Netzwerk Lebenszyklusdaten

Arbeitskreis METHODIK

Studie „Semantisches Mapping“ im Netzwerk Lebenszyklusdaten

Projektbericht

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

Ifu Hamburg GmbH

Forschungszentrum Karlsruhe, Institut für Angewandte Informatik

Forschungszentrum Karlsruhe, Institut für Technische Chemie, Zentralabteilung Technikbedingte Stoffströme

Hamburg Karlsruhe – Dezember 2007

Hrsg.: Forschungszentrum Karlsruhe
Institut für Technikfolgenabschätzung und Systemanalyse –
Zentralabteilung Technikbedingte Stoffströme

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

Bundesministerium für Bildung und Forschung

Studie
„Semantisches Mapping“ im Netzwerk Lebenszyklusdaten

Autoren:

Jan Hedemann, Peter Müller-Beilschmidt
ifu Hamburg GmbH

Dr. Clemens Düpmeier
Institut für Angewandte Informatik, Forschungszentrum Karlsruhe (FZK-IAI)

Achim Stadtherr
Zentralabteilung Technikbedingte Stoffströme, Institut für Technische Chemie, Forschungszentrum Karlsruhe (FZK-ITC-ZTS)

Kontakt:
Jan Hedemann
ifu Hamburg GmbH
Große Bergstraße 219
D-22767 Hamburg
T: +49 40 480009-15
F: +49 40 480009-22
E-mail: j.hedemann@ifu.com

Inhalt

1 Einleitung

Im Rahmen des Netzwerks Lebenszyklusdaten werden von verschiedenen Partnern Daten in unterschiedlichen Formaten bereitgestellt. Ebenso haben die Netzwerkpartner und die Zielgruppe des Netzwerks unterschiedliche Anforderungen an die Formate, in denen die Daten bereitgestellt werden sollen. Hier entsteht der dringende Bedarf unterschiedliche Formate ineinander zu konvertieren.

Wenn LCA Daten zwischen verschiedenen Formaten konvertiert werden, dann kann zwischen syntaktischer und semantischer Konvertierung unterschieden werden. Die syntaktische Konvertierung ist Gegenstand einer anderen Studie im Netzwerk. Dort wurden erste Überlegungen zur semantischen Konvertierung vorgestellt. Darauf aufbauend werden in dieser Studie diese Überlegungen zum semantischen Mapping weiter ausgebaut.

Ausgangspunkt der Überlegungen ist das Handbuch Methodik des Netzwerks Lebenszyklus-daten. Werkzeuge sind hier Konzepte und Methoden (z.B. Ontologien) und deren Anwendung (Ontologie mittels Reasoner), sowie intelligente Hilfesysteme.

1.1 Vision

Ziel ist es, die Voraussetzungen für die Verwendung von LCA Datensätzen aus unterschiedlichen Quellen in einem System zu schaffen. Dazu ist es erforderlich, dass die Datensätze auch inhaltlich zueinander passen. Wichtiges Kriterium ist dabei die Nomenklatur der Elementarflüsse.

Es ist unrealistisch, dass sich weltweit *die eine* Referenzliste etabliert. Damit bleibt als Lösung das Mapping der verschiedenen Nomenklaturen aufeinander.

Die Vision einer weltweiten für den Anwender transparenten LCA Datenbank knüpft sich an die Vision des semantischen Webs (Tim Berners-Lee et al., The Semantic Web, Scientific American 5/01, S.34), die davon ausgeht, dass die Informationen und Daten im Internet künftig maschinenlesbar und maschinenauswertbar werden.

1.2 Syntaktisches Mapping

Die syntaktische Konvertierung von LCA Datensätzen bedeutet, die Daten technisch in eine andere Struktur zu bringen. Bspw. von einer Speicherung in einer Textform (SPOLD99) in eine XML-Form (EcoSpold) oder von einer XML-Form in eine andere (EcoSpold nach ELCD). Hierbei findet eine Abbildung (Mapping) von Datenfeldern im einen Format auf die passenden Datenfelder im anderen Format statt.

Dabei ist natürlich die Bedeutung (Semantik) der Datenfelder selbst zu beachten. Ist das Mapping einmal definiert, so findet eine syntaktische Umwandlung von einem Format in ein anderes statt. Der konkrete Inhalt der Datenfelder wird dann nicht mehr beachtet.

1.3 Semantisches Mapping

Damit die konvertierten Datensätze im Zielsystem genutzt werden können, ist es erforderlich, dass auch eine semantische Konvertierung der Daten stattfindet. Das heißt, dass zwischen den Formaten bedeutungsgleiche Begriffe ineinander umgewandelt werden. Hierfür ist also eine semantische Abbildung notwendig. Herausragendes Beispiel ist die unterschiedliche Benennung oder Schreibweise von Elementarflüssen, die jedoch semantisch gleich sind. In jedem LCA System ist die inhaltliche Bedeutung von *"Kohlendioxid-Emissionen in die Luft aus der Verbrennung von fossilen Brennstoffen"* gleich. Es wird deutlich, dass es nicht einfach um den Namen "Kohlendioxid" geht, sondern dass weitere Eigenschaften eine Rolle spielen. Man spricht auch von einem semantischen Objekt. Das semantische Objekt *"Kohlendioxid-Emissionen in die Luft aus der Verbrennung von fossilen Brennstoffen"* wird nun in den verschiedenen Systemen und Formaten auf ganz unterschiedliche Weise repräsentiert. Die syntaktische Umwandlung interessiert uns hier nicht, sondern die Möglichkeit zu beschreiben, wie die Gleichheit semantischer Objekte in verschiedenen Systemen beschrieben werden kann.

1.4 Warum brauchen wir semantisches Mapping?

In den folgenden Abschnitten werden anhand von konkreten Beispielen die Fallstricke aufgezeigt, die bei der Konvertierung von LCA Daten lauern. Daran wird deutlich, dass eine Konvertierung nicht auf der Syntaxebene stehen bleiben kann, wenn eine sinnvolle Verwendbarkeit von LCA Daten über die Grenzen einzelner Datenbanken hinweg möglich werden soll.

Die Notwendigkeit der Nutzung unterschiedlicher Datenbanken lässt sich auf unterschiedliche Ebenen fokussieren.

- Der Anwender möchte sich aus den verfügbaren Datensätzen, die für den Zweck seiner Studie am besten geeigneten Datensätze heraussuchen können. Keine Datenbank kann bislang für sich in Anspruch nehmen, die Bedürfnisse der Anwender umfassend zu befriedigen.
- Auch in Zukunft wird es sinnvoll bleiben, nicht *die eine* Datenbank mit einer umfassenden Liste von LCA Daten zu erstellen. Abgesehen von unterschiedlichen methodischen Vorgaben und Interessen, wird es Datenbanken geben, die lokale und branchenspezifische Schwerpunkte setzen und damit ihren Beitrag zum weltweiten Datenpool leisten.
- Für die Aktualisierung von Datensätzen ist anzustreben, dass LCA Datensätze auf der Basis von regelmäßig aktualisierten Datenbanken, auch aus nicht-LCA Bereichen, spezifiziert werden. Damit wird die Aktualisierung automatisierbar. Zu solchen Basisdatenbanken gehören z.B. die Daten des Zentralen Systems Emissionen (ZSE) des Umweltbundesamtes, die aufgrund von nationalen Reportingpflichten regelmäßig fortgeschrieben werden. Auch diese Datenbanken haben eine eigene Nomenklatur und die Daten müssen automatisiert konvertiert werden.
- Die Anbieter von LCA Softwareprodukten sind bestrebt, ihren Anwendern eine umfassende Datenbasis zu bieten. Auch hier ist deshalb der Rückgriff auf unterschiedliche Datenquellen von Interesse. Dem Anwender soll für eine komfortable Nutzung eine konsistente Datenbasis bereitgestellt werden. Dazu ist

derzeit noch ein aufwändiges manuelles Mapping durch den Softwarehersteller notwendig oder das Mapping wird dem Anwender überlassen. Hier lassen sich mit semantischem Mapping deutlich bessere Ergebnisse erzielen.

2 Grundlagen

2.1 Semantisches Objekt

Ein *semantisches Objekt* ist ein Objekt, welches eine eigene Bedeutung hat. Unterschiedliche Bezeichnungen können für das gleiche semantische Objekt stehen. Die Schwierigkeit besteht darin, zu erkennen und zu definieren, wann zwei Objekte semantisch gleich sind. Dazu lassen sich verschiedene Eigenschaften des Objektes heranziehen.

Objekte, die sich durch gleiche Eigenschaften (nicht Eigenschaftswerte) auszeichnen, lassen sich durch eine Klasse beschreiben.

Objekte sind die Instanzen einer Klasse mit konkreten Eigenschaftswerten.

Beispiel "semantisches Objekt"

Ein Objekt der Klasse "Elementarfluss" lässt sich durch folgende Eigenschaften beschreiben:

- Name
- CAS-Nummer
- Kompartiment
- ID
- Chemische Formel
- *usw.*

Zusätzlich ist zu definieren, welche Eigenschaften darauf schließen lassen, dass zwei Objekte semantisch gleich sind.

2.2 Ontologie

Eine Ontologie ist ein formales Modell für einen speziellen Anwendungsbereich, der die Kommunikation zwischen Kommunikationspartnern unterstützt und dadurch den Austausch und die Teilung von Wissen erleichtert#[1]. Somit ist eine Ontologie eine Repräsentationsform, welche durch verschiedene Menschen und durch den Computer interpretiert und weiterverarbeitet werden kann. Durch die Verwendung eines gemeinsamen Vokabulars findet eine Standardisierung von Sprache und Kommunikation über die festgelegte Terminologie statt.

Ontologien bestehen aus vier Teilen:

- **Lexikon**: beinhaltet alle Begriffe und semantische Relationen
- **Begriffe**: beschreiben ein Konzept des Anwendungsbereichs

- **Semantische Relationen**: setzen die Begriffe zueinander in Beziehung
- **Regelhafte Zusammenhänge**: erfassen zusätzliche Bedeutungsinhalte von Begriffen und Relationen (z.B. invers, transitiv). Mit regelhaften Zusammenhängen kann man Antworten auf Fragen bereitstellen, ohne dass diese Antworten explizit in der Ontologie beschrieben sind.

Mit dem Einsatz eines Reasoners (logisch Denkender) wird die Konsistenz einer Ontologie überprüft und die Regeln werden durch Schlussfolgern ausgewertet, wodurch neues Wissen gewonnen werden kann, welches nicht explizit in die Ontologie geschrieben wurde.

Eine Möglichkeit, eine Ontologie formal zu beschreiben ist die **Web Ontology Language (OWL)**, eine Spezifikation des W3C. OWL wurde als Ontologiesprache für das Web entwickelt. Es existieren 3 Untersprachen von OWL mit zunehmender Ausdrucksmächtigkeit:

- OWL Lite: zum Erstellen von Ontologien, die einfache Klassenhierarchien mit einfachen Regeln enthalten.
- OWL DL: ist für Benutzer, die das Maximum an Ausdrucksmächtigkeit haben möchten und behält dabei die vollständige Verarbeitbarkeit und Entscheidbarkeit#[2]
- OWL Full: hier werden dieselben Konstrukte wie bei OWL DL benutzt, es existieren hier aber nicht die Einschränkungen von OWL DL. Dadurch sind die OWL Full Ontologien nicht entscheidbar.

Der Nutzen einer Ontologie liegt im Bereich der LCA vor allem in der Integration heterogener Daten. Bei heterogenen Daten gibt es oft das Problem der Synonyme (verschiedene Begriffe, gleiche Bedeutung) und der Homonyme (gleiche Begriffe, verschiedene Bedeutungen). Dies erschwert die Weiterverarbeitung. Durch eine Integration verschiedenster Datentypen in eine Ontologie wird eine Interoperabilität in verschiedenen Systemlandschaften hergestellt.

Ein weiterer Verwendungszweck ist die Ontologie als Basis für eine fachliche Dokumentation, z.B. als semantisches Wiki, zu benutzen.

2.3 Basisbegriffe

Begriffe werden in unterschiedlichen Domänen unterschiedlich definiert und verwendet. Zur Begriffsbestimmung wird im folgenden festgelegt, wie die wichtigsten Basisbegriffe hier verwendet werden.

2.3.1 Referenzliste

Eine Referenzliste erhebt den Anspruch, eine Vorgabe für einen bestimmten Geltungsbereich zu machen. In unserem Kontext bedeutet dies meist, eine Namensliste von Elementarflüssen, Kategorien oder Prozessen aufzustellen.

Referenz ist hier im Sinne von Vorgabe gemeint, eine Liste auf die referenziert werden kann.

So ist der Umberto-Materialbaum eine Referenzliste für alle Prozesse in der Umberto Datenbank. Ebenso hat Ecoinvent eine Referenzliste entwickelt, die für die Ecoinvent-Datenbank verbindlich ist.

Offen bleibt die Frage, ob das deutsche Netzwerk Lebenszyklusdaten eine eigene Referenzliste aufbauen sollte.

Synonym für Referenzliste ist Masterliste.

Sinnvoll erscheint es, Referenzlisten für eine Anwendungsdomäne aus bestehenden breiter definierten Referenzlisten abzuleisten. So könnte die Liste aller Substanzen mit CAS Nummern (siehe http://www.cas.org) eine Ausgangsbasis für die Ableitung einer Elementarflussliste sein. Ähnliche Ausgangspunkte sind statistische Klassifikationen, wie NACE.

2.3.2 Referenzierungsliste

Eine Referenzierungsliste oder Abbildungsliste oder Mappingliste unterstützt die Konvertierung von einer Referenzliste in eine andere. Soweit eine 1:1 Abbildung von Einträgen einer Referenzliste auf Einträge einer anderen Referenzliste möglich ist, lässt sich dies bereits heute leicht bewerkstelligen. In dieser Studie wird vor allem auf die Fälle fokussiert, die darüber hinausgehen.

2.3.3 Fluss

Für Flüsse oder Ströme zwischen Prozessen und zwischen den Systemgrenzen und Prozessen gibt es unterschiedliche Bezeichnungen. Das Erkennen von Synonymen und Relationen zwischen den Begriffen ist eine wichtige Grundlage für das semantische Mapping.

Wichtige Begriffe sind:

Elementarfluss, Technosphärenfluss, Material, Strom, Stoff, Produktfluss, Zwischenprodukt und weitere.

Für eine ausführliche Diskussion und Aufnahme dieser Begriffe in eine Ontologie sollte auf bestehende Glossare zurückgegriffen werden. Quellen sind hier das Wiki des Netzwerks Lebenszyklusdaten, DIN- und ISO-Normen, verschiedene Projekte und die Wikipedia.

Für diese Studie ist die Unterscheidung zwischen Flüssen zwischen dem betrachteten System und dessen Umwelt, sowie innerhalb des Systems von Bedeutung. Diese werden im folgenden definiert.

2.3.4 Elementarfluss

Ein Elementarfluss (Synonyme: to/from nature, elementary flow, Stoff) ist ein Input oder Output in das betrachtete System. Er fließt also über die Systemgrenzen hinweg. Zudem wird er üblicherweise als nicht weiter unterteilbar betrachtet.

2.3.5 Technosphären-Fluss

Ein Technosphären-Fluss ist das Gegenteil vom Elementarfluss. (Synonyme: Nicht-Elementarfluss, Economic flow, technosphere flow, to/from technosphere, process flow). Ein Technosphären-Fluss fließt zwischen den Prozessen des betrachteten Systems.

2.4 Angrenzende Aktivitäten

2.4.1 Syntaktisches Mapping

In der Studie des Netzwerks Lebenszyklusdaten „Konzept zur Unterstützung der Konvertierung von Datensätzen für das Netzwerk Lebenszyklusdaten“ von Cieroth et.al. werden die Probleme der syntaktischen Konvertierung und deren Lösungsmöglichkeiten ausführlich diskutiert. Es wird deutlich, dass der Übergang zwischen syntaktischem und semantischem Mapping fließend ist. Das semantische Mapping beginnt bereits, wenn für ein Datenfeld eine Nomenklatur vorgeschrieben ist und diese Nomenklatur im Zielformat anders definiert ist. Die Abbildung der Nomenklaturwerte aufeinander ist streng genommen bereits ein semantisches Mapping, da die Bedeutungen der Einträge relevant sind.

2.4.2 Nomenklatur-Mappings

Folgende Nomenklaturen-Mappings gibt es bereits oder es wird an ihnen gearbeitet. Diese Arbeiten sind – soweit möglich - zu berücksichtigen, um Doppelarbeit zu vermeiden und eine breite Anwendbarkeit zu sichern

- In der Stoffstrommanagement-Software Umberto wurde ein Mapping der Ecoinvent-Daten auf die Umberto Daten und in die Gegenrichtung gemacht.
- In der LCA Software GaBi wurde ein Mapping der Ecoinvent Daten auf die GaBi Daten gemacht.
- Der OpenLCA Konverter implementiert ein syntaktisches Mapping zwischen ELCD-Daten und Ecoinvent-Daten in beide Richtungen.
- In der UNEP/SETAC Life Cycle Initiative, Taskforce 2 wurde an der Definition eines gemeinsamen weltweiten Austauschformats gearbeitet. Dabei wurden umfassend Fragen des semantischen Mappings diskutiert.
- COST Action 530, Working Group Databases hatte das Ziel der Harmonisierung eines europäischen Austauschformats. Hierbei wurden auch Mapping Fragestellungen diskutiert.

2.4.3 Nomenklatur Definitionen

Neben bestehenden Mappings sind auch wichtige bestehende Nomenklaturen zu berücksichtigen. Auf diese sollte sich ein semantisches Mapping mindestens erfolgreich anwenden lassen.

- Die Ecoinvent-Datenbank definiert eine umfangreiche Nomenklatur für ihre Datensätze.
- Die JRC-LCA Gruppe hat einen Auftrag zur Erarbeitung einer Nomenklatur der ELCD-Datenbank vergeben.
- Die Software Umberto umfasst eine Datenbank mit Ökobilanzprozessen, die einer einheitlichen Nomenklatur folgen. Diese ist im Umberto "Materialbaum" definiert.

- Die Software GaBi umfasst eine Datenbank mit Ökobilanzprozessen, die einer einheitlichen Nomenklatur folgen.
- Im Zentralen System Emissionen (ZSE) des Umweltbundesamtes werden Daten mit einer Nomenklatur bereitgestellt.

2.4.4 Existierende Ontologien

Wünschenswert ist es, bereits bestehende Ontologien unverändert und per Referenz zu importieren, um auf Arbeiten in anderen Anwendungsfeldern aufzusetzen. Beispiele für solche Ontologien sind Beschreibungen von Maßeinheiten mit ihren Umrechnungen, CAS Nummern, Übersetzungen und Ergebnisse internationaler Forschungsprojekte.

2.5 Anwendungsfälle für semantisches Mapping

Bei der Entwicklung eines semantischen Mappings sind die verschiedenen Anwendungsfälle (Use Cases) zu berücksichtigen. Soweit heute erkennbar, sollen diese im folgenden kurz beschrieben werden.

2.5.1 Konvertierung

Die Konvertierung von Datensätzen kann unterschieden werden nach der Menge der Datensätze, also ganze Datenbanken oder Einzeldatensätze zu konvertieren. Weiterhin kann unterschieden werden nach den Quellen und Zielen der Konvertierung, insbesondere ob es um Konvertierungen zwischen Datenbanken, Softwarewerkzeugen oder individuellen Anwendern geht.

2.5.2 LCA Software

Zusätzlich zu den anderen Anwendungsfällen umfasst der Einsatzbereich des semantischen Mappings in Softwaretools noch speziell:

- Import- und Exportroutinen aus Datenbanken oder anderen LCA Werkzeugen.
- Konvertierung von Datensätzen zu Anzeigezwecken (z.B. in standardisierten Dokumentationsformaten)
- Übersetzungen von Datensätzen in Fremdsprachen

2.5.3 Validierung

Bestehende Datensätze können daraufhin validiert werden, ob sie problemlos einem semantischen Mapping unterworfen werden können. Ist also beispielsweise jeder Elementarfluss des Datensatzes von der Definition des semantischen Mappings erfasst. Ist das nicht der Fall, so könnte bspw. Tippfehler in der Namensgebung vorliegen.

2.5.4 Bewertung (LCIA)

Bewertungssysteme haben wiederum eine eigene Nomenklatur. Um diese auf eine Sachbilanz (LCI) anwenden zu können, ist ein Mapping erforderlich.

2.5.5 Integration von Datenbanken

Bereits verfügbare Datenquellen sollen in eine Datenbank integriert werden, um dem Anwender einen konsistenten und komfortablen Zugriff zu ermöglichen.

2.5.6 manueller vs. automatischer Ablauf

Jeder der vorstehend beschriebenen Anwendungsfälle kann noch einmal danach unterteilt werden, ob ein manueller Eingriff des Anwenders zulässig und sinnvoll ist oder nicht. Die Anforderungen für einen vollautomatischen Ablauf sind andere, als wenn bestimmte Entscheidungen an den Anwender delegiert werden können.

3 Semantisches Mapping von LCA Datensätzen

Die zentrale Frage für das Mapping von Nomenklaturen sind:

- Wann sind zwei Objekte semantisch gleich?
- Wie kann die Abbildung von semantischen Objekten zweier Datenquellen aufeinander möglichst vollständig erfolgen?

3.1 Teilbereiche des semantischen Mappings

Um LCA Datensätze komplett aufeinander abzubilden, sind verschiedene Bereiche zu bearbeiten. In den folgenden Abschnitten wird auf jeden Bereich detailliert eingegangen.

3.2 Elementarflüsse

Das Mapping der Elementarflüsse spielt eine zentrale Rolle, da die Elementarflüsse eines LCA Datensatzes die Träger der wichtigsten Informationen sind. Das Mapping scheint auf den ersten Blick einfach zu sein. So ist es simpel die Texte "CO2", "Kohlendioxid" und "carbon dioxide" als semantisch gleich zu erkennen. Die Probleme beginnen jedoch sofort, wenn über den Text hinaus das gesamte Objekt "Kohlendioxid Emission" betrachtet wird. Die physikalische Einheit könnte unterschiedlich sein, das Kompartiment (Luft, Boden, Wasser) in welches emittiert wird, ist zu betrachten und möglicherweise ist auch noch ein Raumbezug zu berücksichtigen. Bei Kohlendioxid tritt dann noch die spezielle Eigenschaft der Entstehung auf, die u.a. mit den Zusätzen "fossil" und "regenerativ" gekennzeichnet wird.

Folgende prototypischen Fälle sollen beispielhaft die Problematik des Mappings zwischen verschiedenen Nomenklatur-Systemen für Elementarflüsse veranschaulichen. Die Beispiele basieren auf Testfällen und Mapping-Versuchen von Datensätzen aus folgenden Datenbanken:

- ELCD
- Ecoinvent v1.3
- Umberto 5.5
- GaBi

- GEMIS

Die Beispiele beziehen sich auf einen Anwendungsfall in dem Datensätze aus einer "neu hinzukommenden Nomenklatur" in eine "existierende Nomenklatur" abgebildet werden.

3.2.1 Prototypischer Fall 1: 1:1 Mapping

Beispiel 1:1 Mapping	
Existierende Nomenklatur: Umberto	Neu hinzukommende Nomenklatur Ecoinvent v1.3
Benzo(a)pyrene	Benzo(a)pyrene
Type: - Unit: kg Category: a Subcategory: Location: - Infrastructure: N (default)	Type: Elementary Flow Unit: kg Category: Air Subcategory: unspecified Location: - Infrastructure: N

Die Namen sind gleich, sie unterscheiden sich nur durch Groß- und Kleinschreibung. Die Einheiten sind gleich. Die Kategorie lautet zwar unterschiedlich, ist aber eindeutig, und kann somit abgebildet werden ("a" ←→ "air"). Die existierende Referenzliste hat keine Festlegung bei den Feldern "Elementary Flow", "Subcategory", "Location" und "Infrastructure", diese Angaben können also ergänzt oder ignoriert werden. Letztlich eine Situation die durch syntaktisches Mapping erfolgreich abgebildet werden kann.

3.2.2 Prototypischer Fall 2: Elementarfluss nur in einer Nomenklatur vorhanden

Beispiel	
Existierende Nomenklatur: Ecoinvent v1.3	Neu hinzukommende Nomenklatur ELCD
-	1-Butoxypropanol

Der Fluss ist nicht in existierender Nomenklatur vorhanden.

Es wird im Namen und in den Synonymen in der gleichen Kategorie geprüft. Es wird im Namen und in den Synonymen der anderen Kategorien geprüft. Wenn keine Übereinstimmung gefunden wird, kann davon ausgegangen werden, dass es sich um einen neuen Fluss handelt. Er muss ergänzt werden. Hierbei müssen Regeln (wie Einordnung in Kategorien) der zu ergänzenden Nomenklatur beachtet werden.

3.2.3 Prototypischer Fall 3: ähnliche Bedeutung, Eigenschaft Ortsbezug unterschiedlich präzise

Beispiel	
Existierende Referenzliste ELCD	Neu hinzukommende Nomenklatur Ecoinvent
Name: Crude oil South Africa	Oil, crude, in ground
Type: exchange Unit: kg Category: Resources Subcategory: Non-renewable energy resources Location: -	Type: Elementary Flow Unit: kg Category: resource Subcategory: in ground Location: RAF (Region Africa)

Fall 1:1 Relation, unterschiedlicher Name oder Schreibweise, die sich auflösen lässt. ELCD Datensatz enthält geographischen Bezug im Namen des Elementarflusses. Allerdings handelt es sich um eine unterschiedlich grobe geographische Auflösung.

Das Mapping müsste in diesem Fall, nach Identifikation des Namensbestandteil „South Africa“ ein Mapping mit einer Länderliste der Region Africa (RAF) bewerkstelligen → Class in semantischem Mapping. Mapping South Africa → RAF ist eindeutig, in die Gegenrichtung jedoch nicht, da es auch Crude oil Zimbabwe geben könnte

Bei mehreren möglichen Zuordnungen müsste der User entscheiden.

3.2.4 Prototypischer Fall 4: N:1 Mapping auflösbar

Beispiel Xenon	
Existierende Referenzliste Ecoinvent v1.3	Neu hinzukommende Nomenklatur Umberto
Xenon, in air	xenon (Xe)
Unit: kg Category: resource Subcategory: in air Location: - Infrastructure: N	Unit: kg Category: r Subcategory: Location: - Infrastructure: N (default)
Xenon-133 Xenon-133m Xenon-135 Xenon-135m Xenon-137 Xenon-138	
Unit: kBq Category: air Subcategory: high population density	
Unit: kBq Category: air Subcategory: low population density	
Unit: kBq Category: air Subcategory: low population density, long term	
Unit: kBq Category: air Subcategory: lower stratosphere + upper troposphere	
Unit: kBq Category: air Subcategory: unspecified	

Dieser Fall einer N:1 Relation ist reduzierbar auf ein 1:1 Mapping bei allerdings noch ein unterschiedlicher Name bzw. Schreibweise berücksichtigt wird. Die Namen sind unterschiedlich. Die Einheit stimmt überein. Die Kategorie lautet zwar unterschiedlich, ist aber eindeutig, und kann somit abgebildet werden (“r“ ←→ “resource“).

Es kommt jedoch hinzu, dass es mehrere weitere Elemente in der existierenden Referenzliste gibt. Diese haben einen gleichen Namensstamm jedoch eine unterschiedliche Einheiten und mehrere unterschiedliche Kategorien. Es muss hier nach einer Vorrangregel Name – Unit - Kategorie – Subkategorie geprüft werden.

3.2.5 Protypischer Fall 5: 1:N Mapping mit Aggregation

Beispiel	
Existierende Nomenklatur: Ecoinvent 1.3	Neu hinzukommende Nomenklatur Umberto
Chlorinated solvents, unspecified	hexachloro-1,3-butadiene heptachlor 1,2,3,4-tetrachlorobenzene 1,2,3,5-tetrachlorobenzene 1,2,3-trichlorobenzene 1,2,4,5-tetrachlorobenzene 1,2,4-trichlorobenzene 1,2-dichlorobenzene 1,3,5-trichlorobenzene 1,3-dichlorobenzene 1,4-dichlorobenzene 1-chloro-4-nitrobenzene 2,3,4,6-tetrachlorophenol 2,4,5-trichlorophenol 2,4,6-trichlorophenol 2,4-dichlorophenol 2-chlorophenol 3,4-dichloroaniline 3-chloroaniline 4-chloroaniline benzylchloride pentachlorobenzene pentachloronitrobenzene pentachlorophenol
Unit: kg Category: water Subcategory: ocean Location: - Infrastructure: N	Unit: kg Category: sw Subcategory: Location: - Infrastructure: N (default)

Es gibt in der neu hinzukommenden Referenzliste mehrere Elemente, die in der Zielnomenklatur nicht jeweils ein eigenes Äquivalent haben. Statt dessen wird ein entsprechender Sammelname verwendet.

Die neuen Bezeichnungen können als Attribute (Synonyme) für die Zuordnung in Richtung N→1 eingetragen werden, eine Rückübersetzung (Rückmapping) ist aber nicht mehr möglich.

Bei einer Zuordnung zum Sammelnamen ist der Umgang mit Materialmengen zu klären. Folgenden Fälle sind möglich:

- Addition der Mengen
- gewichtete Aggregation der Mengen anhand von Eigenschaften

3.3 Kategorien

Aufgrund der Vielzahl von Elementarflüssen und Prozessen in Ökobilanz-Datenbanken hat sich die Nutzung von Kategorien eingebürgert, um eine zusätzliche Strukturierung einzuführen.

Kategorien werden dabei sehr unterschiedlich genutzt und sie unterscheiden sich teilweise deutlich in ihrer Semantik.

Kategorien für Prozesse, also für Ökobilanz-Datensätze, sollen vor allem für eine übersichtliche Gruppierung sorgen und sind meist ohne relevante Semantik. Sie werden hier deshalb nicht weiter betrachtet.

Das Mapping von Elementarfluss-Kategorien hat eine zentrale Bedeutung, da in verschiedenen Nomenklaturen die Kategorie in die eindeutige Bezeichnung des Flusses eingeht und eine relevante Semantik besitzt. Die Kategorien für Elementarflüsse werden im Folgenden anhand konkreter Ausprägungen untersucht.

3.3.1 ecoinvent

Ecoinvent beschränkt sich auf eine zweistufige Kategorisierung aus Haupt- und Unterkategorie (category/subcategory).

Die Hauptkategorien in ecoinvent lauten "resource", "air", "water", "soil". Gemeint ist immer "from resource", "to air", "to water", "to soil". Damit sind bei den Emissionen die Kompartimente als Hauptkategorien abgebildet. Auf der Ressourcenseite (Inputseite) ist das nicht so eindeutig, da die Ressourcen eigentlich ebenfalls nach Kompartimenten unterschieden werden könnten.

Unterhalb der Hauptkategorien sind meist räumliche Informationen als Unterkategorien angeordnet. Diese haben zumindest teilweise Bedeutung für die Bewertung (LCIA).

Sowohl die Haupt- als auch die Unterkategorien sind schnittmengenfrei.

3.3.2 ELCD

ELCD beschränkt sich faktisch auf eine zweistufige Kategorisierung aus Haupt- und Unterkategorie (top category/sub category 1).

In ELCD sind auf der obersten Ebene teilweise Unterteilungen im Sinne der Bewertung gemacht (sea water, fresh water). Auf der zweiten Ebene sind immer die gleichen Kategorien angeordnet, die jedoch keinerlei identifizierende Bedeutung haben. Problematisch ist, dass diese Subkategorien von ihrer Bedeutung her nicht schnittmengenfrei sind.

Daher sind für die Einordnung von Flüssen in diese Kategorien noch zusätzliche Regeln erforderlich.

3.3.3 Umberto

In Umberto ist eine beliebig tiefe Kategorisierung möglich. Die Kategorien heißen hier Materialgruppen (material groups). Sie haben keine Bedeutung für die Identität und dienen auch bei Elementarflüssen lediglich der übersichtlichen Strukturierung für den Anwender.

3.3.4 Mappingansätze

Als mögliche Lösung wurde die Idee diskutiert, für das Mapping einen Kategorienbaum zu entwickeln, der die Kategorien vom Quell- und Zielsystem vereint. Ein Elementarfluss muss dann an verschiedenen Stellen des Baumes eingehängt werden können.

Für die semantische Eindeutigkeit ist es wichtig, dass die Kompartiment Information von der Kategorie separiert wird. Für das korrekte Mapping ist das richtige Kompartiment relevant, die Einordnung in Kategorien ist hingegen ein Ordnungskriterium, welches in unterschiedlichen Systemen und Anwendungsfeldern unterschiedlich aussehen kann ohne die Bedeutung zu verändern.

3.4 Wo treten weitere Herausforderungen auf?

Der Vollständigkeit halber sollen hier weitere mögliche Problemfelder beschrieben werden.

Die Herausforderungen beim syntaktischen Mapping (wie Feldlängen, Feldtypen, Felddifferenzierung) werden in der Studie von Cieroth et.al. behandelt.

Weitere Herausforderungen und „Stolpersteine“ können sein:

- Eindeutigkeit (wie ist diese definiert?)
- Sprachspezifische Fragen (de, en, ..)
- Modell/Modellierungsspezifisch (ELCD, EcoSpold, …)
- Technosphärenflussliste (Prozessflüsse, Zwischenprodukte, …)
    - Kategorien
    - Raumbezug
    - Infrastrukturkennzeichnung
    - Unterschiedliche Namenskonventionen
- Schlüsselwortlisten
- Eigenschaften von Flüssen
- Das Thema LCIA stellt nochmals eigene Anforderungen.

3.5 Prioritäten

Für die Umsetzung eines semantischen Mappings schlagen wir folgende Prioritäten vor, mit denen eine Umsetzung anzugehen wäre.

1. Flusslisten mit Kategorien
2. Prozessnamen mit Kategorien
3. Schlüsselwortlisten

4 Architekturüberlegungen

Es erscheint attraktiv einen Mapping-Services auf Basis von semantischem Mapping und Ontologien als zentrale Dienstleistung auf einem Webserver anzubieten. Die Vorteile der breiten Nutzbarkeit und guten Wartbarkeit liegen auf der Hand. Empfehlenswerte Technologien sind die Bereitstellung als Web-Services, um eine plattformunabhängige Nutzung zu ermöglichen.

Die zentrale Wartung und Pflege erlaubt einen schrittweisen Ausbau der Mappingbeschreibungen mit dem Ziel eine Konvertierung zu 100% automatisiert abzuwickeln.

Der breite Einsatz eines solchen Dienstes auch in Softwarewerkzeugen wird angestrebt. Zu Bedenken ist dabei, dass die Funktionalität dann auch prinzipiell ohne Internetzugang zu Verfügung stehen sollte. Wünschenswert ist dabei ein für den Anwender transparenter Cachingmechanismus, der bei verfügbarem Onlinezugang den zentralen Dienst mit den aktuellsten Daten nutzt und im Offline-Modus den letzten verfügbaren Stand im Zugriff hat. Als Vorbild kann ein IMAP Emailserver und die zugehörige Email-Clientsoftware dienen. Der IMAP Client bietet die Möglichkeit Emailordner für die Offline-Verfügbarkeit zu kennzeichnen. Dies führt dazu, dass Kopien der Emails auf dem Clientrechner gespeichert werden. Geht der Anwender offline, hat er diese Mails in der gleichen Ansicht zur Verfügung, als wenn er online wäre.

Bei der nächsten Online-Verbindung zum Server werden die Emails wiederum synchronisiert. Bei Bedarf ohne weiteren Eingriff des Anwenders.

Für die organisationsübergreifende Akzeptanz eines Mapping-Dienstes kann es wichtig werden, dass es nicht DEN EINEN Hoster des Dienstes gibt. Deshalb sollte ein verteiltes System konzipiert werden, bei dem sich mehrere Server miteinander synchronisieren können. Zudem könnten Teile der zugrundeliegenden Ontologien auch auf unterschiedlichen Servern liegen und jeweils dort gepflegt werden.

5 Zusammenfassung und Ausblick

In der Landschaft der LCA Datenbanken existieren sowohl technisch, als auch aus Sicht der Benutzer bisher ausschließlich Insellösungen (ecoinvent, ELCD, NIRE, NREL,…), die darauf warten miteinander kombiniert zu werden. Der Benutzer wünscht sich, dass Beste aus jeder der Datenbanken in seinen Untersuchungen und Modellen verwenden zu können. Die

Herausforderung besteht darin, diese Dateninseln trotz unterschiedlicher Semantik zu vernetzen.

Hier sehen wir ein erhebliches Innovationspotential durch den Einsatz geeigneter IT-Methoden und –Werkzeuge. Komplexe Herausforderungen erfordern meist auch komplexe Lösungen, die für den Anwender dann einfach zu handhaben sein müssen. Die Studie zeigt, dass es Informatik-Methoden gibt, die darauf warten, in Feld der Ökobilanzierung angewendet zu werden.

Die Entwicklung des semantischen Mappings auf hohem Niveau bringt uns der Vision eines weltweiten Datenpools von Ökobilanzdatensätzen näher.

6 Literatur

[1] Aufsatz: Marc Ehrig, Rudi Studer: Wissensvernetzung durch Ontologien in: Semantic Web Wege zur vernetzten Wissensgesellschaft, 1. Auflage, Springer, Berlin 2006

[2] W3C Recommondation: OWL Web Ontology Language – Overview.
2004, http://www.w3.org/TR/owl-features/, Abruf am 19.11.2007

Aufsatz: Daniel Schober: Ontologien in den Biowissenschaften.
http://www.bioinf.mdc-berlin.de/~schober/bio-ontologien.htm, Abruf am 22.11.2007