



Netzwerk Lebenszyklusdaten

Arbeitskreis METHODIK

AP 9 Konzept zur Unterstützung der Konvertierung von Datensätzen für das Netzwerk Lebenszyklusdaten

Projektbericht

im Rahmen des Forschungsvorhabens FKZ 01 RN 0401 im Auftrag
des Bundesministeriums für Bildung und Forschung

GreenDeltaTC GmbH

Forschungszentrum Karlsruhe GmbH

Berlin Karlsruhe – September 2007

greendeltaTC



Forschungszentrum Karlsruhe
in der Helmholtz-Gemeinschaft



Vorwort

Der vorliegende Projektbericht wird herausgegeben vom Netzwerk Lebenszyklusdaten (www.netzwerk-lebenszyklusdaten.de).

Das Netzwerk Lebenszyklusdaten ist die gemeinsame Informations- und Koordinationsplattform aller in die Bereitstellung und Nutzung von Lebenszyklusdaten in Deutschland involvierten Gruppen – von Wissenschaft und Wirtschaft über Politik und Behörden hin zu Verbraucherberatung und allgemeiner interessierter Öffentlichkeit. Ziel des Netzwerks Lebenszyklusdaten ist es, das umfangreiche Knowhow auf dem Gebiet der Lebenszyklusdaten innerhalb Deutschlands zusammenzuführen und als Basis zukünftiger wissenschaftlicher Weiterentwicklung und praktischer Arbeiten für Nutzer in allen Anwendungsgebieten von Lebenszyklusanalysen bereitzustellen.

Das Netzwerk Lebenszyklusdaten wird getragen vom Forschungszentrum Karlsruhe. Die vorliegende Studie wurde im Rahmen der Projektförderung (2004 – 2008) des Bundesministeriums für Bildung und Forschung (BMBF) „Förderung der Wissenskooperation zum Aufbau und Umsetzung des deutschen Netzwerks Lebenszyklusdaten“ erstellt. Weitere im Rahmen dieser Projektförderung erstellte Studien sind erhältlich unter <http://www.netzwerk-lebenszyklusdaten.de/cms/content/Projektberichte>.

Kontakt Netzwerk Lebenszyklusdaten:

E-Mail: info@netzwerk-lebenszyklusdaten.de

Anschrift: Forschungszentrum Karlsruhe GmbH
Institut für Technikfolgenabschätzung und Systemanalyse,
Zentralabteilung Technikbedingte Stoffströme (ITAS-ZTS)
Postfach 3640
76021 Karlsruhe
www.netzwerk-lebenszyklusdaten.de



Das Netzwerk Lebenszyklusdaten wird gefördert durch das
Bundesministerium für Bildung und Forschung



AP 9

Konzept zur Unterstützung der Konvertierung von Datensätzen für das Netzwerk Lebenszyklusdaten

Autoren:

Dr. Andreas Ciroth
GreenDeltaTC

Michael Srocka
GreenDeltaTC

Dr. Clemens Döpmeier
Forschungszentrum Karlsruhe

Oliver Kusche
Forschungszentrum Karlsruhe

Achim Stadtherr
Forschungszentrum Karlsruhe

Kontakt:

Dr. Andreas Ciroth
GreendeltaTC GmbH
Raumerstraße 7
D-10437 Berlin
T: +49 30 4849 6030
F: +49 30 4849 6991
E-mail: gdtc@greendeltatc.com

Inhalt

1	Einführung.....	5
1.1	Probleme bei der syntaktischen Konvertierung und Lösungsmöglichkeiten	5
1.2	Anforderungen an einen Konverter für das Netzwerk Lebenszyklusdaten	7
2	Gegenwärtige Umsetzung im Netzwerk Lebenszyklusdaten	8
3	Konzept zur dauerhaften Verbesserung der Konvertierung von LCI Datensätzen durch semantische Konvertierung	10
3.1	Grundlagen der semantischen Konvertierung	11
3.2	Aufbau von Referenz- und Mapping-Listen	12
3.3	Ausblick: Nutzung von Ontologien für die semantische Konvertierung	14
4	Zusammenfassung und Empfehlungen	15
5	Anhang.....	17

1 Einführung

Datensätze für Sachbilanzen oder Life Cycle Inventories (im Folgenden kurz LCI Datensätze) sind zurzeit in unterschiedlichen Formaten verfügbar und werden von Anwendern und LCA Software in unterschiedlichen Datenformaten benötigt. Derzeit relevant sind das EcoSpold Format, entwickelt vom Schweizer eocinvent centre, das ELCD Format des Joint Research Centre der EU Kommission, das ISO14048/IMI Format der IMI Gruppe der Universität Chalmers, sowie einige Datenformate von LCA Softwareanbietern.

Die unterschiedliche Semantik der Datenformate verhindert zur Zeit einen vollständig automatisierten und verlustfreien Austausch von Datensätzen. Das Netzwerk Lebenszyklusdaten möchte aber einen weitgehend verlustfreien und format-unabhängigen Austausch von LCI Daten fördern.

Für das Netzwerk ist es daher wichtig, mit unterschiedlichen Datenformaten umgehen und diese verarbeiten zu können, sowie Schnittstellen zu allen in der Datenbereitstellung wichtigen Beteiligten zu schaffen. Eine Webanwendung zum zentralen Datenaustausch, die mit verschiedenen Formaten umgehen kann, ist für das Netzwerk besonders attraktiv, da sie einerseits zentral gepflegt werden kann, und zum anderen von unterschiedlichen Institutionen (eben als Netzwerkanwendung im Wortsinn) angesteuert und genutzt werden kann.

Die Erarbeitung von Schnittstellen zu allen wichtigen Beteiligten des Netzwerks wird im Arbeitsprogramm für die IT-Umsetzung des vom BMBF geförderten Forschungsvorhabens ‚Förderung der Wissenschaftskooperation zum Aufbau und Umsetzung des deutschen “Netzwerk Lebenszyklusdaten“ (FKZ 01 RN 0401)‘ angegangen (im weiteren kurz ‚BMBF-Vorhaben‘).

Ausgangspunkt des Projektes zur Implementierung einer zentralen Formatkonvertierungsschnittstelle für das Netzwerk war der im Rahmen des im openLCA-Projektes der GreendeltaTC GmbH entwickelte Formatkonverter, der im Rahmen des hier beschriebenen Projektes weiterentwickelt wurde. Gleichzeitig wurde daraus eine in eine Webumgebung integrierbare Softwarebibliothek entwickelt. Die folgenden Kapitel beschreiben die aktuelle Situation und Probleme bei der Konvertierung; sie enthalten außerdem ein Konzept für eine Verbesserung der Konvertierung in der Zukunft und Empfehlungen, welche Beiträge dazu das Netzwerk Lebenszyklusdaten leisten kann.

1.1 Probleme bei der syntaktischen Konvertierung und Lösungsmöglichkeiten

Die Konvertierung nimmt letztlich eine Zuordnung von Elementen eines Formats, des Ausgangsformats, zu den Elementen des anderen Formats, des Zielformats, vor.

Bei der Konvertierung treten gegenwärtig verschiedene Probleme auf bzw. sind verschiedene Aspekte zu berücksichtigen. Dabei sind verschiedene Abbildungsformen zu unterscheiden: Die reine Abbildung der syntaktischen Konstrukte bis hin zu dem so genannten semantischen Mapping (d.h. die Transformation der Semantik des einen Formates in die korrespondierende des Zielformates). Das semantische Mapping

berücksichtigt die Bedeutung der Inhalte der einzelnen Elemente; es baut auf dem normalen Mapping auf und erweitert es. Da die inhaltliche Bedeutung erfasst und an das Zielformat angepasst wird, verbessert sich in der Regel der Informationsgehalt der erzeugten Datensätze deutlich (siehe Kapitel 3).

Für das syntaktische Mapping stellen sich folgende Herausforderungen¹:

- **Fehlendes Zielfeld:** Für eine Information des Ausgangsformates existiert im Zielformat kein entsprechendes Feld. Evtl. können diese Informationen in allgemeinen Feldern des Zielformates (z.B. Kommentarfelder) untergebracht werden. Weiterhin bieten die Datenformate teilweise sogenannte ‚Extension Points‘ an, die eine Erweiterung des Zieldatensatzes um die Informationen des Ausgangsdatsatzes ermöglichen. Mit Informationen in den ‚Extension Points‘ können jedoch im Normalfall die auf dem Zielformat basierenden Anwendungen nicht umgehen, so dass die Informationen nicht angezeigt oder weiterverarbeitet werden. Können weder allgemeine Felder noch Extensions genutzt werden, so gehen die Information des Ausgangsformates verloren.

Beispiel: Für ‚ELCD-Process-KeyDataSetInformation-TechnicalPurposeOfProductOrProcess‘ gibt es in Ecospold kein geeignetes Zielfeld. Der Formatkonverter schreibt die Information zusammen mit anderen Informationen daher in das Ecospold Feld `MetaInformation-ProcessInformation-Technology-text`

- **Pflichtangaben:** Wenn ein Element des Zielformats eine Pflichtangabe ist, das entsprechende Element des Ausgangsformates aber nicht, muss ein Defaultwert gesetzt werden; falls für das Zielformats-Element kein Defaultwert existiert, muss ein sonstig als nicht sinnvoller Eintrag erkennbarer Wert eingetragen werden.

Beispiel: `MetaInformation-TSource-year` in EcoSpold ist Pflichtangabe, kann jedoch aus einem ELCD Datensatz nicht gefüllt werden, stattdessen wird ein ‚Nonsens‘-Wert von ‚9999‘ eingefüllt.

- **Häufigkeit:** Wenn ein Element des Ausgangsformats häufiger vorkommt als ein Element des Zielformats vorkommen darf, ist eine direkte Zuordnung zum passenden Element des Zielformats nicht möglich. Die „überschüssigen“ Elemente des Ausgangsformats können möglicherweise in anderen Feldern (Kommentarfeldern o.ä.) aufbewahrt werden oder in einem Feld zusammengefasst werden.

Beispiel: Im EcoSpold Format kann für jeden einzelnen Fluss eines Prozesses eine eigene Allokationsmethode eingetragen werden, im ELCD Format dagegen nur eine Allokationsmethode in den Metainformationen; im ELCD Format können für jedes Flowdataset bis zu 100 allgemeine Kommentare in jeweils einer Sprache (`Flow-DataSetInformationType-generalComment`) abgegeben werden, in EcoSpold dagegen nur höchstens ein Kommentar (Feld `EcoSpold-FlowData-TExchange-generalComment`).

¹ Die Bezeichnungen der Elemente der Datenformate sind der Mapping Dokumentation des Format Konverters entnommen, siehe http://openlca.org/kb/mappingdoc/EcoSpold_to_ELCD.xml und http://openlca.org/kb/mappingdoc/ELCD_to_EcoSpold.xml.

- **Datentyp:** Wenn sich der Datentyp von Ziel- und Ausgangsformat unterscheidet, muss das Ausgangsformat in das Zielformat geändert (gecastet) werden. Das ist oft mit Informationsverlust verbunden.

Beispiel: (1) Für den Namen einer Kontaktperson sieht EcoSpold ein Textfeld mit 40 Zeichen vor (EcoSpold-DataType-TPerson-name), ELCD dagegen ein Textfeld mit 500 Zeichen (ELCD-Contact-DataSetInformationType-name) (2) Die Standardabweichung ist in ELCD ein „decimal“ Format (= ELCD-eigenes Format, Dezimalzahl), in EcoSpold dagegen ein double-Wert (Felder ELCD-Process-ExchangeType-relativeStandardDeviation95In, EcoSpold-FlowData-TExchange-standardDeviation95).

- **Nomenklatur** oder **Muster:** Wenn für das Element des Zielformats eine bestimmte Nomenklatur vorgeschrieben ist oder ein bestimmtes Muster als Wert vorzugeben ist, diese Nomenklatur / dieses Muster aber nicht vom Ausgangsformat geliefert wird, kann das Element im Zielformat nicht belegt werden. Hier sind Zuordnungs- oder Mappinglisten hilfreich, die mögliche Einträge von Ausgangs- und Zielformat einander zuordnen. Diese Mappinglisten ordnen Bedeutungen zu und sind daher streng genommen bereits semantisches Mapping.

Beispiel 1: Die CAS Nummer zur Kennzeichnung von Stoffen ist in EcoSpold (EcoSpold-FlowData-TExchange-CASNumber) ein Muster (Pattern), mit einer vorgegebenen Struktur, in ELCD dagegen ein nicht weiter spezifizierter String mit Feldlänge 500 (Feld ELCD-Flow-DataSetInformationType-CASNumber).

Beispiel 2: Bezeichnungen für Wahrscheinlichkeitsverteilungen unterscheiden sich im Detail zwischen ELCD und EcoSpold Format. Die Konvertierung greift daher auf folgende Mappingliste zurück:

ELCD	EcoSpold
undefined	0=undefined
log-normal	1=lognormal (default)
normal	2=normal
triangular	3=triang
uniform	4=uniform

Abbildung 1: Im Format Konverter verwendete Mapping Liste für Wahrscheinlichkeitsverteilungen von ELCD und EcoSpold

1.2 Anforderungen an einen Konverter für das Netzwerk Lebenszyklusdaten

Eine Webanwendung zur Konvertierung von LCI Datensätzen in unterschiedliche, derzeit relevante Datenformate hat aus Sicht des Netzwerks Lebenszyklusdaten das allgemeine Ziel, zu zeigen, dass eine praktikable Lösung für die Konvertierung von Datensätzen unterschiedlicher Formate „mit Hand und Fuß“ existiert und über das Internet angeboten werden kann.

Anforderungen an eine Webanwendung zur Unterstützung der Formatkonvertierung für das Netzwerk Lebenszyklusdaten sind im Einzelnen:

- Demonstrieren dass eine Konversion von Daten unterschiedlicher Formate prinzipiell über das Netz möglich ist, und vom Netzwerk als Dienstleistung angeboten werden kann;
- Die Anwendung soll praktikabel sein für die Zielgruppe;
- Sie soll daher in einen allgemeinen Workflow der Datensatzannahme, -prüfung, Datensatzaufnahme in die Netzwerk-/ FZK-Datenbank, und Weitergabe von Datensätzen, integriert werden oder sich integrieren lassen;
- Sie soll zeigen dass das Netzwerk (und fortfolgende, darauf aufbauende Organisationsformen) dazu in der Lage ist/sind, mit unterschiedlichen Datenformaten umzugehen und sie zu verarbeiten und somit über die Fähigkeit verfügt Schnittstellen zu allen in der Datenbereitstellung wichtigen Beteiligten zu schaffen. Mit den von der Webanwendung in der ersten Version zu unterstützenden Formaten Ecospold, ELCD und ISOTS14048_IMI verfügt das Netzwerk über Schnittstellen zu den derzeit wichtigsten Datenformaten;
- Die derzeit und zukünftig im Netzwerk entstehenden Datensätze im ELCD Format sollen mit der Anwendung in das EcoSpold Format der Versuchsdatenbank des Netzwerks übertragen werden;
- Anvisierte Nutzergruppen für die Webanwendung sind:
 - Wissenschaft und Forschung (BMBF Projekte, EU Projekte, Hochschulen);
 - LCA practioner.

Ein Grundverständnis der LCA Methodik kann daher vom Nutzer vorausgesetzt werden.

2 Gegenwärtige Umsetzung im Netzwerk Lebenszyklusdaten

Ausgehend von dem Formatkonverter des OpenLCA-Projektes wurden im Rahmen dieses Projektes zwei Möglichkeiten zur Konvertierung von und in die Datenformate Ecospold, ELCD und ISOTS14048_IMI für das Netzwerk Lebenszyklusdaten erarbeitet:

- Der Formatkonverter als stand-alone Anwendung aus dem openLCA Projekt (www.openlca.org):
Dieser ermöglicht für die Geschäftsstelle des Netzwerks Lebenszyklusdaten die Annahme und Bereitstellung der im Zuge des BMBF-Vorhabens erarbeiteten Datensätze in den unterstützten Datenformaten sowie die Speicherung in der auf dem Ecospold-Format basierenden Versuchsdatenbank zur sofortigen strukturierten Bereitstellung und vollständigen Integration in der zukünftigen Datenbereitstellung des Netzwerks.
- Die mit Unterstützung durch dieses Vorhaben entwickelte Bibliothek (API) des Konverters:
Diese ist integrierbar in eigene Anwendungen sowie in das Webportal des Netzwerks, welches die Konversion von Datensätzen über das Web-Portal ermöglicht.
Nach der Integration der API in das Webportal können beliebige Mitglieder des Netzwerks oder nach Freischaltung auch Dritte das Web-Portal dazu nutzen, um

Datensätze online zu konvertieren oder um konvertierte Datensätze des Netzwerks vom Server herunter zu laden. In einem nächsten Schritt wird die API in die Upload und Download-Funktion der Versuchsdatenbank integriert, so dass in Echtzeit jeweils auch Datensätze in anderen Formaten in die Versuchsdatenbank hochgeladen und von dieser wieder heruntergeladen können.

Beide Anwendungen sind in der Konversion auf gleichem Stand. Sie liegen als Open Source Anwendungen vor. Derzeit konvertieren sie die Formate ELCD 1.0.1, EcoSpold (aktuelle Version) und ISO14048/IMI der Chalmers Universität und erzeugen jeweils schema-valide Datensätze für das Zielformat, im batch-Modus (zahlreiche Datensätze) und für einzelne Datensätze. Alle Informationen des Ausgangsformat werden in einem Metadokument gespeichert, das bei einer „roundtrip-conversion“ (von einem Ausgangsformat für einen Datensatz in ein Zielformat und zurück) optional berücksichtigt werden kann und so dafür sorgt, dass der Ausgangsdatsatz vollständig wiederhergestellt werden kann.

Da der Quellcode offen liegt, können die Anwendungen natürlich um weitere Formate ergänzt werden, entweder durch das FZK / IAI, durch das openLCA Projekt, oder durch Dritte.

Abbildung 2 zeigt die Konvertierung des ELCD-Datensatzes „Heavy fuel oil, consumption mix; at refinery“ mit dem Formatkonverter für einen Teil der Felder. Auf der linken Seite ist der Original-Datensatz im ELCD Format dargestellt und auf der rechten Seite der resultierende Datensatz im Ecospold-Format. Die mit den Pfeilen verbundenen roten Felder zeigen an, von welchem Ursprungsfeld aus Inhalte aus dem ELCD-Format in das Ecospold-Format übertragen wurden. Dabei werden die bereits in Kapitel 3 erwähnten Probleme deutlich. So kann aufgrund eines unterschiedlichen **Datentyps** (unterschiedliche Feldlänge) nur der ‚Base name‘ vom ELCD-Format nach Ecospold übernommen werden. Die zusätzlichen Angaben aus ‚Mix type and Location‘ werden in das allgemeine Feld „Technology“ geschrieben (in der Abbildung nicht zu sehen). Im Ecospold-Format ist die Angabe ‚Local name“ eine **Pflichtangabe**. Dafür existiert im ELCD Format keine Angabe. Es wird daher der Default-Wert ‚unspecified‘ gesetzt (rechts oben in Abbildung 2). Für die Information ‚Use advice for data set‘ in ELCD (Fragezeichen in der Abbildung) **fehlt das Zielfeld**. Die Information geht in der Ecospold-Ansicht verloren².

² Der Formatkonverter schreibt in das Feld ‚general Comment‘, das im Prinzip für die Aufnahme des Felds ‚Use Advice‘ geeignet wäre, eine Information zur Konvertierung und belegt damit dieses Feld. Andere Felder werden nicht auf diese Weise belegt. Im während der Konvertierung optional erzeugten Metadokument wird der ‚use advice‘ übertragen und geht daher *nicht* verloren (kann aber im EcoSpold Stylesheet nicht dargestellt werden).

The screenshot displays the conversion of an ELCD dataset into Ecospol format. On the left, the ELCD interface shows the dataset 'Heavy fuel oil, Consumption mix, at refinery' with various metadata fields such as 'Location', 'Reference year', and 'Technical purpose'. A red box highlights the 'Average data, no individual sub-locations explicitly considered' note. On the right, the Ecospol interface shows the converted dataset with fields like 'Name', 'Reference function', 'Category', and 'Technology description'. Red boxes in the Ecospol window highlight the 'Energy carriers' and 'Crude oil based energy carriers' categories. A large text block in the Ecospol window contains the full technology description from the ELCD window, which details the petroleum refinery processes from distillation to desulfurization.

Abbildung 2: Ausschnitt der Konvertierung des ELCD-Datensatzes ‚Heavy fuel oil‘ nach Ecospol

Die Umsetzung der Konvertierung im Formatkonverter bzw. der Konverterbibliothek verbleibt zunächst weitestgehend auf der Ebene der reinen Abbildung der syntaktischen Konstrukte. Die dabei auftretenden Probleme werden so weit wie möglich aufgelöst oder vorerst durch Behelfslösungen angegangen. Die genaue Vorgehensweise bei der syntaktischen Abbildung der beiden Formate Ecospol und Ecoinvent wird in der Mapping Dokumentation des Format Konverters (http://openlca.org/kb/mappingdoc/EcoSpold_to_ELCD.xml und http://openlca.org/kb/mappingdoc/ELCD_to_EcoSpold.xml) sowie dem Dokument ‚Externe Listen_final.doc‘ beschrieben, welches beim Netzwerk Lebenszyklusdaten vorliegt.

3 Konzept zur dauerhaften Verbesserung der Konvertierung von LCI Datensätzen durch semantische Konvertierung

Wie schon angedeutet, lässt sich die Konvertierung von Datenformaten in zwei Abbildungsformen unterscheiden. Bei der ersten Form erfolgt eine reine Abbildung der syntaktischen Konstrukte vom einen Datenformat auf das andere. Bei der zweiten Form soll eine komplette Übertragung der inhaltlichen Bedeutung der Einträge im Ausgangsformat, der „Semantik“ des Ausgangsformates, auf die Semantik des Zielformates erfolgen. Erst damit wird eine vollständige Integration eines Datensatzes in einen anderen Datenbestand möglich, was die Verfügbarkeit von Daten enorm erhöht. In den folgenden Ausführungen

werden grundlegende Gedanken zum semantischen Mapping aufgegriffen sowie ein Konzept für die nächsten Schritte hin zu einem semantischen Mapping vorgestellt. Im Ausblick wird dargestellt, in welche Richtung die weitere Forschung der an diesen Aktivitäten beteiligten Partner im Netzwerk Lebenszyklusdaten geht, um die semantische Konvertierung weiter voranzutreiben.

3.1 Grundlagen der semantischen Konvertierung

Semantisches Mapping wenden der Formatkonverter bzw. die Konverter-API bisher nur eingeschränkt an. Die derzeitigen Grenzen bezüglich des semantischen Mappings zeigt das Beispiel der Zuordnung der Datensätze zu Kategorien (siehe Abbildung 2) auf. Hier werden bei der derzeitigen Konvertierung die Informationen aus den Feldern ‚Category Information‘ im Ecospold-Format einfach in die Felder Category und Subcategory des ELCD Formates übertragen (sowie mangels Internationalisierung des ELCD-Datensatzes in localCategory und localSubCategory – siehe Abbildung). Die Kategorie ‚Energy carriers‘ mit der Unterkategorie ‚Crude oil based energy carriers‘ stammt aus dem ELCD-Datenbestand bzw. Namensraum. Durch Einsetzung einer entsprechenden Kategorie aus dem Ecoinvent Namensraum bzw. Datenbestand (also Mapping der ELCD Kategorien auf Ecoinvent-Kategorien, hier z.B. auf die Ecoinvent-Kategorien oil/fuels), würde die Einsortierung des Datensatzes in den Ecoinvent-Datenbestand direkt neben bereits vorhandenen - inhaltlich ähnlichen - Ecoinvent-Datensätzen ermöglicht. Elemente, die evtl. Bausteine für ein semantisches Mapping sind, aber derzeit noch nicht in einen eventuellen Zielnamensraum übertragen werden, sind z.B. Kategorieinformationen, Elementarflussbezeichnungen, Bezeichnungen für ökonomischen Flüsse bzw. Zwischenproduktflüsse sowie Prozessnamen. Es besteht derzeit noch Forschungsbedarf, inwieweit hier bei einzelnen Elementen zukünftig ein Mapping möglich und sinnvoll ist.

Das semantische Mapping ordnet Objekte gleicher Bedeutung, semantisch gleiche Objekte, einander zu. Es ist zunächst zu klären, wann Objekte als semantisch gleich anzusehen sind. Dabei muss man auch zwischen den verschiedenen Arten von Gleichheit (z.B. zwischen den Begrifflichkeiten gleich und dasselbe [„Ding“]) unterscheiden. Z. B. gibt es sowohl in Ecospold als auch in ELCD mehrere gleiche Felder ‚name‘, die aber nicht immer dasselbe bedeuten. Daher können Namen über die Formate nur bedingt als Indikator für Gleichheit verwendet werden. Eine recht einfache Lösung zum Ausdrücken von Gleichheit besteht darin, eine URI (= unified resource identifier) bei einem Element zu setzen, die als Webadresse sozusagen auf die Beschreibung eines solchen Elementes im Internet weist (z.B. als Bestandteil einer Ontologie, eines semantischen Modells) und Gleichheit von zwei Instanzen genau dann anzunehmen, wenn die URI gleich ist. Eine URI hat typischerweise die Form einer Webadresse wie z.B. http://lca.jrc.ec.europa.eu/lcainfohub/datasets/elcd/processes/Steel_rebar_Eurofer_268a11fb-baf2-4b9e-8867-38bea0e76ef6_01.00.001.xml aus dem ELCD Format. Diese Lösung würde eine Standardisierung verschiedener Konstrukte, z.B. für Elementarflüsse, über ihre Beschreibung im Web voraussetzen, und dann die Verpflichtung an die Tools bzw. Anwender bedingen, die eindeutige URL des Konstruktes als eindeutige ID für die Gleichheit mit in ihre Werkzeuge bzw. in die Austauschformate zu übernehmen, wie dies bei ELCD prinzipiell bereits für einige Konstrukte, wie Flows, möglich ist. Eine URL ist dabei einer GUID vorzuziehen, da GUIDs, obwohl ebenfalls eindeutig, schwerer lesbar sind und auch nicht zumindest potentiell über das Internet direkt aufgerufen werden können.

Ein semantisches Mapping wird oft durch Referenzlisten ermöglicht, die als allgemeingültig akzeptiert werden. Sie ermöglichen eine Zuordnung der unterschiedlichen ‚Dialekte‘ in den Formaten. Ein Beispiel für eine Referenzliste sind die physikalischen Standardeinheiten, die sogenannten SI Einheiten („Système International d'unités“).

In der Praxis hat semantisches Mapping, wie bereits angedeutet, die Auswirkung, dass ein konvertierter Datensatz unterschiedlich weit in einen Datenbestand des Zielformates integriert werden kann. Z.B. bewirkt ein Mapping von Elementarflüssen, dass der konvertierte Datensatz bei der Integration in eine Prozesskette eine konsistente Sachbilanz über die Prozesskette ermöglicht. Ein semantisches Mapping der Produktflüsse ist dagegen Voraussetzung für die Verknüpfbarkeit von Prozessdatensätzen. Ein weiterer Bereich, in dem semantisches Mapping zur Anwendung kommen kann bzw. muss, ist die Abbildung der Quell-Kategorisierung eines konvertierten Datensatzes in das Kategorisierungssystem des Zielformates zur besseren Auffindbarkeit der Datensätze.

3.2 Aufbau von Referenz- und Mapping-Listen

Referenzlisten werden ein wichtiger Baustein für den Aufbau eines semantischen Mappings sein. Sie ordnen Elemente gleicher Bedeutung in den verschiedenen Formaten einander zu und können unabhängig von der eigentlichen Zuordnungslogik aufgebaut und gewartet werden.

Im Idealfall leiten sich die erforderlichen Mappinglisten aus einer semantischen Beschreibung des von den LCA Datensätzen abgebildeten „Informationsraumes“ (Wissensdomain) und einer Abbildung der Konstrukte der Datenformate auf äquivalente Konstrukte in dieser semantischen Beschreibung ab. Eine derartige Beschreibung eines Wissensdomains kann in Form einer Ontologie erfolgen. Sie stellt die im Informationsraum vorhandenen fachlichen Konzepte und ihre Eigenschaften in Form von Ontologieklassen (auch Konzepte genannt) dar, einschließlich ihrer Beziehungen untereinander. Konkrete Ausprägungen von Konzepten (z.B. „Einheit von Masse“) können dann als Individuen (z.B. Kilogramm oder Tonne) in einer Ontologie gespeichert werden. Von diesen derartig erfassten Objekten oder Klassen lassen sich dann Referenzlisten ableiten, die die in den verschiedenen Formaten gleichen Objekte mit ihren Ausprägungen einander zuordnen (z.B. verschiedene Darstellungsformen ein und derselben Einheit in den verschiedenen Formaten).

Mapping Listen sind allerdings auch zum jetzigen Zeitpunkt (ohne Ontologie) denkbar, z.B. für Flussdatensätze (Flow in ELCD, exchanges in EcoSpold / Ecoinvent). Es ist wichtig, für den Aufbau der Listen die allgemeinen Beschreibungen der Formate zu verlassen und sich den konkreten Datensätzen in den Formaten zuzuwenden, also den verschiedenen existenten Bezeichnungen für Flussdatensätze in verschiedenen Formaten.

Im Vorhaben wurde ein erster Vorschlag für die Struktur einer Mapping Liste von Flussdatensätzen aufgestellt (Tabelle 1). Zusätzlich sind ID-Felder für die eindeutige Kennzeichnung der Zuordnung erforderlich, die in der Tabelle nicht dargestellt sind. Ebenfalls nicht dargestellt sind Kategorien und sonstige Klassifikationssysteme in den Formaten, die die Datensätze in das jeweilige Format einbetten und bei der Zuordnung zu berücksichtigen sind.

Tabelle 1: Struktur einer Mapping-Liste für Elementarflüsse aus ELCD und Ecoinvent

Ecoinvent Namen	ELCD Namen	ELCD Sprache	Datum der Zuordnung	Verantwortlichkeit für die Zuordnung	Kommentar
--------------------	---------------	-----------------	------------------------	-----------------------------------------	-----------

Eine solche Mapping Liste für Flüsse bliebe – bei korrekten Zuordnungen – relativ konstant; neu angelegte Flüsse, auf ELCD- oder Ecoinvent-Seite, würden allerdings eine Erweiterung notwendig machen

Der Aufbau und Unterhalt der Listen ist jedoch keineswegs trivial. In der Regel ist hohes Fachwissen erforderlich, um eine korrekte Zuordnung vornehmen zu können. Da sich die Listen ändern können, und Nutzer eine aktuelle Liste erwarten werden, ist ein Mechanismus zum Aufbau und zur Wartung der Listen erforderlich, der auf das oft sehr spezielle Fachwissen von Experten optimal zurückgreift und die Listen für Nutzer optimal bereitstellt.

Aus einer intensiven Diskussion dieser Punkte im Vorhaben ergaben sich folgende Empfehlungen:

- Soweit möglich sollten die Zuordnung und der Abruf der Listen für die Konvertierung softwaregestützt erfolgen. Hierzu ist ein entsprechendes Tool einschließlich eines Nutzungskonzepts zu entwickeln;
- Eine Administration / Moderation der Zuordnungen erscheint in jedem Fall zusätzlich notwendig.
- Die Listen sollten zentral gehalten werden; dezentrale Varianten, mit einem zentralisierten Updateverfahren ähnlich wie bei Open Source Projekten, erscheinen außerdem als interessante, später zu prüfende Variante. Die Entscheidung wird von Zahl und Umfang der Listen und dem dadurch verursachten Datenverkehr abhängen.
- Erstellung und Pflege der Listen sollte „community driven“ erfolgen, im besten Fall auf Basis einer als Grundlage erstellten Zuordnungsliste die durch registrierte Nutzer aktualisiert werden kann.
- Details zur Ablage der Listen, zum Ablauf der Aktualisierung und zur Moderation der Listen sind zu klären, wenn die Zuordnungssoftware erstellt ist und vorliegt, einschließlich der entsprechenden Listen in erster Version.

Für den Aufbau und die Pflege der semantischen Mapping Listen wäre es daher wünschenswert von Seiten des Netzwerks ein Mapping-Tool bereitzustellen, das den Aufbau von Mapping-Listen zur Transformation zwischen Wertelisten der Formate unterstützt und sie für die Konvertierung nutzbar macht. Als konkretes Ergebnis wurde darauf aufbauend für das Netzwerk eine Skizze für einen weiteren Auftrag im Rahmen des BMBF-Vorhabens zum Aufbau einer deutschen Referenzliste für Elementarflüsse sowie eines Tools zur Pflege und Bereitstellung von Referenz- und Mapping-Listen entworfen. Beginn des Projektes ist voraussichtlich im September 2007.

3.3 Ausblick: Nutzung von Ontologien für die semantische Konvertierung

Als langfristiges Konzept für die Verbesserung der Konversion wird der Aufbau eines Ontologiemodells auf Basis von OWL (Ontology Web Language) für den LCI Bereich vorgeschlagen, wie es sich z.B. mit dem frei verfügbaren Tool Protégé (Stanford University) modellieren lässt. Eine solche Ontologie könnte allgemein, unabhängig von einer konkreten Realisierung der Semantik in einem speziellen Datenformat, die in den verschiedenen Datenformaten vorhandenen Informationen strukturieren und miteinander in Beziehung setzen und damit dem Konverter zusätzliche Informationen bei der Konvertierung von Datensätzen liefern, die aus den Datensätzen selbst allein nicht extrahierbar sind. Eine Ontologie könnte außerdem anwendungs- oder rechenalgorithmusspezifische Unterschiede der Datenformate transparent machen und damit dazu beitragen, die Diskussion über Formate transparent zu führen.

Die Ontologie enthält dabei eine Obermenge der strukturellen und semantischen Informationen über die fachlichen Konzepte im Umfeld von LCI, wie sie in den betrachteten Datenformaten zu finden sind. Die Ontologie kann weiter Mengen von Individuen bestimmter Konzepte als „Referenzlisten“ (z.B. eine Referenzliste von Elementarflüssen) beinhalten, die es dem Konverter erlauben, solche Konstrukte (z.B. Flüsse) innerhalb von LCI Datensätzen über die verschiedenen Formate hinweg aufeinander abzubilden.

Als langfristige Idee lässt sich das zu einer „community driven, semantic ontology (of LCA data formats)“ erweitern, die zentral verwaltet und auf einer Website allgemein präsentiert und diskutiert wird und auch das semantische Mapping bereitstellt. Diese Ontologie scheint durchaus aufstellbar, da die Informationselemente in den LCA Formaten sehr ähnlich sind. Damit lässt sich eine formatübergreifende Fachsemantik in der Ontologie abbilden. Die Ontologie ließe sich mit gängigen Werkzeugen wie Protégé auch relativ gut visualisieren und damit von LCA Experten diskutieren. Einzelne Elemente in der Ontologie sind automatisch über URIs eindeutig ausgewiesen und damit in Anwendungen eindeutig referenzierbar.

Eine technische Realisierung zur weitergehenden Entwicklung einer LCA Ontologie wäre über das Jena Semantic Web Framework for Java (<http://jena.sourceforge.net>) und vor allem über die von Jena angebotene API zur Erstellung und Darstellung von Ontologien möglich. Jena unterstützt die gängigsten Ontologie-Sprachen, unter anderem OWL. Damit ließe sich das ontologische Modell zunächst in eine relationale Datenbank exportieren und dann über eine Webanwendung z.B. in Form einer Webdokumentation mit Kommentarfeldern versehen und von LCA Experten, also von „der Community“, kommentieren; bei einer Überarbeitung der Formate sollten diese Kommentare dann einfließen. Weiter könnte die Webanwendung die Ontologie über eine Webservice Schnittstelle anderen Anwendungen zur Verfügung stellen. Mögliche Plattformen für die Webanwendung wären das Netzwerk / FZK für den deutschen Raum, JRC für den europäischen Raum, und das openLCA Projekt für eine allgemeine, weltweite Diskussion mit Unterstützung der UNEP.

Bestimmte Bestandteile des semantischen Modells kommen in allen Datensätzen mehr oder weniger gleich vor, z.B konkrete Objekte der Konzepte

- Unitgroup
- FlowProperty

des ELCD Formates. Solche Konzepte (die Semantik von physikalisch-chemischen Eigenschaften und deren Einheiten) und deren Ausprägungen und Eigenschaften können in einer Ontologie einmalig und zentral gespeichert und dann von einem Datenformat referenziert werden. Auf diese Weise würde man beim ELCD Format z.B. nicht mehr mit jedem Datensatz die entsprechenden Unitgroup- oder FlowProperty-Objekte mitschicken müssen, sondern könnte diese zentral über die Ontologie referenzieren. Bei Wandlung von Ecospold zu ELCD Datensätzen könnte man die benötigten Referenzen direkt aus der Ontologie entnehmen.

Dabei können zum Beispiel die physikalisch-chemischen Größen bereits auf anderen Gebieten der Informationstechnik definiert / standardisiert worden sein. Es sollte angestrebt werden, solche Typbibliotheken dann von dort zu übernehmen. Damit wird eine direkte Anbindung zu den anderen Gebieten ermöglicht. Z.B. könnten für LCA-Datensätze Literaturquellen direkt aus webbasierten Literaturverzeichnissen übernommen werden und die in LCA-Datensätzen definierten Literaturquellen wiederum direkt von den Literaturverzeichnisdiensten mitverarbeitet werden.

Inwieweit eine Ontologie für den LCA-Bereich einen allgemeinen und/oder allgemeingültigen methodischen Oberbau formulieren und welcher weitere konkrete Nutzen daraus gezogen werden kann - sprich welche Anwendungen, bleibt in weiteren Forschungsanstrengungen zu klären. In einem ersten Schritt, wird das Forschungszentrum Karlsruhe, als einer der an diesen Aktivitäten im Netzwerk beteiligten Partner, eine Studienarbeit zur Entwicklung eines ersten Ansatzes für eine solche Ontologie durchführen.

4 Zusammenfassung und Empfehlungen

Mit dem vorliegenden Bericht werden die Ergebnisse eines als Unterauftrag durchgeführten Projektes im Arbeitsprogramm für die IT-Umsetzung des vom BMBF geförderten Forschungsvorhabens ‚Förderung der Wissenschaftskooperation zum Aufbau und Umsetzung des deutschen “Netzwerk Lebenszyklusdaten“ (FKZ 01 RN 0401)‘, kurz ‚BMBF-Vorhaben‘, vorgestellt.

Im Ergebnis verfügt das Netzwerk über den weiter entwickelten Formatkonverter aus dem OpenLCA-Projekt (www.openlca.org) sowie eine Konverter-API, die in das Web-Portal sowie die Versuchsdatenbank integriert wurde. Damit ist zum einen die Geschäftsstelle des Netzwerks in der Lage Datensätze, die im Rahmen des BMBF-Vorhabens in den Formaten ELCD, Ecospold und ISOTS14048_IMI erstellt wurden, anzunehmen und in dem jeweils anderen Format bereitzustellen. Mit der Integration des Konverters in das Webportal ist das Netzwerk zum anderen in der Lage, den Teilnehmern und weiteren Nutzern einen Konvertierungsservice anzubieten, als auch den Upload und Download von Datensätzen in den erwähnten Formaten in und aus der Datenbereitstellung des Netzwerks zu ermöglichen.

Damit verfügt das Netzwerk bei der eigenen Datenbereitstellung bezüglich der Datenannahme und Datenabgabe über höchstmögliche Flexibilität, solange noch keine gemeinsamen Festlegungen bezüglich eines Datenaustauschformates getroffen wurden.

Die Konvertierung von Datensätzen fördert den Austausch von Datensätzen und die Anwendung der Methode der Ökobilanzierung mit qualitativ hochwertigen Datensätzen in

einem heterogenen Umfeld. Sie ist daher ein für das Netzwerk auch bezüglich dieses Aspektes passendes Aktionsfeld.

Seit kurzem verfügbare Werkzeuge für die Konversion von Datensätzen aus dem openLCA Projekt leisten im Prinzip bereits gute Arbeit, sind aber derzeit begrenzt durch die oft sehr unterschiedlichen Semantiken und Strukturen der einzelnen Datenformate. Der Aufbau einer Ontologie und einem daraus abgeleiteten semantischen Mapping mit Mapping Listen, die langfristig unterhalten werden, verspricht einen deutlichen Qualitätssprung.

Aufbau und der Unterhalt dieser Listen stellt daher ein interessantes Betätigungsfeld für das Netzwerk dar, auf dem allerdings auch andere Akteure aktiv werden wollen. Hier ist eine Abstimmung und Abgrenzung zu suchen z.B. zur European Platform on Life Cycle Assessment des Joint Research Centers der Europäischen Kommission und möglicherweise anderen Akteuren. Dies gilt es auch bei dem, weiter oben bereits erwähnten, noch im laufenden BMBF-Vorhaben geplanten Projekt zum Aufbau einer deutschen Referenzliste für Elementarflüsse sowie eines Tools zur Pflege und Bereitstellung von Referenz- und Mapping-Listen zu berücksichtigen.

Die Integration eines Konvertertools hat den Reiz, zum einen einen attraktiven Service auf dem Netzwerkportal anbieten zu können, und zum anderen durch den Upload von Nutzern Zugang zu Datensätzen zu bekommen. Allerdings werden Nutzer aus diesem Grund auch eher zögern, sensible Datensätze über eine Website zu konvertieren. Dies kann durch einen glaubwürdigen „data policy disclaimer“ möglicherweise entschärft werden.

5 Anhang

Das Dokument „Externe Listen“ wurde parallel zum Projekt erstellt und separat übergeben (Abb. Screenshot Inhaltsverzeichnis).

Es enthält sowohl die Darstellung der externen Listen von EcoSpold und ELCD Format als auch ihre Zuordnung in der Konvertierung.

Außerdem listet es, unter den Diskussionspunkten, Änderungsvorschläge für die Formate auf.

Externe Listen	
Andreas Ciroth, Michael Srocka GreenDeltaTC GmbH, Juli 2007	
Externe Listen	1
1 Ziel des Dokuments	3
2 ELCD 1.0.1	3
2.1 Categories	3
2.2 Locations	8
2.3 Contact Data Set	12
2.3.1 ELCD-Contact-DataSetInformationType-UUID	12
2.4 Flow Data Set	12
2.4.1 ELCD-Flow-CompletenessAvailabilityImpactFactorsType-value	12
2.4.2 ELCD-Flow-DataSetInformationType-UUID	13
2.4.3 ELCD-Flow-FlowPropertyType-uncertaintyDistributionType	13
2.4.4 ELCD-Flow-LCIMethodType-typeOfDataSet	14
2.5 Process Data Set	15
2.5.1 ELCD-Process-CompletenessElementaryFlowsType-type	15
2.5.2 ELCD-Process-CompletenessElementaryFlowsType-value	16
2.5.3 ELCD-Process-ExchangeType-dataDerivationTypeStatus	17
2.5.4 ELCD-Process-ExchangeType-dataSourceType	17
2.5.5 ELCD-Process-ExchangeType-exchangeDirection	18
2.5.6 ELCD-Process-ExchangeType-functionType	19
2.5.7 ELCD-Process-DataSetInformationType-UUID	19
2.5.8 ELCD-Process-QuantitativeReferenceType-type	19
2.5.9 ELCD-Process-LCIMethodAndAllocationType-allocationOrSystemExpansionPrinciples	20
2.5.10 ELCD-Process-LCIMethodAndAllocationType-LCIMethodPrinciple	23
2.5.11 ELCD-Process-LCIMethodAndAllocationType-typeOfDataSet	24
2.6 Source Data Set	24
2.6.1 ELCD-Source-DataSetInformationType-publicationType	24
2.6.2 ELCD-Source-DataSetInformationType-UUID	25
2.7 Comment	26
2.7.1 ConformityValues	26
2.7.2 TypeOfOrganisation	26
2.7.3 TypeOfReviewValues	27
2.7.4 WorkflowAndPublicationStatusValues	28
2.7.5 Languages	28
3 EcoSpold	41
3.1 Categories	41
3.2 Units	45
3.3 MetaInformation	45
3.3.1 EcoSpold-MetaInformation-TDataGeneratorAndPublication-countryCode	45
3.3.2 EcoSpold-MetaInformation-TDataGeneratorAndPublication-dataPublishedIn	49
3.3.3 EcoSpold-MetaInformation-TDataGeneratorAndPublication-accessRestrictedTo	49
3.3.4 EcoSpold-MetaInformation-TDataGeneratorAndPublication-companyCode	50
3.3.5 EcoSpold-MetaInformation-TDataSetInformation-type	50
3.3.6 EcoSpold-MetaInformation-TDataSetInformation-languageCode	51
3.3.7 EcoSpold-MetaInformation-TDataSetInformation-energyValues	51
3.3.8 EcoSpold-MetaInformation-TSource-sourceType	51
3.3.9 EcoSpold-MetaInformation-TDataEntryBy-qualityNetwork	51
3.4 FlowData	52
3.4.1 EcoSpold-FlowData-TAllocation-allocationMethod	52
3.4.2 EcoSpold-FlowData-TExchange-uncertaintyType	52
3.4.3 EcoSpold-FlowData-TExchange-inputGroup	52
3.4.4 EcoSpold-FlowData-TExchange-outputGroup	52
3.5 Person	53
3.5.1 TindexNumber	53
4 Mapping	53
4.1 Uncertainty Distribution Types	53
4.2 Input-/Outputgroups	54
4.3 exchangeDirection	54
4.4 QuantitativeReferenceType	54
4.5 Source/ Publication Type	55
4.6 DataSetInformation-type / TypeOfDataSet (Type of Process)	55
4.7 Country Codes	55
4.8 Publication Status	56
4.9 Access Restrictions	56
4.10 Language Code	56
4.11 energyValues	57
4.12 Units	57
4.13 Kein Mapping	59
4.13.1 Aus ELCD	59
4.13.1.1 Categories	59
4.13.1.2 Completeness Types	59
4.13.1.3 CompletenessElementaryFlowsTypeValues	59
4.13.1.4 dataDerivationTypeStatus	59
4.13.1.5 dataSourceType	59
4.13.1.6 CompletenessElementaryFlowsType	59
4.13.1.7 ExchangeType-functionType	60
4.13.1.8 Allocation ELCD	60
4.13.1.9 Process Model Principle	60
4.13.1.10 ConformityValues	60
4.13.1.11 TypeOfOrganisation	61
4.13.1.12 TypeOfReviewValues	61
4.13.1.13 WorkflowAndPublicationStatusValues	61
4.13.2 Aus EcoSpold	61
4.13.2.1 Categories	61
4.13.2.2 companyCode	61
4.13.2.3 allocationMethod	61
5 "Diskussionspunkte"	61