AMERICAN CENTER FOR LIFE CYCLE ASSESSMENT

Requirements for the US LCI Database

Input pursuant to the Camp Long Declaration

01-Mar-10

This is the consensus work of a group of volunteer LCA experts over a year of meetings January – December 2009. It is intended that this input assist the National Renewable Energy Laboratory in its efforts to upgrade the US LCI database.

Contents

Requirements of the US LCI Database

Introduction

The US Life Cycle Inventory Database (US LCI DB) is planning major upgrades, to increase the quality and quantity of the data it holds, and to become more useful for a wider variety of stakeholders. In order to accomplish its goal, major revamping of the database format, quality systems and management systems is necessary. This document lays out the technical and business requirements of the US LCI DB's software implementation.

This document represents one of the outputs of a year-long open effort sponsored by the American Center for Life Cycle Assessment to assist the US LCI Database. It is an outcome of the Camp Long Declaration, and the work of a group of volunteers, listed below.

Alberta Carpenter, NREL

Andreas Ciroth, GreenDelta

Nuno da Silva, PE-Americas

Michael Deru, NREL

Tom Gloria, Industrial Ecology

Mark Goodkoep, PRė Consulting

Neil Huizenga, Microsoft

Pascal LeSage, Sylvatica

Jeff Rice, University of Arkansas

Rita Schenck, IERE

We wish to thank them for their hard work in developing this document.

Requirements not Covered

Following is a list of requirements areas that will not be covered by this document. Where available, the correct reference document is called out.

1. Data quality – This is covered by the data quality plan
2. Data input and output – This is covered by the Input and Output Interface requirements documents
3. Maintenance and ownership
4. Data transfer formats

Definitions

Term	Definition	Reference
Attributional LCA	Attributional LCA (ALCA) provides information about the impacts of the processes used to produce (and consume and dispose of) a product, but does not consider indirect effects arising from changes in the output of a product.	Brander, et al
Consequential LCA	Consequential LCA seeks to analyse [sic]the changes induced by the decision. It aims at describing the environmentally relevant physical flows to and from a life cycle and its subsystems, which are influenced by the decision product.	Ciroth, A., Lundie, S., Huppes, G.: UNEP-SETAC Life Cycle Initiative, Life Cycle Inventory (LCI), Task Force 3, Methodological Consistency: Inventory methods in LCA: towards consistency and improvement. VDM-Verlag, Saarbrücken, 2008, p. 9)
The System	We are using the term "system", "the system", etc. to refer to the database structure and software built to manage the database.	

Business Requirements

The following section lists high level business requirements. These are here to describe the needs of the business and/or users in an implementation agnostic fashion.

1. The database must be able to store LCI data for all unit processes relevant to the US economy
   a. This includes data for processes outside the US
2. The database must be able to export to, and import from, the following existing LCI data formats:
   a. EcoSpold v1
   b. EcoSpold v2
   c. ELCD
   d. The International Reference Life Cycle Data System (ILCD)
3. The database format must be compatible with to ISO/TS 14048 compliant transfer formats
4. The database must serve the needs of the entire ISO 14000 series of standards
5. The database must support consequential LCA
6. The database must use existing international standards wherever possible
7. The database should use existing, published systems to define inclusive nomenclature (see ISO/TS 14048)
8. The database should select freely available systems and standards for nomenclature definition in preference over those that are not free
9. The database must contain sufficient metadata to evaluate changes in data quality
10. The database must be easily updated as the needs of the business change
11. The database needs to be usable by anyone on any platform
12. The database should contain both unit process data and cradle-to-gate

Functional Requirements

The following section describes the functional needs of the system as they relate to the business requirements. Each functional requirement maps to one or more business requirements and must be ranked in order of importance; the lower the rank in the table below, the higher the priority. Details of each requirement are given in the section Detailed Requirements.

Rank	Id	Description	Business Req.
10	R-0100	The system must provide a mechanism for change management	10
20	R-0200	The system must be accessible via open data access methods	11
30	R-0300	The system must contain statistical metadata for each flow	9, 14
40	R-0400	The system data format must conform to ISO/TS 14048	3, 4
50	R-0500	The system data format must be portable to the EcoSpold (v1 and v2) and ILCD formats	2
70	R-0700	The system should use the UNSPSC for identifying technosphere flows	7, 8
80	R-0800	The system must use international standards for field data formats	6, 7
90	R-0900	The database must be able to store data for any unit process regardless of locale, language, and scope	1
100	R-1000	The system must use S.I. units for all units of measure	7
110	R-1100	The system must use CASR numbers for identifying elementary flows	7
120	R-1200	The system must contain data to assess the technological applicability of a unit process	

Detailed Requirements

R-0100: The system must provide a mechanism for change management

In order to support changes in data format, the system will provide an API for access. This API would provide a seamless façade for users of the database to ensure that changes in schema are not too disruptive. By providing this layer of indirection, changes may be made to the underlying database without instantly disrupting all applications dependent on the database.

R-0200: The system must be accessible via open data access methods

The system should be accessible online, and via well-known application programming interfaces (API).

R-0300: The system must contain statistical metadata for each flow

In order to determine statistical uncertainty, some data needs to be included about each flow for each unit process. The table below outlines all the data needed to produce confidence intervals for existing data, which can then be used for validation.

Name	Comments
Number of samples	NULL value indicates that sample size is unknown.
Maximum result	
Minimum result	
Mean	
Standard deviation	
Median	
Statistical distribution	One of: Unknown, Normal, Log-Normal, Uniform, Other. Value of "Other"

Name	Comments
	requires entry of description of distribution.

R-0400: The system data format must conform to ISO/TS 14048

The ISO/TS 14048 provides a definition for a minimal and maximal definition of a data structure for the sharing and storage of LCI data. In order to properly implement the standard, proper definition of the inclusive nomenclature is required.

Nomenclature reference	Name	List of values or Reference to specification
7.3(a)	Process description – Quantitative reference – Type	The quantitative reference will be implemented as a reference to the output flow that represents the functional unit only.
7.3(b)	Process description – Technical scope	Two values: start and end. Start is to be one of Cradle, Gate. End is to be one of Gate or Grave.
7.3(c)	Process description – Valid geography – Area name	Listing of valid 2-letter country codes (per ISO 1366-1)
7.3(d)	Process description – Valid geography – GIS reference	A binary GIS shape file. http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf
7.3(e)	Inputs and outputs – Group	TBD
7.3(f)	Inputs and outputs – Receiving environment specification	See R-0410 and R-0420 for details.
7.3(g)	Inputs and outputs – Name – Reference to nomenclature	Use either UNSPSC or CASRN. See R-0700 and R-1100 for details.
7.3(h)	Inputs and outputs – Amount – Name	TBD
7.3(i)	Inputs and outputs – Amount – Unit – Symbol or name	Use the SI per ISO 31.
7.3(j)	Inputs and outputs – Amount – Parameter – Name	See R-0300 for a listing of statistical parameters to be included.
7.3(k)	Modeling and validation – Modeling principles – Modeling constants – Name	TBD
7.3(l)	Modeling and validation – Validation – Method	
7.3(m)	Units	

R-0410: Definition of receiving and providing environments

The following values are to be used for all receiving and providing environments and serve as the definition of ISO/TS 14048 7.3(f):

1. Air
   a. Urban air
   b. Rural air
2. Water
   a. Fresh water
      i. Surface fresh water
      ii. Groundwater
   b. Salt water
3. Ground
   a. Deep earth
   b. Surface soil
   c. Sediment

R-0420: Definitions of land use

All land use shall be described using the USGS land cover codes. The USGS codes are represented by a two digit unsigned integer. See the USGS LULC Data Users Guide for more details.

For compatibility with other systems, there will be a need for mapping between the USGS values and those used in CORINE or other formats.

R-0430: Definition of processes

Processes will be defined using the following fields:

Field	Description	Type
Location	Describes the location at which the unit process operates.	List of addresses, UTM coordinates and/or Latitude/longitude coordinates
Is full stage	Indicates whether the process represents an entire life cycle stage or not.	Boolean
Stages	A listing of the life cycle stages encompassed by the process.	List of Enumeration: { Extraction, Manufacture, Transport, Use, End of life }
Aggregation	Describes the type of aggregation used in defining the process.	Enumeration: { Vertical, Horizontal, Both }
Flow chart	A file that contains a flow chart of the inner workings of the process. This field is optional.	BLOB
Technologies	The technologies that represent the sub-steps of a unit process.	See R-1200 for more details.

R-0431: Defining the Reference flow

The reference flow for a process is always one of the outputs of the process (see R-0440) and must be expressed using a unitary measure, e.g. *one metric ton* of carbon steel.

R-0432: Definition of process modeling and validation data

The modeling and validation data for a process will be defined using the following fields:

Allocation

The database will be composed of unit processes, which are by definition not allocated. In order to support different allocations and different used of the inventory data, it is only necessary that fields be provided to key on other methods. Thus, for example fields must be provided for the mass of the flows, and the technosphere flows are represented using UNSPSC codes, which permit one to key on NAICS codes for economic allocation as well as mass allocation. Fields must also be provided for other physical units, as appropriate, e.g. area.

Mathematical models

The mathematical models for unit processes are simple ones, such as mass and energy balances, and they are described in the data quality document.

R-0433: Definition of process administrative data

The administrative data for a process will be defined using the following fields:

Provider's contact information. See R-0435.

R-0434: Definition of validation data

Each unit process will have validation data associated with it. This data will be populated by administrators of the database and include the following information:

Field	Description	Type
Verifier	Contact information for the person that verified the process data.	See R-0435.
References checked	The percentage flows for which the references were verified.	Real
Re-calculations	The percentage of the flows that were re-calculated independently by the verifier.	Real
Consistency with published	The percentage of the flows for the process that were verified using other published data.	Real

R-0435: Definition of a appropriate contact information

A person must include the full name, physical address, email address, telephone number, and a listing of credentials.

R-0440: Definition of inputs and outputs

Inputs and outputs, or flows, will be defined using the following data fields:

Field	Description	Type
Statistical information	Statistical metadata data. See R-0300 for details.	Complex
Flow sphere	What sphere the flow is	Enumeration: { Ecosphere,

	connected to.	Technosphere Sociosphere}
Flow direction	Whether the flow is entering or leaving the process boundary.	Enumeration: { Input, Output }
Collection dates	A date range representing the time over which the data for this flow was collected.	Date Range
Applicable dates	A date range representing the time over which the data is expected to be valid.	Date Range
Technical coverage	Describes the technique used in collection or aggregation.	Enumeration: { Unknown relative performance, Average for technology, Best case example, Worst case example }
Data source	Describes the source of the data. If it is not "Unknown" or "Primary", then the *Data source reference* field is required.	Enumeration: { Primary, Secondary, Tertiary, Unknown }
Data source reference	Describes the source of the data. This field is required if and only if the *Data source* field is set to "Secondary" or "Tertiary".	Citation
Calculation model	Describes the techniques and model used to calculate the value of the input or output.	BLOB
Internal validation	Describes the type of internal validations that were done for the input data.	Enumeration: { Mass balance, Recalculation, Check with other study, Proof-reading, Re-sampling, Other }
Internal validation description	Additional detail about the internal validation performed. This field is required when the value of *Internal validation* is either "Check with other study" or "Other".	Text
Quantity	The amount of the substance.	Real
Substance	The substance the flow represents.	See R-0441 for details.
Technologies	The technology or technoplogies that this flow is related to.	See R-1200 for details.

R-0441: Defining a substance

A substance is the *what* portion of a flow definition and is defined using the following data:

Field	Description	Type
Product code	The UNSPSC code for the product produced or consumed. This is only required for Technosphere flows.	String
CAS Number	The CASRN number for the compound emitted or	String

	consumed. This is only required for Ecosphere flows.	
Attributes	A listing of key/value pairs or *tags* [1]that further describe the product produced or consumed. This is only required for Technosphere flows.	List of key/value pairs

R-0500: The system data format must be portable to the EcoSpold and ILCD format

This section describes the method for mapping between EcoSpold or ILCD, and the US LCI format.

R-0700: The system should use the UNSPSC for identifying technosphere flows

The UNSPSC is to be used for all technosphere inputs and outputs.

R-0800: The system must use international standards for field data formats

When not formatted as binary data, all data fields must be formatted as the appropriate string representations, in accordance with relevant ISO standards.

R-0900: The database must be able to store data for any unit process regardless of locale, and scope

This is largely covered in the R-0400 section. Note: the language of the database is English.

R-1000: The system must use S.I. units for all units of measure

All units shall be S.I. units as laid out in ISO 31.

R-1100: The system must use CASR numbers for identifying elementary flows

All elementary (i.e. ecosphere) flows must have the most accurate CASRN associated with them.

R-1200: The system must contain data to assess the technological applicability of a unit process

Each process must reference one or more technologies. Each technology will consist of a description, and information about its level of development and maturity. Additionally, technologies will be classified into groups to ease discoverability. When more than one technology is included with a unit process, each flow for the process must also be associated with one or more technologies.

The fields that define a technology are:

Field	Description	Type
Name	A unique name for the technology	String
Description	A complete description of the technology, its outputs and inputs. This information must be sufficient for a layperson to identify a technology in use.	String
Development date	The date that the technology entered the early development phase of use. This phase includes development and experimental up-scaling of the technology.	Date
Early adoption date	The date that the technology entered the early adoption	Date

[1] A tag is effectively key/value pair with the key of the tag, and the value of "true"

Field	Description	Type
	phase of use. This is the date of early use at scale, or the beginnings of widespread use.	
Mature date	The date that the technology entered the mature phase of use. Mature technologies are assumed to change slowly until they are phased out.	Date
Phase out date	The date at which the technology begins to be phased out.	Date

This requirement is directly related to the data quality requirements for use in determining the data quality of a data source.

References

Consequential and Attributional Approaches to LCA: a Guide to Policy Makers with Specific Reference to Greenhouse Gas LCA of Biofuels by Brander, M., Tipper, R., Hutchison, C., Davis, G. (http://www.ecometrica.co.uk/wp-content/uploads//2009/05/consequential-and-attributional-approaches-to-lca3.pdf)

USGS Land Cover Data Set Users Guide: http://www.vterrain.org/Culture/LULC/Data_Users_Guide_4.html

Data Quality Plan for the US LCI Database

1. Introduction

The US Life Cycle Inventory Database is in the process of upgrading to develop more and better data that is fully validated. The appropriate level of quality and quality control for data depends entirely on the use to which the data is being put. This is as true for life cycle inventory data as for any other kind of data. Of course, one wishes always to have complete, accurate and precise data, but this is rarely the case in life cycle studies, where data gaps are the rule rather than the exception, and where the cost of getting new data can be very high.

This document describes an approach to the quality measurement and management portion of the overall US LCI Database plan. It begins with a formulation of low, medium and high levels of decision-support and examples of the potential uses of the data at these levels. Then it describes how to measure the quality of a given unit process, and how to validate that quality.

Data quality refers to the reliability and statistical variability of each ecosphere and technosphere flow. Sometimes it is possible to know the absolute numbers for technosphere flows, e.g. the number of units produced over the life of a facility. But one typically has figures that are a sample, not a full accounting. Thus, one can measure production in three consecutive years, or production from several locations, and then these sample results can be analyzed statistically. This is especially useful for analyzing ecosphere emissions, because there are very strong statistical tests for evaluating the summation of emissions that are a key feature of life cycle impact assessment.

2. Uses of Life Cycle Inventory Data

Unit process inventories are the building blocks of any life cycle assessment. They describe and characterize both the flows to and from the ecosphere, and the flows that connect to another unit processes (technosphere flows). A full life cycle assessment links together all the unit processes until there are no technosphere processes left except the flows to the economy (the product(s) or service(s) being sold). The full life cycle study can give us information about the environmental impacts implied by any decision. It truly is a holistic yardstick of environmental performance.

2.1. Scope of decisions:

Life Cycle Assessments support several different kinds of decisions, and the needed level of quality of the data depends on the scope of those decisions. We have identified three levels.

2.1.1. Level one:

Decisions that affect few outcomes, a small area or few people:

Examples: Process modifications, screening studies, design for environment projects, educational illustration, single studies, and single location environmental management systems.

2.1.2. Level two: Decisions that affect a moderate number of outcomes, a moderate area or many people

Examples: Environmental marketing, product stewardship programs (especially for widely-used products), corporate policy making, local ordinances and sector-specific regulation.

2.1.3. Level three: Decisions that affect many and cover a large area

Examples: international, state and national policy-making, legislation and regulation of the entire economy.

3. Data Quality Parameters

The data quality parameters of interest are defined in the ISO 14040 series standards:

- Age
- Geographic appropriateness
- Technological appropriateness
- Completeness
- Availability of statistical data (e.g. number of samples, distribution and standard deviation)
- Validation

The technology represented by a given unit process is a pivotal point for all LCI data quality characteristics. It can be characterized by its stage in adoption (early, mature or phase out), and the normal life span of a particular process, which may be on the order of a decade (example: chip manufacturing technology), a century (example: automated glass blowing technology), or millennia (example: agriculture). For unit processes that represent a unit process in the chemical engineering, single step sense of the word (i.e., not system processes), the technology is inherent to the unit process. However, many if not most LCA unit processes are derived from multi-step processes, and the ISO standards define the unit process as the lowest level at which data are available. Although this flexibility is a strength of LCA, it creates a challenge in characterizing data quality, especially as it relates to technological appropriateness.

For a single step unit process during its mature phase of adoption, changes in process efficiency are slow, and the age of the data is important only when they represent a substantial percentage of change of the mature phase of the technology. For example, chip manufacturing technology primary data figures are of excellent quality and can be considered current when they are two years or less old, of good quality and slightly aged when they are three to five years old, of fair quality but old when they are five to seven years old, and unacceptable when over seven years old.

For mature technologies that represent several combined steps, the situation is much more difficult to evaluate. For example, the direct environmental impacts of agricultural tillage practices can be considered to be constant per unit area, while the impact of the use of pesticides changes decade to decade as new chemicals are developed. Where unit processes represent combined steps, the technological characteristics (age, lifetime and maturity) must be specified for each flow. The overall level of quality for such unit processes must be evaluated with an eye to the relative impacts of each flow. Thus, for example, it is important to keep track of actual pesticides used, but it may not be as important to know the actual kind of fertilizer used (e.g. urea vs. ammonia) because the environmental impacts of producing and using these different nitrogen sources are very similar. The only way to evaluate the quality of these data is to calculate the sensitivity of the potential impacts to expected changes in inputs of the materials in question.

For any given unit process, there will be both ecosphere and technosphere data, and that data may be derived from any of the following sources. Each of these data sources has certain quality characteristics that speak to the requirements of the ISO standards.

1. Primary data collected for the study is always appropriate, as long as the technology in question is current. When it was collected for a previous purpose, it may become inappropriate as it ages.
2. Data based on first principles are always appropriate, as long as the technology in question is current.
3. Emissions factors and other similar data are always appropriate as long as the technology is current.
4. Statistical data is appropriate only as long as it matches the technology.
5. Estimates based on professional judgment are appropriate only when they represent a small proportion of the expected impacts.

The table below lists the quality measurement and required level of quality for each of these parameters for different decision scopes. They represent the system requirements, but level 1 can also be viewed as minimum requirements for new data in the database.

At issue is the availability and accuracy of inventory data relative to the proposed use. Since the database includes many different kinds of unit processes with different quality characteristics, these parameters can only be evaluated in use. No data should be allowed into the database unless the appropriate meta-data for a level 1 analysis is included and the data validated.

Matrix for data quality estimates for system processes

	Measure/ units	Level 1	level 2	level 3
Age	Years collected, technology stage and lifetime	Aged for < 50% of flows	Current for > 50% of flows (Slightly aged)	Current for > 90% of flows
Data Source	Which of 5 noted above	All flow data sources must be documented	Professional judgment limited to <5% of energy, mass or toxics	Professional judgment limited to <1% of energy, mass or toxics
Geographic appropriateness/identity	National or smaller unit ID'd	Geography Must be known	OECD/non-OECD	Same country/region
Technological appropriateness	Unit process description	Must be clear	Relevant to particular product/process, but need not represent the average	Must represent average or median data for age & technology
Completeness (mass balance, process knowledge)	% flows missing on a mass basis	<10	<5	<1
Availability of statistical data	% available on a mass basis	0	>30	>60
Validation of Data	% fields validated	>25	> 50	>90
Validator's qualifications		Self-declared	LCACP or certified validator	LCACP or certified validator

Note: the figures in this table are not fixed, but are a first attempt to provide guidance. They need to be validated by a wider group of CLA experts.

Technological appropriateness must be evaluated in the context of whether the technology represents the best, worst or average technology for the time and geography in question.

4. Steps for validation

Validation can occur in either an automated or a manual fashion. Automated validation should occur as much as possible during data input, while manual validation can be performed either before or after the data enters the system. Until software applications are developed, all the automated validation will have to be performed manually.

4.1. Automated data validation

Automated data validation includes:

- Verification that all required fields are complete and properly formatted.
- Where chemical equations are used for the calculation, the software should check to verify that all chemicals in the equation are present as inputs or outputs to the system.
- A mass balance on the unit process.

Where these basic tests fail, the system should automatically send a message to the provider that the dataset failed validation and needs to be fixed prior to becoming part of the database.

4.2. Manual data validation

Manual data validation should perform the following tests

1) Checking references (can they be readily found, are they appropriate?)
2) Re-calculating flows (must re-calculate energy, GHGs, a selection of others).
3) Checking other LCI data/studies to see if they are comparable. Comparable means that when the unit processes are roughly equivalent, the flows should agree within two standard deviations, approximately the 95% confidence interval).
4) From specific knowledge of the process, are there missing flows? This analysis is done by inspection, but may also be performed using a direct comparison from other data sets.
5) Is missing or ageing data significant? This should be evaluated using estimates of variability and a sensitivity analysis.
6) Is the characterization of the unit process correct (single/system; maturity)? This evaluation must be based on process knowledge.

If changes are made in the dataset as a result of the validation exercise, a new validation should be performed.

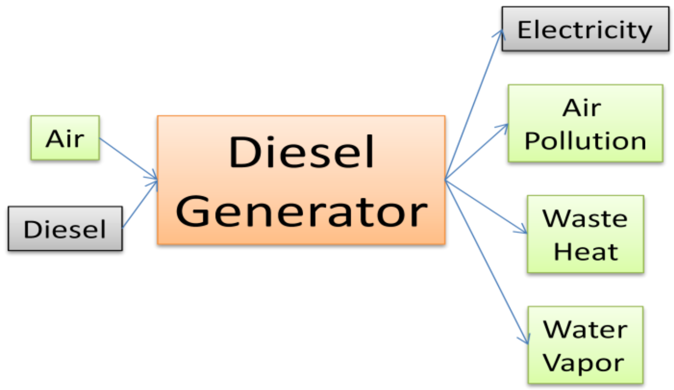
4.3.Qualifications of Validators

Validators should at a minimum be Life cycle Assessment Certified Professionals (LCACPs) in good standing, per the IERE program or its equivalent. Ideally they should also be certified to a program derived from the requirements of this data quality plan (to be created based on approval by the committee and NREL).

4.4. Example 1: calculated flows

In order for manual validation to occur, the documentation of the data calculation must be very clear. Ideally, the calculation should be described for each flow using an equation that links the technosphere flow to all relevant ecosphere flows, and identifies the source of the conversion factors. In a perfect world, one would have all primary data, but this is rarely if ever the case.

A good example is one of combustion. Typically, one knows the type of combustion (e.g., boiler, open air, internal combustion) plus the amount of fuel consumed, and possibly the amount of energy produced. All other flows to and from the system are calculated using mass or energy balances and emission factors. The case of electricity derived from a diesel internal combustion generator is shown below. In this case, the reference flow is the electricity produced (the technosphere output).



The general chemical equations are:

$CH_2 + O_2 \rightarrow CO_2 + H_2O$

$\frac{1}{2} N_2 + O_2 \rightarrow NO_2$

One can obtain air emission factors from EPAs AP-42 documents, as follows.

Substance	Emission Factor Lbs per MMBtu (fuel input)
NOx	3.2
CO	0.85
SOx	1.01
CO2	165
PM	0.1
TOC (as Methane)	0.09
< 1u	.0478
< 3u	.0479

Substance	Emission Factor Lbs per MMBtu (fuel input)
<10u	.0496
Total Filterable	.0620
Condensed	.0077
Total PM-10	.0573
Total Particulate	.0697
Benzene	7.76 E-04
Toluene	2.81 E-04
Xylenes	1.93 E-04
Propylene	2.79 E-03
Formaldehyde	7.89 E-05
Acetaldehyde	2.52 E-05
Acrolein	7.88 E-06
Naphthalene	1.30 E-04
Acenaphthylene	9.23 E-06
Acenaphthene	4.68 E-06
Fluorene	1.28 E-05
Phenanthrene	4.08 E-05
Anthracene	1.23 E-06
Fluoranthene	4.03 E-06
Benz(a)anthracene	6.22 E-07
Pyrene	3.71 E-06
Chrysene	1.53 E-06
Benzo(b)fluoranthene	1.11 E-06

Also from AP-42: “the average heating value of diesel was assumed to be 19,300 Btu/lb with a density of 7.1 lb/gallon.”

From the Energy Information Agency one finds that the conversion efficiency of diesel to electricity is 33%.

All the measurements must be converted to the proper metric units, and then the following calculations can be made:

1) Air pollution for substance i = fuel consumed * emission factor for substance i
2) Mass Oxygen used = mass CO_2 produced * Molecular weight O_2/Molecular weight CO_2 + Mass CO produced * atomic weight O/molecular weight CO + mass NO_2 produced * molecular weight O_2/molecular weight NO_2
3) Mass Nitrogen used = Mass NOx produced * Atomic weight N/Molecular weight NO_2
4) Mass Water Produced = Mass CO_2 produced * Molecular weight H_2O/Molecular weight CO_2
5) Waste Heat = energy of fuel used * 0.67
6) Electricity = energy of fuel used * .33

Molecular weights can be found at http://www.chemie.fu-berlin.de/cgi-bin/molform

Normally unit conversion factors are not supplied, since they are readily available from many sources. A validator will be able to check that the units are properly converted. Decimal point errors are among the most common input errors for LCI data.

4.5. Missing Data

The output inventory for the unit process would still have some things missing. For example, there is no information about the emissions of naturally occurring radiation or of noise. There is no information about land use. The TOC emissions figure indicates that some of the air emissions are not fully characterized. There is likely to be some solid waste produced, but we do not have these data either. It is the job of the validator to decide whether the absence of these is significant, or merely represents normal and acceptable LCI practice.

Note that for this example, we do not have any estimates of the variability of the inventory data. That is because we only know the fuel consumed and emission factors for each ecosphere flow, and neither of these is presented with any statistical information.

UNSPSC ID	**Technosphere Output**	
83101803	Supply of three phase electricity	
	Reference Flow = 1 MJ Electricity	
UNSPSC ID	Technosphere input	Grams
15101505	Diesel Fuel	67.6

CAS RN	Ecosphere Input	
7782-44-7	Oxygen (air)	160
7727-37-9	Nitrogen (air)	1.27
CAS RN	Ecosphere Output to Air	Grams
10102-44-0	NOx as NO2	4.17
630-08-0	CO	1.11
05-09-46	SOx as SO2	1.32
22541-53-3	CO2	215
	PM	0.130
74-82-8	TOC (as Methane)	0.117
	< 1u	0.0623
	< 3u	0.0625
	<10u	0.0647
	Total Filterable	0.0809
	Condensed PM	0.0100
	Total PM-10	0.0747
	Total Particulate	0.0909
71-43-2	Benzene	0.00101
108-88-3	Toluene	0.000366
1330-20-7	Xylenes	0.000252
115-07-1	Propylene	0.00364
50-00-0	Formaldehyde	0.000103
75-07-0	Acetaldehyde	3.29E-05
107-02-8	Acrolein	1.03E-05
91-20-3	Naphthalene	0.000170
208-96-8	Acenaphthylene	1.20E-05
83-32-9	Acenaphthene	6.10E-06
86-73-7	Fluorene	1.67E-05
85-01-8	Phenanthrene	5.32E-05
120-12-7	Anthracene	1.60E-06
206-44-0	Fluoranthene	5.26E-06
56-55-3	Benz(a)anthracene	8.11E-07
129-00-0	Pyrene	4.84E-06
218-01-9	Chrysene	2.00E-06
205-99-2	Benzo(b)fluoranthene	1.45E-06
7732-18-5	Water (vapor)	88.1
	Energy outputs to ecosphere	MJ
	Waste heat	2.00

4.6. Example 2: Measured flows (Primary Data)

Suppose that you are lucky enough to have measured flows from continuous emission monitoring or some other method. You are in the excellent position of having data that can be fully statistically characterized. Let's imagine that you are producing glass, removing and returning water from a river, and here is your data.

Technosphere Inputs, 2007	Mean metric tons/ton glass	Standard Deviation	Number of samples	Kurtosis
Sand	.72	.0055	12	.98
Crushed l.imestone	.18	.00015	12	1.02
Soda Ash	.15	.0025	12	1.05
Electricity (MWH)	1.47	.102	12	1.4
Powdered graphite	0.000001	N/A	1	N/A
Technosphere outputs	Mean ton/ton			
Mixed solid waste	0.01	.00032	12	.91
Contaminated cullet	.0015	.0005	50	1.6
Ecosphere Inputs	Mean kg/ton glass			
Water	1300	3500	12	1.1
Ecosphere Output	Mean Kg/ton			
PM-2.5	.05	0.025	12	.98
Residual chlorine	0.0026	0.0031	12	0.5
Suspended solids	0.051	0.030	12	0.53
Water	1300	3500	12	1.1
	Mean MJ/ton			
Waste heat	1.27	0.23	50	1.02

The measurements here show that the distribution of most but not all quantities is normal. Chlorine and suspended solids have a uniform distribution (implying that their emission are not closely related to the production), while electricity consumption and production of contaminated cullet is log-normally distributed. Having this information allows one to make estimates of total errors based on Monte Carlo simulations and other statistical methods.

4.7.Default Statistical Assumptions

Normal distributions are less common in nature than long normal distributions. In the absence of statistical data, one can assume that the distribution of both inputs and outputs are log-normally distributed, with relative standard deviations of 1.0. This permits validators to estimate whether the missing data or the data for which no statistical data is available is important.

Applications needed

Input applications

An input a verification application and an output module are needed. Several of the input application features are mentioned above. The validation application should include information on the validator (contact and credential information), date the dataset was validated, tests performed, and the outcome of those tests.

	Application	Description
1	Primary data input tool	-Allow data providers to input data into the system providing step by step guidance on required fields and formatting -Streamlined auto formatting data collection
2	Nomenclature tool	When entering the type of unit process being entered, the data provider can search for the nomenclature for matching unit process in database
3	Unit converter	Converts all data inputs into SI units
4	Graphical interface input tool	input application that allows people to use a graphical interface to: -draw the unit process, showing ecosphere and technosphere inputs and outputs, -describe the nature of the unit process (e.g. chemical reaction, forming, mining) -input the formulas and tables linking the flows, -calculate all the flows, -making the unit conversions as necessary
5	Validation tool	-verifies all required fields are complete -verifies all required fields are properly formatted -verifies proper nomenclature is utilized -verifies proper units are utilized -performs mass balance of unit processes

	Application	Description
		-performs energy balance of unit processes -Notifies data provider of any validation test failures and corrections are required prior to inclusion in the database. -Compare inputs & outputs to similar processes in database -Where chemical equations are used for the calculation, the software should check to verify that all chemicals in the equation are present as inputs or outputs to the system.
6	QA tracking tool	Track data QA process, assign reviewers, track comments responses, action items and responsibility
7	Data aggregating tool	Collect data from different providers for like processes providing averages, ranges and standard deviations.
8	Upload tool	When unit process data has been validated (automatic and manual), the tool will upload files into database for public access

Output Applications

An output application that allows the user to sort the data according to data requirements would be useful, but not essential. It could ask the user what level of data quality it needed in terms of proposed use (level 1,2,3) and geography, then provide only those data that fit the parameters. The question of technological appropriateness can only be answered in terms of the study to be undertaken, and cannot be evaluated at the level of the unit process

	Application	Description
9	Historical data averaging tool	Tool to provide historical averages for a data set. (e.g. electricity data)
10	Format conversion	Convert to EcoSpold, ILCD formats (from GreenDelta?)
11	Website data viewing tool	Allows user to view data set online
12	Data roll-up tool	Tool to combine unit processes that are connected upstream and downstream to provide the user with a cradle to gate perspective
13	Graphical roll up tool	Tool to create a process tree / network diagram for upstream processes
14	Auto update tool	Tool to automatically update data set with frequently recurring update (i.e. public energy data like electricity)
15	Data set update tracker	Tool to track when data sets are due to be reviewed
16	Data search tool	Tool to search for data based on industry type, process, material, region, other meta data categories
17	Help tool	Providing guidance as needed to users, data providers, etc as they are utilizing the database, entering data, etc.