	0.223
Bituminous coal, combusted in industrial boiler (kg) [RNA]			■	■	4.04e-06
Gasoline, combusted in equipment (l) [RNA]				■	2.25e-09
Liquefied petroleum gas, combusted in industrial boiler (l) [RNA]			■		7.57e-10
Transport, barge, average fuel mix (t*km) [RNA]			■		5.66e-07
Foreground Emissions	**0**	**1**	**2**	**3**	$\tilde{b_f}$
Output: Lead [air, unspecified]	■				2.16e-07
Output: Suspended solids, unspecified [water, unspecified]	■				2.56e-05
Output: Particulates, unspecified [air, unspecified]	■		■	■	1.78e-07
Output: NMVOC, non-methane volatile organic compounds [air, unspecified]	■				4.7e-05
Output: BOD5, Biological Oxygen Demand [water, unspecified]	■				1.35e-09
Output: COD, Chemical Oxygen Demand [water, unspecified]	■				8.3e-07
Output: Sulfur dioxide [air, unspecified]			■		3.52e-09
Output: Dissolved solids [water, unspecified]	■				2.38e-08
Output: carbon dioxide [air, unspecified]			■		1.8e-05
Input: Limestone [resource, ground-]				■	4.39e-05
Output: Heavy metals, unspecified [water, unspecified]	■				2.25e-05
Output: Organic substances, unspecified [water, unspecified]	■				1.29e-06
Output: Acids, unspecified [air, unspecified]	■				4.31e-05

Figure 4: A structured product system model for secondary aluminum, drawn from US LCI. Exchange values are replaced with black squares for clarity.

resource consumption (input flow) is reported – the extraction of limestone. The direct emission modeling is limited in scope and includes a number of "unspecified" flows that may not be well characterized in many impact methods.

The structured publication of the secondary aluminum model includes a total of 47 data points about 36 entities. The A_f, A_d, and B_f matrices are reported in sparse format, along with the aggregation vectors. In the publication, LCIA scores are reported for several impact indicators drawn from the Ecoinvent reference LCIA implementation.[1] The E matrix is also partially reproduced: entries are only reported if they match flows that appear in the foreground.

3.2 Ecoinvent: Organic Potatoes

Ecoinvent version 3 is provided in three different system models that reflect different linking strategies. The example is drawn from the "Allocation at the point of substitution" or APOS model, which includes 11,420 processes that produce 12,966 product flows. Of these, 10,282 are background flows and the rest are foreground flows.

The example system reports production of organic potatoes supplied to the global market (Table 5). The foreground includes nine nodes, of which six (nodes 3 through 8) form a cyclic dependency associated with the production of potato seeds. The Ecoinvent database's use of "markets" as mixer processes is evident in the table: node 0 is a mixer process that combines Swiss ("CH") production (node 1) weighted at 2.4% with rest-of-World ("RoW") production (node 2) weighted at 97.6%. That same market split (2.4 / 97.6) can also be seen in nodes 7 and 8.

Nodes 4 and 5, which make seeds ready "for setting," are mixed by node 8. Although nodes 4 and 5 are geographically distinguished (CH vs RoW), both can be seen to consume potato seed from the global market.

Looking at the dependency and emission lists, the different "signatures" of different kinds of pro-

[1] The Ecoinvent LCIA factors are not published, but are available from the Ecoinvent Centre. The name of the file used in the current study is `LCIA implementation v3.1 2014_08_13.xlsx`.

market for potato, organic [GLO]

(node) Foreground flow	0	1	2	3	4	5	6	7	8	
(0) potato, organic (kg) [GLO]	⬚									
(1) potato, organic (kg) [CH]	0.024	⬚								
(2) potato, organic (kg) [RoW]	0.976		⬚							
(3) potato seed, organic, at farm (kg) [RoW]				⬚				0.976		
(4) potato seed, organic, for setting (kg) [CH]					⬚				0.024	
(5) potato seed, organic, for setting (kg) [RoW]						⬚			0.976	
(6) potato seed, organic, at farm (kg) [CH]							⬚	0.024		
(7) potato seed, organic, at farm (kg) [GLO]					1	1		⬚		
(8) potato seed, organic, for setting (kg) [GLO]		0.11	0.11	0.16			0.16		⬚	
Foreground Node Weights $\tilde{x}$	1	0.024	0.976	0.128	0.00315	0.128	0.00315	0.131	0.131	

Background Dependencies	0	1	2	3	4	5	6	7	8	$\tilde{a_d}$
transport, freight train (metric ton*km) [GLO]	■							■		0.138
irrigation (m3) [CH]		■					■			0.000426
tillage, hoeing and earthing-up, potatoes (ha) [GLO]		■	■	■			■			0.000101
potato planting (ha) [GLO]		■	■	■			■			5.04e-05
solid manure loading and spreading, by hydraulic loader and spreader (kg) [GLO]		■	■	■			■			0.717
electricity, low voltage (kWh) [CH]					■					0.00023
potato grading (kg) [GLO]		■	■	■			■			1.13
transport, freight, light commercial vehicle (metric ton*km) [GLO]	■							■		0.0258
tillage, harrowing, by spring tine harrow (ha) [GLO]		■	■	■			■			5.04e-05
tillage, ploughing (ha) [GLO]		■	■	■			■			5.04e-05
electricity, low voltage (kWh) [GLO]						■				0.00933
transport, freight, lorry, unspecified (metric ton*km) [GLO]	■							■		0.404
green manure, organic, until March (ha) [GLO]		■	■	■			■			5.04e-05
building, multi-storey (m3) [GLO]					■	■				5.24e-06
liquid manure spreading, by vacuum tanker (m3) [GLO]		■	■	■			■			0.000565
application of plant protection product, by field sprayer (ha) [GLO]		■	■	■			■			0.000222
transport, freight, sea, transoceanic ship (metric ton*km) [GLO]	■							■		0.525
potato haulm cutting (ha) [GLO]		■	■	■			■			5.04e-05
transport, tractor and trailer, agricultural (metric ton*km) [GLO]		■	■	■			■			0.00113
harvesting, by complete harvester, potatoes (ha) [GLO]		■	■	■			■			5.04e-05
tillage, currying, by weeder (ha) [GLO]		■	■	■			■			0.000101
transport, freight, inland waterways, barge (metric ton*km) [GLO]	■							■		0.0857
copper oxide (kg) [GLO]		■	■	■			■			0.000107
tillage, harrowing, by rotary harrow (ha) [GLO]		■	■	■			■			5.04e-05
irrigation (m3) [GLO]			■	■						0.0173

Foreground Emissions	0	1	2	3	4	5	6	7	8	$\tilde{b_f}$
Input: Occupation, construction site [natural resource, land]					■	■				2.1e-06
Input: Transformation, from unspecified [natural resource, land]					■	■				1.05e-06
Input: Transformation, to industrial area [natural resource, land]					■	■				1.05e-06
Input: Energy, gross calorific value, in biomass [natural resource, biotic]		■	■	■			■			3.87
Output: Phosphate [water, ground-]		■	■	■			■			6.25e-06
Output: Cadmium, ion [water, ground-]		■	■	■			■			1.14e-09
Output: Zinc, ion [water, ground-]		■	■	■			■			1.35e-06
Output: Nitrate [water, ground-]		■	■	■			■			0.00696
… (31 rows omitted)										

Figure 5: A structured product system model for organic potato production, drawn from Ecoinvent v3.2 (APOS). Exchange values are replaced with black squares for clarity.

cesses can be seen: nodes 1, 2, 3, and 6 are clearly agricultural processes that require irrigation, tillage, manure and so on. Nodes 0 and 7 are visible as market processes, their only requirements being transport processes. Nodes 4 and 5 each consume electricity and require use of a "multi-storey building." Taken together, the CH-locality processes appear to use CH irrigation and electricity supply but are otherwise similar to their RoW counterparts. The product model includes no cutoff flows.

The structured publication of the organic potato model reports the A_f, A_d, and B_f matrices in full rather than sparse format to enable a more visually immediate review. In addition, the E matrix is reported in full, for all emissions found in the model, rather than only the emissions found in the foreground. This enables the publication to be used to visually evaluate different implementations of the same impact assessment method in the context of the inventory model (Ecoinvent vs ILCD implementations of ReCiPe Midpoint marine eutrophication; see supplementary materials for details).

4 Discussion

An LCA foreground study can be published concisely by distinguishing foreground components from background components, explicitly identifying all entities involved in the study, and then reporting the data values that make up the submatrices in Eq. 5. As the examples demonstrate, a structured publication can be created that allows a reader to critically review, validate, and accurately reproduce the product system model without any additional information. The entity map provides the foundation for a linked-data publication by explicitly naming all externally-referenced entities involved in the foreground model. Ultimately it should be possible to name entities using linked data identifiers or URIs, allowing unambiguous, automatic retrieval of metadata about the entities.

The input matrices (A_f, A_d, and B_f) or aggregation results ($\tilde{\mathbf{x}}$, $\tilde{\mathbf{a}}_d$, $\tilde{\mathbf{b}}_f$), together with the entity map, form a research object that can be easily generated and shared with colleagues. By separating foreground and background, the research object satisfies the objective for attribution, because only the information attributable to the study author is included in the research object; whereas information that came from the background database is excluded from the publication.

This framework allows an author a great deal of flexibility in publishing their model. Any of the components or intermediate results visualized in Figure 2 may be selected for publication at any level of detail. By publishing a set of linked data identifiers corresponding to the rows and entries of the matrices, the author ensures that the data user will be able to identify the foreground and determine the background processes used. Again because the background is excluded, the size of the research object is kept down, and licensing limitations associated with the distribution of proprietary data are avoided. Use of the research object to selectively exclude private data is discussed in the supplementary materials.

4.1 Validating LCIA Results

As shown in Eq. 9, the LCIA results for a foreground system are the sum of the foreground and background scores. The foreground scores are computed from the foreground emissions, while the background scores are computed from LCI results included in the background database. These two components require distinct approaches for study validation.

For foreground scores, the LCIA results depend on the information provided in B_f combined with the characterization factors for the LCIA methods selected. Because LCIA methods are maintained independently from LCI databases and LCA software providers, the redundant implementation of these methods by different providers and/or software makers is a significant source of irreproducibility (Speck *et al.*, 2015; Herrmann & Moltesen, 2015). In the context of a structured publication, the most transparent approach is for the publication to include the characterization factors used in the study, which allows a reader to inspect them and check the results.

For background scores, the results cannot be validated without access to the background database. Again, the characterization factors used would need to be disclosed along with the background LCI results. Because background databases are occasionally updated, it is important for a background dependency reference to include information about which version was used. Assuming datasets within a particular background database do not change, it would be possible to create an archive of unit impact scores for each background process and LCIA method supported by the database. These values could then be independently shared and validated.

In the context of structured publication, a reader of the publication could obtain the unit LCIA scores for each background dataset referenced, and use them to validate the background LCIA result. If the background database is subject to licensing restrictions, it may be necessary to require readers or data users to independently obtain background LCIA scores. However, it may also be possible for the author to publish unit impact scores for background processes used in the model without violating licensing terms. In this case, it is much easier to reproduce published results, though the reader may still wish to check the author's published unit scores for accuracy. If unit impact scores for background processes are included in the publication, then it is a simple matter for a reader to perform a contribution analysis and inspect the sensitivity of the results to foreground parameters. This is exemplified in the "organic potato" publication provided in the supplementary materials.

4.2 Avoiding Redundancy in LCA Computation

Considering the discussion of LCIA score reproducibility, it is evident that relying on study authors to individually perform flow characterization, background LCI computation, and background LCIA computation generates a tremendous duplication of effort in the LCA community. Although background LCI and LCIA results are static as long as the background database is not revised, the conventional approach to LCA requires study authors to individually re-compute background inventory results and LCIA scores for every study. This has weighty implications:

- If the database is provided in unit process form, then the practitioner must have access to the entire database in order to compute any LCI result. This means that every licensee must download the full database upon purchasing a license.
- Different LCA software providers must independently ensure that their software can support every LCI database their users may wish to use. This causes redundancy and can introduce errors and inconsistencies.
- Database updates become inconvenient, because every update requires a complete replacement of the database. This inhibits the deployment of incremental updates, even when errors are found.
- Because of the increased burden of data updates, software providers may be slow to apply updates from LCI providers. Additionally, it is more likely that users will postpone updating their software, leading to the possibility of version mismatches between users or software systems.
- Software providers must also independently implement LCIA methods, which introduces the same possibilities of errors, inconsistencies, and version mismatches.

In contrast, a computational approach based on foreground study publication separates the concerns of study authorship from software and database maintenance. In principle, the product system models presented as examples in this paper could remain unchanged even if background databases and LCIA methodologies were completely revised. A study publication could be easily updated simply by changing which entities are pointed to by the entity map. This exercise is well defined and easily reviewed. It would also allow for a straightforward comparison of the changes to study results after a database update, a task that is currently challenging to perform.

A further implication of the linked data approach suggested here is a shift away from stand-alone computation of background data results, and toward a system where authoritative data providers compute aggregation and characterization results as a service for their users. In such a system, data users would supply a query in the form of a particular reference to a background process or elementary flow, and service providers would answer the query with LCI results, LCIA results, or characterization factors. Answers could be provided on-demand to study authors, readers, reviewers, or downstream data users.

This approach, which could be termed "aggregation by reference" and "characterization by reference," could dramatically reduce the complexity of authoring LCA studies, in principle allowing authors to construct whole product system models without procuring any data. Before computation of results could be performed, authors would have to obtain access to the desired datasets. It would also situate the data providers as the sole authoritative sources for results derived from their data. This has the dual effects of giving readers increased confidence in the correctness of the results, and in ensuring that the data providers continue to occupy an indispensible role in study preparation and review.

Finally, it would allow data providers to directly control access to their intellectual property. A given query for an aggregation result could be answered in different ways depending on the credentials of the user making the query, and depending on the context of the query. For example, an authorized critical reviewer of a published study may be given free access to the requested dataset, whereas a member of the general public would be given only the impact score.

4.3 Interlinking Foreground Models

Current practice does not provide a way for study authors to easily provide their models to collaborators, or for independently created models to be linked together. However, structured publication of foreground studies could allow independent authors to work together on complex models. Large foregrounds can be easily partitioned into fragments simply by selectively including or excluding columns of A_f, A_d, and B_f in different publications, allowing for independent maintenance of different subsystems. Similarly, an isolated product system described in a structured format could easily be combined with other models of complementary systems to form a larger system. Independent foreground systems that both reference the same shared cutoff flows could be automatically linked without any intervention by the authors.

In both cases, the interlinking of foreground models rests on the principle of shared reference to common entities. Linked semantic data provides an enabling technology, both allowing standalone foreground publications to be automatically linked to data resources and to one another, and by providing interpretive services such as dataset discovery, flow matching, unit conversion, and metadata retrieval. This represents an opportunity for further research and development.

5 Conclusion

An LCA study, as formulated in Eq. 5, can be published in a structured format by providing data users with sufficient information to unambiguously construct the foreground matrices A_f, A_d, and B_f, and also to clearly identify the meaning of the nonzero row and column entries in each matrix. This can be accomplished by providing sparse, tabular data that make up the matrices, and an entity map which relates the rows and columns to references on the Web. A study author can provide varying degrees of transparency in reporting their study, but the structured publication format ensures that a reader or data user can interpret the information in the way the author intended. The entity map also serves as a formal statement of the functional scope of the study.

Before the goal of easily reproducible foreground models can be realized, the linked data foundation must be laid. The four entity classes found in foreground studies, namely foreground flows, background processes, elementary flows, and characterization quantities, must shift from being recorded locally to being stored on the Semantic Web, where their shared meaning can be curated. Fortunately, this is easy to do: authoritative sources simply need to maintain static hyperlinks to their content, so that licensees and data users can make shared reference to the same entities, even if the data may not be accessible to the general public. Similarly, impact assessment providers should publish their characterization data in a way that allows for automatic query and validation.

Both of these tasks can be accomplished independently through the efforts of members of the LCA community; however, the accuracy and validity of the data would be improved if the work were done by the data providers themselves. A shared semantic framework for foreground flows would also support collaborative development of inventory data resources.

Foreground studies are distinct from process inventory data sets, about which there is much ongoing work to establish file formats and standards. To avoid the need for a similar enterprise for foreground publications, a minimal, inclusive standardization of information should be sought that makes the fewest possible prescriptions on its content. The publication model presented here is meant to identify the minimal requirements that a study can be both reproduced and interpreted accurately: an *entity map*, a list of references to entities used in the study; and *tabular data* to fill in the model in Eq. 5. From such a simple prescription, a framework for distributed, platform-independent LCA computation can be imagined.

Acknowledgments

This work draws on the efforts of the “Inventory Model Description and Revision” discussion group of SETAC North America, whose members include Ben Mourad Amor (University of Sherbrooke), Bill Bernstein (NIST), Paula Bernstein (PRé Sustainability), Marcos Esterman (Rochester Institute of Technology), David Evers (Hexion), Karl Haapala (Oregon State University), Troy Hawkins (Eastern Research Group), Wesley Ingwersen (US EPA), Christoph Koffler (Thinkstep), Lise Laurin (EarthShift Global), Antonino Marvuglia (Luxembourg Institute of Science and Technology), David Meyer (US EPA), KC Morris (NIST), Christopher Mutel (Paul Scherrer Institut), Tomas Navarrete (Luxembourg Institute of Science and Technology), Devarajan Ramanujan (Massachusetts Institute of Technology), Barclay Satterfield (Eastman Chemical), and the author. This work was supported by the US National Science Foundation (CCF-1442966).

References

Bapat, R.B., & Ghorbani, E. 2014. Inverses of triangular matrices and bipartite graphs. *Linear Algebra Appl.*, **447**(0), 68–73. Special Issue - Memorial Issue for Michael Neumann and Uriel Rothblum. doi:10.1016/j.laa.2013.03.002.

Bechhofer, Sean, Buchan, Iain, De Roure, David, Missier, Paolo, Ainsworth, John, Bhagat, Jiten, Couch, Philip, Cruickshank, Don, Delderfield, Mark, Dunlop, Ian, Gamble, Matthew, Michaelides, Danius, Owen, Stuart, Newman, David, Sufi, Shoaib, & Goble, Carole. 2013. Why linked data is not enough for scientists. *Future Generation Computer Systems*, **29**(2), 599–611. doi:10.1016/j.future.2011.08.004.

Bizer, Christian, Heath, Tom, & Berners-Lee, Tim. 2009. Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, **5**(3), 1–22. doi:10.4018/jswis.2009081901.

Brandão, Miguel, Clift, Roland, Cowie, Annette, & Greenhalgh, Suzie. 2014. The Use of Life Cycle Assessment in the Support of Robust (Climate) Policy Making: Comment on "Using Attributional Life Cycle Assessment to Estimate Climate-Change Mitigation ...". *Journal of Industrial Ecology*, **18**(3), 461–463. doi:10.1111/jiec.12152.

Buckheit, Jonathan B., & Donoho, David L. 1995. WaveLab and Reproducible Research. *Pages 55–81 of:* Antoniadis, Anestis, & Oppenheim, Georges (eds), *Wavelets and Statistics*. New York, NY: Springer New York.

Ciroth, Andreas, & Srocka, Michael. 2014. *OpenLCA Schema*. https://github.com/GreenDelta/olca-schema.

Cleary, Julian. 2009. Life cycle assessments of municipal solid waste management systems: a comparative analysis of selected peer-reviewed literature. *Environment international*, **35**(8), 1256–1266. PMID: 19682746. doi:10.1016/j.envint.2009.07.009.

Clift, R., Frischknecht, R., Huppes, G., Tillman, A.-M., & Weidema, B. 1998. *Toward a Coherent Approach to Life Cycle Inventory Analysis. Report of the Working Group on Inventory Enhancement*. Tech. rept. SETAC-Europe.

Curran, Mary Ann, & Young, Steven B. 2014. Critical review: a summary of the current state-of-practice. *Int. J. Life Cycle Assess.*, **19**(10), 1667–1673. doi:10.1007/s11367-014-0778-2.

Fet, Annik Magerholm, & Skaar, Christofer. 2006. Eco-labeling, Product Category Rules and Certification Procedures Based on ISO 14025 Requirements (6 pp). *The International Journal of Life Cycle Assessment*, **11**(1), 49–54. doi:10.1065/lca2006.01.237.

Fomel, Sergey, & Claerbout, Jon F. 2009. Guest Editors' Introduction: Reproducible Research. *Computing in Science & Engineering*, **11**(1), 5–7. doi:10.1109/mcse.2009.14.

Heath, Garvin A., & Mann, Margaret K. 2012. Background and Reflections on the Life Cycle Assessment Harmonization Project. *Journal of Industrial Ecology*, **16**, S8–S11. doi:10.1111/j.1530-9290.2012.00478.x.

Heijungs, Reinout, & Suh, Sangwon. 2002. *The computational structure of life cycle assessment*. Vol. 11. Springer Science & Business Media.

Henriksson, Patrik John Gustav, Guinée, Jeroen Bartholomeus, Heijungs, Reinout, Koning, Arjan, & Green, Darren Michael. 2014. A protocol for horizontal averaging of unit process data-including estimates for uncertainty. *The International Journal of Life Cycle Assessment*, **19**(2), 429–436. doi:10.1007/s11367-013-0647-4.

Herrmann, Ivan T., & Moltesen, Andreas. 2015. Does it matter which Life Cycle Assessment (LCA) tool you choose? – a comparative assessment of SimaPro and GaBi. *Journal of Cleaner Production*, **86**(jan), 163–169. doi:10.1016/j.jclepro.2014.08.004.

Ingwersen, Wesley W., Hawkins, Troy R., Transue, Thomas R., Meyer, David E., Moore, Gary, Kahn, Ezra, Arbuckle, Peter, Paulsen, Heidi, & Norris, Gregory A. 2015. A new data architecture for advancing life cycle assessment. *Int. J. Life Cycle Assess.*, **20**(4), 520–526. doi:10.1007/s11367-015-0850-6.

ISO 14044. 2006. *Environmental management — Life cycle assessment — Requirements and guidelines*. ISO, Geneva, Switzerland.

Janowicz, K., Krisnadhi, A., Hu, Y., Suh, S., Weidema, B., Rivela, B., Tivander, J., Meyer, D.E., Berg-Cross, G., Hitzler, P., Ingwersen, W., Kuczenski, B., Vardeman, C., & Ju, Y. 2015 (October). A Minimal Ontology Pattern for Life Cycle Assessment Data. *In: Proceedings of 6th Workshop on Ontology and Semantic Web Patterns (WOP2015) co-located with the 14th International Semantic Web Conference (ISWC 2015)*.

Khan, Huda, Caruso, Brian, Corson-Rikert, Jon, Dietrich, Dianne, Lowe, Brian, & Steinhart, Gail. 2011. DataStaR: Using the Semantic Web approach for Data Curation. *International Journal of Digital Curation*, **6**(2), 209–221. doi:10.2218/ijdc.v6i2.197.

Kuczenski, Brandon. 2015. Partial ordering of life cycle inventory databases. *The International Journal of Life Cycle Assessment*, **20**(12), 1673–1683. doi:10.1007/s11367-015-0972-x.

Kuczenski, Brandon, Davis, Christopher B., Rivela, Beatriz, & Janowicz, Krzysztof. 2016. Semantic catalogs for life cycle assessment data. *Journal of Cleaner Production*, **137**(nov), 1109–1117. doi:10.1016/j.jclepro.2016.07.216.

Laurent, Alexis, Clavreul, Julie, Bernstad, Anna, Bakas, Ioannis, Niero, Monia, Gentil, Emmanuel, Christensen, Thomas H., & Hauschild, Michael Z. 2014. Review of LCA studies of solid waste management systems – Part II: Methodological guidance for a better practice. *Waste Management*, **34**(3), 589–606. doi:10.1016/j.wasman.2013.12.004.

Lloyd, Shannon M., & Ries, Robert. 2007. Characterizing, Propagating, and Analyzing Uncertainty in Life-Cycle Assessment: A Survey of Quantitative Approaches. *Journal of Industrial Ecology*, **11**(1), 161–179. doi:10.1162/jiec.2007.1136.

Madin, Joshua S., Bowers, Shawn, Schildhauer, Mark P., & Jones, Matthew B. 2008. Advancing ecological research with ontologies. *Trends in Ecology & Evolution*, **23**(3), 159–168. doi:10.1016/j.tree.2007.11.007.

Majeau-Bettez, Guillaume, Wood, Richard, & Strømman, Anders Hammer. 2014. Unified Theory of Allocations and Constructs in Life Cycle Assessment and Input-Output Analysis. *Journal of Industrial Ecology*, **18**(5), 747–770. doi:10.1111/jiec.12142.

McManus, Marcelle C., Taylor, Caroline M., Mohr, Alison, Whittaker, Carly, Scown, Corinne D., Borrion, Aiduan Li, Glithero, Neryssa J., & Yin, Yao. 2015. Challenge clusters facing LCA in environmental decision-making—what we can learn from biofuels. *The International Journal of Life Cycle Assessment*, **20**(10), 1399–1414. doi:10.1007/s11367-015-0930-7.

Mesirov, J. P. 2010. Accessible Reproducible Research. *Science*, **327**(5964), 415–416. doi:10.1126/science.1179653.

Mila i Canals, L., Asselin, A.-C., Cruypenninck, H., Liu, S., Ingwersen, W., & Macdonald, T. 2015. Towards a Global Network of Interoperable LCA Databases. *In: Proceedings of SETAC Europe 25th Annual Meeting (SETAC Europe 2015)*.

Modahl, Ingunn Saur, Askham, Cecilia, Lyng, Kari-Anne, Skjerve-Nielssen, Camilla, & Nereng, Guro. 2012. Comparison of two versions of an EPD, using generic and specific data for the foreground system, and some methodological implications. *The International Journal of Life Cycle Assessment*, **18**(1), 241–251. doi:10.1007/s11367-012-0449-0.

Pauliuk, Stefan, Majeau-Bettez, Guillaume, & Müller, Daniel B. 2015. A General System Structure and Accounting Framework for Socioeconomic Metabolism. *Journal of Industrial Ecology*, **19**(5), 728–741. doi:10.1111/jiec.12306.

Pelletier, Nathan, Ardente, Fulvio, Brandão, Miguel, Camillis, Camillo De, & Pennington, David. 2014. Rationales for and limitations of preferred solutions for multi-functionality problems in LCA: is increased consistency possible? *The International Journal of Life Cycle Assessment*, **20**(1), 74–86. doi:10.1007/s11367-014-0812-4.

Plevin, Richard J., Delucchi, Mark A., & Creutzig, Felix. 2013. Using Attributional Life Cycle Assessment to Estimate Climate-Change Mitigation Benefits Misleads Policy Makers. *Journal of Industrial Ecology*, **18**(1), 73–83. doi:10.1111/jiec.12074.

Rainville, Anne, Hawkins, Richard, & Bergerson, Joule. 2015. Building consensus in life cycle assessment: the potential for a Canadian product category rules standard to enhance credibility in greenhouse gas emissions estimates for Alberta's oil sands. *Journal of Cleaner Production*, **103**(sep), 525–533. doi:10.1016/j.jclepro.2014.06.067.

Reap, John, Roman, Felipe, Duncan, Scott, & Bras, Bert. 2008. A survey of unresolved problems in life cycle assessment Part 1: goal and scope and inventory analysis. *Int. J. Life Cycle Assess.*, **13**(4), 290–300.

Recchioni, Marco, Mathieux, Fabrice, Goralczyk, Malgorzata, & Schau, Erwin M. 2013. *ILCD Data Network and ELCD Database: current use and further needs for supporting Environmental Footprint and Life Cycle Indicator Projects*. Tech. rept. EUR 25744 EN. European Commission, Joint Research Centre, Luxembourg.

Speck, Ricky, Selke, Susan, Auras, Rafael, & Fitzsimmons, James. 2015. Life Cycle Assessment Software: Selection Can Impact Results. *Journal of Industrial Ecology*, **20**(1), 18–28. doi:10.1111/jiec.12245.

Subramanian, Vairavan, Ingwersen, Wesley, Hensler, Connie, & Collie, Heather. 2012. Comparing product category rules from different programs: learned outcomes towards global alignment. *The International Journal of Life Cycle Assessment*, **17**(7), 892–903. doi:10.1007/s11367-012-0419-6.

Suh, Sangwon, Weidema, Bo, Schmidt, Jannick Hoejrup, & Heijungs, Reinout. 2010. Generalized Make and Use Framework for Allocation in Life Cycle Assessment. *J. Ind. Ecol.*, **14**(2), 335–353. doi:10.1111/j.1530-9290.2010.00235.x.

Turconi, Roberto, Boldrin, Alessio, & Astrup, Thomas. 2013. Life cycle assessment (LCA) of electricity generation technologies: Overview, comparability and limitations. *Renewable and Sustainable Energy Reviews*, **28**(dec), 555–565. doi:10.1016/j.rser.2013.08.013.

UNEP/SETAC. 2011. *Global Guidance Principles for Life Cycle Assessment Databases*. Tech. rept. United Nations Environment Programme.

van der Harst, Eugenie, & Potting, José. 2013. A critical comparison of ten disposable cup LCAs. *Environmental Impact Assessment Review*, **43**(nov), 86–96. doi:10.1016/j.eiar.2013.06.006.