The Contribution of Open Frameworks to Life Cycle Assessment

by

Bianca Sayan

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Environmental Studies

in

Environment and Resource Studies

Waterloo, Ontario, Canada, 2011

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

ABSTRACT

Environmental metrics play a significant role in behavioural change, policy formation, education, and industrial decision-making. Life Cycle Assessment (LCA) is a powerful framework for providing information on environmental impacts, but LCA data is under-utilized, difficult to access, and difficult to understand. Some of the issues that are required to be resolved to increase relevancy and use of LCA are accessibility, validation, reporting and publication, and transparency.

This thesis proposes that many of these issues can be resolved through the application of open frameworks for LCA software and data. The open source software (OSS), open data, open access, and semantic web movements advocate the transparent development of software and data, inviting all interested parties to contribute.

A survey was presented to the LCA community to gauge the community's interest and receptivity to working within open frameworks, as well as their existing concerns with LCA data. Responses indicated dissatisfaction with existing tools and some interest in open frameworks, though interest in contributing was weak. The responses also pointed out transparency, the expansion of LCA information, and feedback to be desirable areas for improvement.

Software for providing online LCA databases was developed according to open source, open data, and linked data principles and practices. The produced software incorporates features that attempt to resolve issues identified in previous literature in addition to needs defined from the survey responses. The developed software offers improvements over other databases in areas of transparency, data structure flexibility, and ability to facilitate user feedback.

The software was implemented as a proof of concept, as a test-bed for attracting data contributions from LCA practitioners, and as a tool for interested users. The implementation allows users to add LCA data, to search through LCA data, and to use data from the software in separate independent tools..

The research contributes to the LCA field by addressing barriers to improving LCA data and access, and providing a platform on which LCA database tools and data can develop efficiently, collectively, and iteratively.

ACKNOWLEDGEMENTS

I'd like to thank many people that have been invaluably supportive. I'd like to thank Steve Young for being my advisor, Goretty Dias and Geoff Lewis for providing me with so much help and guidance, Leonardo Bonanni and Matthew Hockenberry for inviting me into their project and supporting me, Jorge Zapico for being my partner on footprinted.org, Victoria Allenye for being my thesis-whisperer, Michael Fagan for consistently encouraging me, Nadia Charania and the other MERS 2011 students for being great people to learn with, and my parents (Harry Sayan and Pauline Guindon).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

Information and Communication Technologies (ICT) are providing useful tools for making progress in sustainability (Fuchs, 2006; Tomlinson, 2010). Specifically, ICT is proving to be useful in providing tools that deliver more accurate and detailed environmental metrics that play a key role in sustainability decision-making (Zapico et al, 2010). These environmental metrics play a significant role in behavioural change, policy formation, education, and industrial decision-making.

ICT have also been well-utilized to develop and present Life Cycle Assessment (LCA) studies, which are developed using the LCA framework for environmental impact assessment (Moreno et al, 2011; Ciroth et al, 2007). The LCA framework is used to estimate the environmental impacts of a product, service, or activity through resource extraction, manufacture, use, and disposal. LCA data has the potential to be very useful to multiple stakeholders because: 1) the process has a clearly defined framework with standardized guidelines, resulting in consistent application of methods and presentation of results; 2) LCA data are suitable for addressing commonly asked questions like "What is the water footprint of a PET bottle?" or "How do the environmental impacts of plastic and paper bags compare?", and 3) the LCA framework facilitates detailed data that can appeal to LCA experts and non-experts. LCA inventory and impact data has the potential to become a dominant source of sustainability information, and the use of ICT to improve upon LCA could significantly increase its relevance as valuable sustainability information and environmental metrics.

ICT not only provides the tools needed to manage information, but it has also indirectly created new ways for people to collaborate and organize. Open source software (OSS) and other related frameworks promote a project model that can improve results by requiring public software source code, publicly shared data and documents, and the linking together of data across organizations and sources. These frameworks may improve LCA by making LCA tools and data easier to develop, access, share, and improve upon. These two fields of LCA and open frameworks have prompted this study's exploration of *how open frameworks can make life-cycle assessment more relevan*t.

The LCA framework and related ICT tools are also discussed at length in Chapter 2. OSS and related tools and approaches, including OSS, free software, and linked data, are discussed in Chapter 2. The intersection of LCA and OSS will be explored through three projects:

1) a survey to establish current and potential needs for new and open frameworks in the LCA community (which is explained in Chapter 4) and

2) the proposition and development of an open source data-sharing platform for LCA data, or LCA databases (Chapter 5), and

3) the implementation of this open source application as a proof of concept (Chapter 5).

Conclusions and recommendations are discussed in Chapter 6.

2 LITERATURE REVIEW

This literature review introduces life cycle assessment (as well as LCA data) and open frameworks. It explains their use and discusses the current and potential applications of open frameworks to LCA.

2.1 Life Cycle Assessment

The advent of the modern environmental movement in the 1960s initiated the need for tools and frameworks to help analyze the multitude of product systems. The systems being explored varied widely from product packaging (von Falkenstein et al, 2010; Singh et al, 2011) to facilities (Kannan et al, 2004; Bieda, 2007) to agricultural commodities (Wang, 2007; Meisterling, 2009; Hishchier et al, 2005; Brentrup, 2004; Kim and Dale, 2005; Ometto et al, 2009). Governments, firms, and academics needed to be able to analyze systems with greater complexity, scope, and accuracy in order to understand how to make the most environmentally beneficial decisions.

Life Cycle Assessment (LCA) was one of the tools able to deal with complex environmental data. LCA began in the 1970s as an approach to assessing the full environmental impact of packaging (Klopffer, 1997). It was expanded in the 1980s to other products (Klopffer, 1997). Refinement of an LCA framework continued under the facilitation of the Society of Environmental Toxicology and Chemistry (SETAC), which culminated in the publication of a code of practice in 1993 (Klopffer, 2006). The International Organization for Standardization (ISO) then took on the role of facilitating the development of LCA guidelines/practice and is now responsible for publishing the LCA standards and facilitating revisions. LCA (including, but not limited to the ISO standards) is used as a general framework by companies, policy-makers and academics, and has a sizable academic community (including a number of journals and multiple annual conferences) devoted to its refinement and evolution.

2.1.1 Definition of Life Cycle Assessment

Life Cycle Assessment is a framework for assessing the potential environmental impacts associated with a "product system," which may be a specific product, a service, facility, or technology. The most powerful aspect of LCA is its thoroughness, both in its approach to defining and analyzing the whole system in question and also the breadth of environmental impacts that relate to the system. LCA analyzes a whole system by inventorying all the inputs, outputs, and processes through the raw material extraction to manufacturing, distribution, retail, use, and disposal stages of the system.

The LCA framework is general enough that it can be applied to a product, facility, technology, or an industry. This method is also flexible enough that it can be applied to different system models like specific case studies, regional estimates, and hypothetical scenarios.

2.1.2 Description of the ISO LCA Framework

LCA shapes studies through the general application of LCA concepts or the specific application of standards like PAS 2050 and the ISO 14040 series. The ISO published standards provide a rigorous and generally agreed-upon framework, through ISO 14040 (Environmental Management – Life Cycle Assessment – Principles and Framework), and 14044 (Environmental Management – Life Cycle Assessment – Requirements and Guidelines) standards. These will be described here and used as the model for LCA in this thesis. LCA is divided into four stages of analysis: 1) scope and goal definition, 2) inventory analysis, 3) life-cycle impact assessment, and 4) interpretation.

2.1.2.1 Goal and Scope

The purpose of the goal and scope definition stage is to describe both the purpose of undertaking the LCA and the boundaries of the system that is being analyzed. This is an essential step, because all LCA studies, even those for very similar products, can make different decisions on what to measure, how to measure it, and what sources of information are used to provide secondary data. These decisions can result in markedly different results between different LCA studies, and it is important that it is made explicitly clear what methodological factors or values contributed to that difference.

Defining the goal requires the inclusion of why the LCA is being undertaken, how it will be used, who will use it, and whether it will be used in comparative analysis (ISO 14040, 2006). Defining the scope requires a description of the system. It also requires the specification of a functional unit, which is a measure or expression of the product's performance. An example of a functional unit could be one mile driven by a mid-sized sedan on a highway using diesel fuel or the toasting of a thousand pieces of toast by a toaster. The reference flow is then the amount of the product needed to perform that function. In this case it may be a quarter of a gallon of diesel fuel or a toaster, respectively. While many LCA studies may simply use the reference flow of the product as a point of reference for the study (for example, a study could examine the environmental impact of a gallon of diesel fuel), there are a few advantages to using a functional unit as a point of reference. The functional unit helps us to associate impacts with the benefits we care about and receive from the system. For instance, fuel is not important to an economy, rather it is the transportation it provides. This helps in the decision-making process, as the discussion will address how to improve transportation, of which fuel is a factor. Additionally, there are impacts associated with the function that may not be covered by just addressing the product. For instance, examining the impact of fuel may not reveal significant impacts in the use stage of transportation (i.e., fuel combustion) that overshadow impacts in the production stage of fuel.

The system boundary must also be defined to specify which life-cycle stages, inputs, and processes will be included in the study. While it seems like system boundaries would be consistent, they may vary by focusing on segments of the life-cycle or eliminating parts of the system that are deemed minimal or negligible. An example of the former may be a study of fuel that excludes the distribution and use stages. An example of the latter may be the exclusion of the production of lubricant for machinery in an appliance factory.

The scope must also include a discussion of allocation procedures; in many cases, industrial processes and activities have co-products. Co-products are products that result from the same raw material and energy inputs and processes. An example of this is the by-product Dried Distillers Grains and Solubles (DDGS) from producing ethanol, made during ethanol production, which can be sold as feed (Kim et al., 2008). A study must make it clear whether the environmental impact of the whole system is completely attributed to the input or output that is the focus of the study, or whether part of the impact can be attributed to the other resulting products of the system. The allocation is usually decided based on the usefulness of the co-products, more specifically based on the sale value, mass, energy content, or replacement value of the co-product (Shapouri, 1995).

The scope must also outline the life-cycle impact categories, the methods for impact assessment, and subsequent interpretation to be used. While LCA can be used to describe the complete environmental impact of a system, it can and often is used to answer a question about a specific environmental impact (such as effects of agriculture on water systems) with a specific sub-set of environmental impact categories (for example, eutrophication).

The scope should also specify data requirements for the study, any and all assumptions that are made, limitations of the study, initial data quality requirements, plans for a critical review, type, and format of the report required for the study. These details are all important for assuring an accurate interpretation and contextualization of the study for readers.

2.1.2.2 Inventory Analysis

The inventory analysis stage consists of the measurement, or synthesis, of inputs and outputs for the system and the subsequent calculation of that data. This stage should begin with the preparation for data collection. It requires specific documentation of the descriptions, interrelations, and associated data categories of the unit processes in the system. The units of measurement used for all documentation and the chosen techniques to be used for measurement and calculations should be defined at this point. A procedure should be established for reporting with inconsistencies and problems.

Once preparation is completed data collection can be performed. It should take into account the possibility of double-counting and the allocation issues of multiple flows. Data validation is also an important step; data should be checked for consistency and completion according to the plan for data collection established earlier. The data must then be related to the unit process it represents and then to the chosen functional unit. A number of software packages are available to facilitate these calculations.

At this point, boundaries of the system can be limited based on what data and factors contribute more heavily to the environmental impact of the system. This may include the inclusion or exclusion of unit processes and other elements. Issues pertaining to the appropriateness of the initial scope and goal definitions, data quality, and uncertainty that arose in the data collection process should be clearly noted.

2.1.2.3 Impact Assessment

The purpose of the impact assessment stage is to understand the potential environmental impacts of the system and which impacts can be attributed to which processes and inputs. Once impact categories have been chosen the data must be classified and characterized, which means that the impacts are calculated based on the normalized inputs and outputs.

The assessment stage has optional components: normalization, grouping, and weighting. Quality analysis may also be performed at this stage. Normalization recalculates the quantitative value of an impact category into a relative value contributing to the absolute regional, national or global impacts. This can be useful depending on the purpose and goals of the LCA; for instance, if the purpose of an LCA is to explain the impact of a type of agriculture to global climate change, normalizing the impact values better explains the results of the LCA within the greater context. Grouping involves sorting impacts in order to find more general areas of significance. These groups may be also ranked by importance. Alternatively, to make results more pertinent weighting can be used; results from impact categories are modified according to specific values and concerns.

2.1.2.4 Interpretation

The purpose of the interpretation stage is to explain and present the results of the assessment. This should be done by first identifying significant issues that arise in the previous stages (Skone, 2000). The interpretation should include a reevaluation of the data and an explanation of its ability to help fulfill the study goals.

The data presented in the previous stages should also be evaluated for completeness and other issues that may compromise data reliability, as a study may encounter obstacles of missing or low-quality data not anticipated in the original plan (Skone, 2000).

The information also needs to be framed appropriately for the intended audience, as the audience may be technical or non-technical, policy-oriented, or business-oriented. The data from a study must be re-interpreted and explained so that it answers the questions for the intended audience. For instance, a LCA on agriculture in a region will require interpretation in order to help regional planners formulate water consumption policy.

2.1.2.5 ISO 14048

ISO has developed a technical specification for data documentation formats for the Life Cycle Inventory Stage entitled "Environmental Management – Life Cycle Assessment – Data Documentation Format" (ISO/TS 14048, 2002). The specification outlines the information that should be recorded during an inventory; it is intended to be used to shape the design of the LCI process and data storage. It is platform-independent and has already been used to create the foundation for data formats, as all LCA data formats (see section 2.1.4.1) have been built with the intent to be compliant with ISO 14048. While this assures that almost all LCA datasets have the same types of information included, it does not assure that they have the same structure or format. This assures consistent information for a reader, but the inconsistent formats pose complications from a software standpoint.

2.1.3 Issues in the Relevancy of LCA

With all of its usefulness as an analytical framework for environmental impacts, LCA has many methodological and implementation issues that have so far prevented it from becoming more relevant in the field (Finkbeiner, 2009; Kumar et al, 2010; Reap et al, 2008).

1. Facilitating Iterative Improvement

An LCA study is itself a process that requires continual, iterative improvement (ISO 14040, 2006), because LCA is a way of *estimating* environmental impacts, not conclusively assessing impacts. This requires the LCA process to be flexible, responsive, and open to critical review, even *during* the assessment study process. Academics may provide feedback and criticism to a study through academic journals.

While there are numerous LCA studies in the academic literature, there is little evidence of iterative improvements on LCA studies (Weidema, 2000). This is also apparent in LCA databases. Section 2.1.4.4 notes that 6 out of 40 reviewed LCA databases have some evidence of updates beyond initial dataset contributions. There is no correct way to make decisions within the LCA framework like drawing boundaries or calculating values, and, there is a great need to facilitate discussion and perform multiple LCA studies in the hope of decreasing uncertainty and improving upon primary and secondary data. LCA studies can also be performed again when any new changes or differences in the system are revealed. All of these needs require a better way to facilitate review and revision.

2. Appeal to Wider Audiences

The use of the LCA method, although broadly understood and widely referred, is fairly limited to LCA professionals and larger organizations. While the framework appeals to a wider audience, there are barriers to its widespread adoption. Undertaking an LCA is a complex, intensive, and often expensive process. Small and medium sized enterprises (SMEs), Non-Governmental Organizations (NGOs), as well as the general public could benefit greatly to increased accessibility to the LCA process. The Co-ordination Action for Innovation in Life-Cycle Analysis for Sustainability (CALCAS) project notes that the results of LCA research are increasingly being used by non-state stakeholders (Vagt et al, 2008). CALCAS also notes that there are several issues that prevent the accurate and useful interpretation of LCA research, including the proper framing of questions, the description of relevant information needs, and clear explanations for how to weight and aggregate LCA data (Vagt et al, 2008).

One of the limitations to use is the monetary and labour investment in performing a LCA study. It requires intensive data collection, expertise in the LCA process, and dedicated time for analysis. Because of this required investment, many smaller organizations cannot afford to undertake an LCA. This has severely limited the potential of LCA to assist many companies. The creators of openLCA, an open source LCA software, point out that one of the implications of making their software free is that it allows for "*opening applications where license fees are critical*" (Ciroth, 2007).

Consequently, much of the discussion and use of LCA has been targeted to a narrow audience of professionals. This is partially due to the technical understanding required to implement a LCA. Again, this limits the application of LCA. While it may be difficult to translate the LCA process into a public vernacular, it is worth exploring whether this is feasible. There have been several successful projects to expose life-cycle approaches to high school students. For instance, Powers et al. (2011) described the use of the LCA process to discuss sustainable transportation with high-school students. Sourcemap.org, an open source environmental supply-chain website, is an example of making a high-level concept like supply-chain mapping accessible to the public. It has done this by simplifying and guiding users through mapping out supply-chains (Bonanni et al, 2010). Anex and Focht (2002) also presented the case for public participation in LCA, pointing out that "*LCA is conducted, and often its results are used, through a process that excludes key interested and affected parties.*"

3. Reporting, Collection and Comparison

Approximately 25,000 LCA datasets are currently available (calculated from the review of databases in section 2.1.4.4). Tools for organizing and comparing this large and scattered body of data will become increasingly necessary. For many similar systems there are multiple published LCA studies. Ethanol production is one of these popular subject systems, with studies in the last several years now numbering in the dozens. There are no existing tools for comparing and organizing these reports.

4. Presentation and Visualization

LCA has not experienced much sophisticated development in presentation or visualization of LCA information and study results. This is not a minor issue, as the visualization of this data can better express and communicate important results to decision-makers—and many LCA studies are intended to serve decision-makers in policy, design and engineering. Weidema (2000, p 63) comments "*Long tables of figures and obscure environmental indicators are not facilitating this. New tools are needed for presenting the information from LCAs in a form readily understandable for the audience*."

5. Data Standardization

Existing ISO LCA standards facilitate consistent processes for conducting LCAs, but do not extend to standardizing the results of an LCA. This means that LCA studies often have to be read and interpreted in order to extract figures and compare studies. This could be improved upon by adhering to a standardized way of presenting and sharing LCA results. This has already been recognized and acted on by part of the LCA community culminating in the creation of the Technical Specification ISO 14048, which outlines a common data format for the LCI stage. While most LCA databases adhere to some form of this specification (see section 2.1.4.4 for how many do and do not), they still diverge by file format and data format. While many journal articles are based on the use of LCA software like SimaPro and GaBi, which are compatible with the ISO specification, publication of the data itself is not common. Sub-sets of the LCA community have created more detailed formats that provide sufficient standardization for comparison between studies and easy use in LCA software, yet these have not experienced widespread adoption either. See *Section 2.1.4.1* for a discussion of these formats and *Section 2.1.4.4* for a review of their adoption.

The lack of standardized data can be considered a subset problem of the open data movement in the academic community, of which a significant number of academics and scientists have voiced a desire for standardized, freely available data as a catalyst for accelerating scientific progress (Seringhaus, 2007).

6. Documentation and Transparency

The issue of standardization is also related to issues of raw data sharing and transparency, both concerns of the scientific Open Data movement. LCAs are rarely published with the raw data or a full inventory analysis (Murray-Rust, 2008). Transparency with secondary data is also an issue. ISO instructions are clear about the need for sourcing all information in a study [emphasis added]:

> *When data are collected from published literature, the source shall be referenced. For those data collected from literature which are significant for the conclusions of the study, the published literature which supplies details about the relevant data collection process, the time when data have been collected* ***and about further data quality indicators, shall be referenced.*** *If such data do not meet the initial data quality requirements, this shall be stated. (ISO 14044, 2006)*

While this seems arduous, sourcing reduces uncertainty and increases credibility by allowing readers to examine secondary data to make their own assessment of data quality. The importance of this is obviously recognized by many to be included in ISO standards, yet there must be some barrier that results in lackluster uptake of these recommendations.

7. Data Size and Variation

The number of LCA data sets (including those provided in LCA databases but not LCA academic papers) is currently in the thousands (with over 25,000 counted between 40 LCA databases surveyed in Section 2.1.4.4), describing a multitude of projects, technologies, services, products, and systems. Often, multiple LCA studies address similar systems, creating many estimations of the environmental impact of a similar system. For instance, there are several LCAs that address the environmental impact of US-produced corn ethanol (Kim and Dale, 2005; Liska, 2009; Nielsen, 2005). This is the case for many reasons: 1) results can be general or based on more case-specific inputs based on facility or batch, geography, and time; and 2) results are also the result of subjective input by study contributors, who make decisions on the scope, model choices, impact categories, and other essential factors (Anex and Focht, 2002). The enormous size and variation of LCA studies is beneficial to the field, as interested parties could find studies and data that best resembled their own system and scenario. Though this, in turn, also can create confusion and tremendous difficulty in evaluating results.

8. Validation and Verification

Validation and Verification are two very valuable concepts that do not play a strong role in LCA studies. In fact the ISO 14040 series documents were not designed as auditable guidance, like ISO 14001 (Environmental Management Systems) or ISO 14064-2 (Environmental Management – Greenhouse Gases – Part 2: Specification with guidance at the project level for quantification, monitoring and reporting of greenhouse gas emission reductions or removal enhancements). Validation is a process wherein a judgment is made on whether a model adequately represents a real system (Ciroth and Becker, 2006). ISO 14064-1 describes validation as the "*systematic,* *independent and documented process for the evaluation of an [greenhouse gas] assertion in a plan against agreed [validation] criteria*" (ISO 14064-1, 2006, p 3). Verification is the process of examining whether a project yields the results that the plan proposed (ISO 14064-2, 2006). While both of these processes play a strong role in the ISO 14064 series (which make up guidelines for greenhouse gas measurement) that standardizes processes and reporting for greenhouse gas (GHG) reduction projects, they are not included or discussed in LCA (ISO 14044, 2006). The process of conducting and reporting a GHG project bears many similarities to LCA, and the same argument for validation can be extended. (The case for verification is weaker, as most systems that are the subject of LCA studies cannot be measured.) Ciroth and Becker (2006) argue the processes that LCAs model are complicated and ever-changing, making validation very desirable. There are a couple of possible routes for incorporating validation into the LCA process. Young (paper forthcoming) notes the similarities between the ISO standards for LCA (ISO 14040) and the ISO standards for carbon footprinting, and suggests that the validation standards for carbon footprinting (ISO 14064-2), which are necessary to support a valid carbon market, could be adapted for LCA studies. PAS 2050, a standard for life-cycle assessment of greenhouse gas (GHG) emissions, also prescribes an auditing process. Given its incorporation of an LCA approach, it could also be a source for adaptation in general, non-GHG LCA studies. Whichever way it is implemented, there must be a mechanism by which validation can easily be implemented. It would require a modification to LCA data structures as well as a tool that could facilitate feedback in the form of validation. Current LCA data formats and databases are too inflexible to easily incorporate validation into the process of sharing and commenting on LCA studies.

2.1.4 LCA Tools

The LCA field has already yielded many computer-based tools to assist in building, organizing, and reporting LCAs. To understand how an open approach may be useful to the LCA field, it is constructive to examine how technology is currently used (both effectively and ineffectively) in the LCA field. Section 2.1.4 reviews available tools and the functions they do and do not provide.

2.1.4.1 LCA Data Formats

Many data formats have arisen out of a need for LCA data standardization. Most of these formats are compliant with ISO 14048, which means that they all cover the prescribed areas of information in the standard. ISO 14048 could also be interpreted loosely as a suggested format structure, with a main division of data into process description, administrative information, and modeling/validation. All of the following formats have been developed to be ISO 14048 compliant (which means they cover all of the information that ISO 14048 recommends), but none match the ISO 14048 recommended structure, rather rearranging and expanding on the information.

It is important to discuss these formats for a couple of reasons. The first is that the format can limit or facilitate better data quality and transparency. Users are less able to provide detailed and transparent data unless they use a format that allows for deeper description of data and meta-data. The second is that the way LCA data is stored affects how it is used. If data is in an easily-readable, common format users will experience less barriers to reading and using data. The four dominant formats are discussed below.

Ecospold is a data exchange format improved upon by the ecoinvent Centre (previously called the Swiss Centre for Life Cycle Inventories), which is a multi-institution organization that is also responsible for the ecoinvent LCI database. Ecospold has been through multiple iterations over the past fourteen years. The most current version is Ecospold v2. The project started with Spold 97 in 1997 (created by SETAC's Society for the Promotion of Lifecycle Development), then revised to create Spold 99 in 1999. Ecospold v1 was created subsequently to incorporate recommendations from ISO/TS 14048. Most recently, Ecospold v2 was created to incorporate enhanced data features like more complex ways of storing and presenting data, including Universally Unique Identifiers (UUIDs), images, and formulas. Ecospold v2 also included additional data fields to allow for quality indicators, multiple languages, and flexibility for combination or difference in model type. Consideration for some compatibility with ILCD (described below) was also considered in its design.

Spine@ISO14048 is a data format. Specifically, it describes a set of relational databases that can be used to store LCA data. There have been no significant revisions to this format, so it has not adopted some of the technological enhancements that other formats have.

ELCD is the European Life Cycle Database data exchange format. The acronym ELCD is used as well for the European database that uses this format. ELCD began in 2005 and has recently been revised and renamed as ILCD. The revision had many addressed issues in common with the revision of Ecospold (between versions 1 and 2), including support for enhanced media like charts and photos, multiple language support, reduced redundancy, and the use of UUID. ILCD also addressed support for review documentation, greater compatibility with other types of LCA studies, and better quality indicators.

Earthster (a US consultancy on supply chains and LCA) has developed a pilot linked data ontology called ECO, which is a description of how data should be stored in the semantic web. ECO is Ecospold compatible, meaning that it should be able to retain all data in an Ecospold v2 dataset. As with all linked data ontologies, it has a free use license.

2.1.4.2 File and Data Types

While a file format will specify how information is organized, a file type is the way in which information is encoded so that it can be recognized by computers and read by various software applications. The selection of a file type is usually based on the type of data in question and its intended use. For instance, storing a digital photo in any format other than those that encode graphics will not work. A digital image has many appropriate formats such as PNG, JPEG, and GIF. However, LCA data can be represented by many data types: proprietary formats (like SimaPro), spreadsheet, text, web page, marked-up text. Existing LCA datasets use all of these data types to represent LCA data. For each of these data types, there are multiple file types from which to choose. If the decision is made to communicate LCA data as text, for instance, the file type can be txt, rich text format (rtf), doc (Word Document format), or PDF. Table 2-1 explains the data types and file types that are currently used to represent and store data from LCA studies.

Table 2-1 Data Types and File Types Used for LCA Data

	File Types	Description	Queryable	Proprietary
Text				
	txt	simple text format. No formatting Permitted	No	No
	Rich Text Format (RTF)	simple format. Limited formatting	No	No
	Microsoft Document Format (doc/docx)	allows for conplex documents with tables, charts, and pictures	No	Yes
	Portable Document Format (pdf)	complex document that retains exact formatting and layout across all computer platforms and programs	No	No
Database				
	Sima-pro database		Yes	Yes

format (nx1)			
Structured Query Language (sql)	Can be used to manage and store relational databases	Yes	No
Data Markup			
Extensible Markup Language (xml)	format that uses markup to structure and identify data in the file	Yes	No
Linked/Semantic Web Data			
Triple-N (n3)	format that structures and identifies data in a file using subject, predicate, object form	Yes	No
Resource Description Framework (rdf)	format that uses markup and semantic web principles to identify data	Yes	No
Turtle (ttl)	format that structures and identifies data in a file using subject, predicate, object form	Yes	No
Spreadsheet			
Microsoft Excel Spreadsheet (xls/xlsx)	Microsoft Excel spreadsheet file-type	No	Yes

The choices in both data type and file type have a number of characteristics that result in trade-offs. The use of a proprietary data type may allow software developers to add more functionality to a software application, but it will limit the use (viewing and editing) of the data within that file to the software. The use of spreadsheet files allow for flexibility of data organization at the expense of compatibility with any other LCA dataset or software. The presentation of data on a web page allows for maximum access for everyone, however storing data in a web page eliminates the possibility of importing or converting the data, as a web page only presents data visually and does not retain any of the structure of the data. Database-specific file-types are often developed for a particular piece of software, whereas xml files are a universally recognized file type, simple to work with, and flexible.

Another important characteristic is design for human-readability versus machine-readability. When information is presented with human-readability in mind, the data type and file type is chosen that will help explain a LCA study to a human in a natural language like English. When a human agent opens a text file, they will understand the information found inside. However, it would not be feasible to write a software program that could open a text file and detect the subject of the file, the components contained in it, or the significance of the quantitative figures inside.

Often, an LCA format uses a particular file type. For instance, Ecospold files are stored as a XML file type. ECO can be represented in any of Notation 3 (N3), Terse RDF Triple Language (TTL), and Resource Description Framework (RDF). However, SPINE@ISO14048 has not chosen a particular file format, rather presenting a proscriptive structure that is more generalizable to many file types. It specifies that a dataset should be stored in a relational database and what that database should look like. It does not prescribe or require a particular kind of database. Databases are a good way of storing data but not a good way of sharing data. Data in a database is stored and accessed using database software. This data must be queried from the database and processed into a format in order for others to see and access datasets, which *can* limit sharing and accessibility. ECO is also a more generalized format; it is a linked data ontology, but does not prescribe sharing the linked data in any of the multiple file formats that can be used to store and share linked data (n3, ttl, rdf).

2.1.4.3 Software

In order to assist in the LCA process, a handful of proprietary and open software programs have been developed. They are all almost wholly proprietary and range upwards of thousands of dollars for licenses. The field consists of widely applicable software products like SimaPro (Pre Consultants, 2006) and GaBi (Spatari et al, 2001), but also includes industry or area specific software like the GREET model, which focuses on transportation (Wang, 1999). OpenLCA is one of the first open source software solutions for the LCA field (Ciroth, 2007). Several reviews on LCA software already exist (see for example Jönbrink et al, 2000), and a review of LCA software in this thesis would be redundant and would deviate from the focus on LCA data itself.

2.1.4.4 Databases

Several information databases have been compiled to provide secondary data for studies. LCA studies often require the use of secondary data (which is data that the developers of LCA studies use to estimate their impacts) to help quantify what cannot reasonably or will not be measured in a study. Since the collection of LCA data is costly, time-intensive, and dependent on access to proprietary datasets, databases are not always free and have internal restrictions on viewing metadata and other details. Hisheir et al. (2005) note that existing databases have had to retain strong levels of privacy by hiding the process details because industries (e.g., in the case the chemical industry) are reluctant to share too much detail. However, they also note that the collection and presentation of data on *basic* chemicals presents a very viable opportunity; this could be extended to other generic products.

The field has produced many disparate databases, often divided by geographic boundaries and industry. These databases can provide the life cycle inventory, impact indicators, or both. Despite (or perhaps because of) multiple efforts to establish data guidelines, standardized formats and preferred file types for LCI/LCA data, these databases have many differences that make it difficult to navigate the complete "sea" of LCA data, such as different data formats, file types, and licensing agreements that affects access and uptake. This section (2.2.4) consists of a review of 40 LCI/LCIA databases, used to discuss the current state of LCA data. Table 2-2 is a list of the databases that were reviewed (the full collection of collected metadata is provided in Appendix A). This review extends the reviews of Curran and Notten (2006) and the European

Commission Joint Research Centre list of databases (EU-JRC, 2010) to discuss characteristics of accessibility and transparency.

Table 2-2 Databases Reviewed

Name	URL
Australian Life Cycle Inventory Database Initiative	http://www.auslci.com.au/
Australian Life Cycle Inventory Data Project	http://www.cfd.rmit.edu.au/programs/life_cycle_assessment/life_cycle_inventory
BUWAL 250	http://www.umwelt-schweiz.ch/buwal/eng
Canadian Raw Materials Database	http://crmd.uwaterloo.ca
Ecoinvent	http://www.ecoinvent.ch
EDIP	http://www.lcacenter.dk
Franklin US LCI	http://www.pre.nl
German Network on Life Cycle Inventory Data	http://www.lci-network.de
Japan National LCA Project	http://www.jemai.org.jp/lcaforum/index.cfm
Korean LCI	http://www.edp.or.kr/lcidb/english/lcidb/lcidb_intro.asp
LCA Food	http://www.lcafood.dk
Spine@CPM	http://www.globalspine.com
Swiss Agricultural Life Cycle Assessment Database (SALCA)	http://www.reckenholz.ch/doc/en/forsch/control/bilanz/bilanz.html
Thailand LCI Database Project	http://www.thailcidatabase.net/
US LCI Database Project	http://www.nrel.gov/lci
CPM LCA Database	http://cpmdatabase.cpm.chalmers.se/
DEAM™	http://www.ecobilan.com/uk_deam.php
DEAM™ Impact	http://www.ecobilan.com/uk_team05.php
EIME V11	http://www.codde.Fr
esu-services database v1	http://www.esu-services.ch/inventories.htm
GaBi databases 2006	http://www.gabi-software.com/
KCL EcoData	http://www.kcl.fi/eco
Option data pack	http://www.jemai.or.jp/CACHE/lca_details_lcaobj6.cfm
PlasticsEurope Eco-profiles	http://www.plasticseurope.org/content/default.asp?PageID=392
ProBas	http://www.probas.umweltbundesamt.de
SALCA	http://www.agroscope.admin.ch/
SimaPro database	http://www.pre.nl/simapro/inventory_databases.htm
LEGEP	http://www.legep.de
The Boustead Model	http://www.boustead-consulting.co.uk
Waste Technologies Data Centre	[Under Construction]
IDEMAT	http://www.idemat.nl/
Inventory of Carbon & Energy (ICE)	http://people.bath.ac.uk/cj219/
Foodprint	http://foodprint.awardspace.com/foodprintmethods.pdf
Okala	http://www.sustainableminds.com/product/methodology
USDA National Agricultural Library Digital Commons	http://riley.nal.usda.gov/nal_display/index.php?info_center=8&tax_level=1&tax_subject=757
European Aluminium Association	http://www.eaa.net/
Deutsches Kupferinstitut/ European Copper Institute	http://www.kupfer-institut.de/lifecycle/
European Federation of Corrugated Board Manufacturers (FEFCO)	http://www.fefco.org/
International Iron and Steel Institute (IISI)	http://www.worldsteel.org/
Nickel Institute	http://www.nickelinstitute.org/index.cfm/ci_id/11.htm

One of the major issues of access is the cost to access the data. Some databases are free to use, some are strictly for use with purchased software, and some datasets may be purchased independently. Of the databases reviewed, twenty-two are free, seven are made available with a fee, two databases have tiered availability, and eight are included as part of purchased software. Free databases are dominant. Given the effort required to develop a database, it would be useful to know which type of organizations are willing and interested in this investment. Table 2-3 shows the relationship between type of facilitator of the database and the monetary condition of access. Free databases are mostly undertaken by government and academic groups. Consulting groups do not develop free databases, rather they predominantly develop databases for software. Associations mostly offer free databases. While it is known that corporations do perform LCA studies on their own products and services, it does not appear that any has published their results in their own database, though they often cooperate in the formulation of datasets for aggregated databases. The reasons behind this are an area for future inquiry.

Table 2-3 Database Survey – Relationship Between Access and Facilitator Type

Cost	Government	Academic	Consulting	Association
Free	6	7	0	3
Fee	1	1	3	1
with Software	1	0	7	0
Tiered	0	0	1	0

Another accessibility issue is the lack of clear establishment of rights, license, attribution, and terms of use. Most of the databases (36 of the 40) reviewed did not establish a statement conspicuously giving potential users clear instructions on accepted use. This is important because their absence leaves users to make assumptions for themselves. The very purpose for the existence and availability of these databases are for use by others. However, the conditions of use have not been made clear. The data may be incorporated into other academic studies or possibly be modified and reissued. The LCI database offered by the German Network on Life Cycle Inventory Data clearly states the condition of use: LCI data may be used or modified as long as a reference back to the database is included. They go so far to say that it is acceptable to pass the data on to third parties, provided that they also give reference. The conditions of use for ESU Services Ltd. are quite different; they state that the data are for use in internal studies and distribution of any kind is not allowed without express permission. Both organizations are clear about their conditions and conditions are quite different, which indicates the difficulty faced by users when no terms of use are given; users may incur the ire of organizations if they make the wrong assumption about terms of use, and organizations can lose control of their data with no legal recourse.

LCI, LCA without the LCIA component, datasets are used in other LCA studies or environmental analyses. These databases allow LCA practitioners to conduct LCA studies quickly because they can benefit from existing figures for component materials. It is expected that users will apply impact assessment methods of their own choice to the whole system. There are also LCIA databases that provide adjusted indicators. Some databases provide both. Table 2-4 shows the distribution between LCI, LCIA, and LCI/LCIA databases.

Table 2-4 Database Survey - Information Type

LCI/LCIA		**Percentage**	**Count**
LCI	|||||||||||	28%	11
LCIA	|||||	13%	5
Both	|||||||||	23%	9
Total (Databases that showed an inventory and/or IA values)	||||||||||||||||||||||||||||	70%	28

While the data in most databases are screened for quality, there still may be a need to update. Errors may be discovered after initial publishing, improvements may be made, or changes in the system (like a significantly different electricity grid) may be incorporated. This iterative approach is an accepted and important part of the LCA process. However, most LCA databases do not update data in an iterative manner. Of the 40 databases reviewed, only 6 showed any sign of iterative improvement.

While, compliance with ISO 14048 will not assure that databases have similar data formats or files, it does assure that LCA datasets are described similarly across different databases. Out of the 40 databases reviewed, at least 8 stated or implied compliance with ISO 14048 standards and at least 14 did not take into consideration ISO 14048 in the structure of their data.

Database formats include Ecospold, ILCD, proprietary formats, and custom formats. The distribution among the databases reviewed is shown in Table 2-5. Note that, because some of the databases provide LCA data in multiple formats, the total reflects the number of databases studied with multiple responses in the categories above. Many databases appear to be complying with ISO 14048, but only eight databases are using standardized formats to represent the data. This means that most databases are using their own interpretation of ISO 14048. While the software OpenLCA allows for the conversion between Ecospold and ILCD, publishing in a custom format makes conversion and comparison at best prohibitively time consuming and at worst impossible. As well, database files utilized by proprietary LCA software like SimaPro are not intended for free distribution and are not comparable outside the software environment. It is interesting that, with the effort that has gone into the development of Ecospold and ILCD, most databases are electing not to adopt those formats.

Table 2-5 Database Survey - Use of LCA-specific Data Formats

Formats		Percentage	Count
Simapro	||	5%	2
Ecospold	|||||	13%	5
SPINE	||	5%	2
ILCD	|	3%	1
Total	||||||||	20%	8

These databases also have different approaches to LCA. Even databases that have implemented a standardized ISO 14048-compliant format vary on other data included on top of ISO 14048 data or different labels. For instance, the CPM database has included quality indicators and a label that denotes the scope of the LCA (cradle-to-grave, well-to-wheels, cradle-to-cradle). Other formats may or may not have this information, or may have it in a different field with different labels. Some have added attributes like production or materials costs, or whole data segments like allocation information. These data are almost always useful and relevant, yet the differences in amounts of information lead to difficulty in designing software tools that can accept varying data structures.

Databases are using a wide variety of file types. These include proprietary database types, Extensible Markup Language (xml), Hypertext Markup Language (HTML), Portable Document Format (PDF), and the Microsoft Excel proprietary file-type. Table 2-6 shows the frequency of use of file types to represent LCA studies. Note that the total reflects the number of databases that disclosed file types, with multiple types for some of the databases included. There does not appear to be a strong trend toward using a particular file type as can be seen in Table 2-6. The implications of file type on usability have already been discussed in *Section 2.1.4.2*.

Table 2-6 Database Survey – Frequency of Use of File Types

File Types	Percentage	Count
nx1	5%	2
Xls	20%	8
Pdf	15%	6
Html	15%	6
Xml	10%	4
Total	45%	18

Language availability does play a role in access. Table 2-7 Language Availability of LCA Databases shows the languages that LCA databases use and the frequency of use. English very clearly dominates. While some databases make the data available in multiple languages (13), most do not.

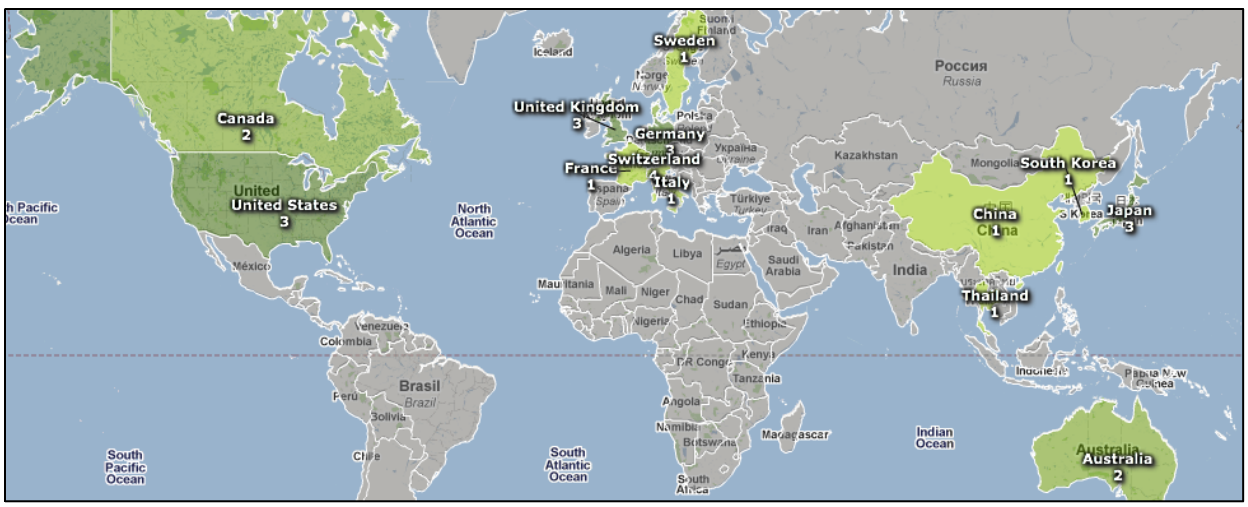
Table 2-7 Language Availability of LCA Databases

Languages	Percentage	Count
English	85%	34
German	33%	13
French	18%	7
Dutch	3%	1
Japanese	10%	4
Taiwanese	0%	0
Chinese	3%	1
Korean	3%	1
Thai	5%	2
Swedish	8%	3
Spanish	13%	5
Italian	3%	1
Afrikaans	3%	1
Catalan	3%	1
Czech	3%	1
Hungarian	3%	1
Malay	3%	1
Moldavian	3%	1
Norwegian	3%	1

Portuguese	5%	2
Punjabi	3%	1
Russian	3%	1
Slovak	3%	1
Turkish	3%	1
Twi	3%	1
Ukrainian	3%	1
Vietnamese	3%	1
Total (Databases that have clearly disclosed languages)	100%	40

The databases reviewed have varying geographic coverage, including six global databases and five Europe-wide databases. Table 2-8 shows the coverage of country-specific databases.

Table 2-8 Database Survey - Geographic Coverage



The databases were also examined for indications of an iterative process applied, for example, in the form of updates, revisions, and the reissue of datasets. Six databases were identified as having these attributes. Three databases were identified as being open for submissions, with clear invitations to submit data to the database.

The review of global LCA databases reveals a few concerns. The disparate LCA databases offer a wealth of data. However, many of these datasets cannot be used with more than one or two specific software programs. Some cannot easily be used with any software program. Due to the individual, siloed approaches of presenting each database, it is also difficult to use multiple databases for purposes that do not involve the use of common LCA software.

It appears that most of the databases are created and presented with consideration for one or two narrow goals. For instance, the Boustead database was developed for the Boustead model and likewise the DEAM databases were developed for the TEAM model. This approach, development for a particular narrow purpose, may be directing a design that does not contribute to the wider field as much as it could. The data within the databases is valuable to many, and it would require little effort to alter the databases to make the data more accessible to the wider LCA community. This could take the form of easy conversion to a common LCA data storage format. Moreover, the use of an individual database approach may lead to obscurity and inattention if the original use of the database falls out of favour. On the other hand, if a database serves a greater community and number of purposes, there will be greater interest in updating it and maintaining its utility.

These independent, siloed projects are also leading to siloed tools. Most databases have not borrowed or made use of the code and presentation tools of other databases. While there are already dozens of existing libraries or applications written to supply LCA data, one for each of the existing databases currently online, none of the responsible organizations have actively made any form of software application or library available for others. These useful tools that parse, organization, search through, and visualize data will only be useful for the database that they were developed for.

The issue of additions to common LCA information meets with the same problem. Beyond the standard components like flows, metadata, and impact assessments are additional modules like quality assessment or physical properties of the materials and chemicals involved. The CPM database has an additional section describing quality that many other databases may have benefited from. However, benefits to one database remain with that database. There is little in the way of tools, documentation, plugins, or any kind of help that would make these improvements modular and easier to incorporate into other databases.

It is also useful to note that of the 40 databases reviewed, at least 4 have been discontinued, resulting in loss of data, loss of currency and/or absence of support. While this may be for good reasons (data obsolescence) the loss of the data, documentation, and technical tools is not ideal, especially if the reasons for the abandonment have more to do with reaching the end of funding for a project. Addressing some of the problems in sustaining the results of these database projects would benefit the field.

There are several gaps in LCA that are preventing the field from being as relevant as it could be, meaning greater use in a more widespread set of interested parties. The field yields detailed data on the inventory and environmental impacts, yet it is difficult to access, read, and use. This prompts the question of whether there are alternative frameworks that can provide solutions to these gaps.

2.2 Open and Linked Development Frameworks

In the 1970s an unnamed movement in software development arose as an alternative to proprietary development. This movement was concerned with the problems caused by private control over software: users of software could not fix software bugs, they could not adapt or enhance software, and they could not provide their own support when software (and corresponding support) was discontinued. This begat a group of software developers that believed that it was imperative to assure access to the source code of software programs, allowing for collaborative improvements on the code by users. This spawned multiple movements with differing motivations and demands for transparency and freedom.

The distinction has been made repeatedly about the use of the term “free”. Most use the comparison of “free-as-in-speech” not “free-as-in-beer” (or *gratis* versus *libre*) to explain that “free” pertains to the absence of license restrictions on accessing, sharing, modifying, and distributing software. It does not pertain to being able to use software without paying. Software that is available for no charge is called “freeware”, which does not relate to or include source code and protective copyrights. This is, confusingly-enough, unrelated to “Free Software”, the concept discussed in section 2.2.2.

The intent of this section is to introduce the dominant concepts and philosophies in open frameworks and the motivations behind their creation and use. The chapter covers the concepts of open source software (2.2.1), free software (2.2.2), open data (2.2.3), open access (2.2.4), and linked data (2.2.5).

2.2.1 Open Source Software

In 1998 a more formal, organized branch of the non-proprietary movement christened itself "Open Source Software" (OSS). It is a development framework in software development. Unlike advocates for "Free Software" (See section 2.2.2), which are driven by moral and social imperatives, the OSS movement is based on economic benefits and optimizing code. It advocates making software code transparent, accessible, and available for adaptation/modification.

The OSS movement contends that open source software development is a superior development model resulting in better code and better programs. Software developed in an open source framework often selects for the "best" code and is subject to rigorous peer editing (Fogel, 2009). This means that code that is released is considered "good" (organized and efficient) code. In contrast, proprietary code will only be developed and edited by a few programmers, so the quality of the software will be limited to their collective skill, knowledge, and effort. Another benefit of software developed within an open source framework is that the community can and often does contribute bug fixes and improvements to the software, increasing software performance and improving the experience for software users.

2.2.2 Free Software

The free software movement is the name for another non-proprietary approach to software development with the same roots as the OSS movement. The branch arose out of a few key differences. It is generally associated with the philosophical leanings of Richard Stallman, who has become the public representative for the movement and was one of the first leaders in starting a community around the development of non-proprietary software (Bretthauer, 2002).

Unlike the OSS movement, which advocates for the practicality and effectiveness of open source development, the free software movement pushes for open development as a philosophical and social good. Stallman believes that the open development of software creates the need and desire for participants to contribute knowledge and work back into the community, allowing the community to grow as a whole and all participants to benefit. The other schism with the OSS movement is their leniency towards allowing an overlap with proprietary software; OSS is not concerned with the inclusion of open source code into proprietary software, whereas free software is often released under a GNU General Public License (GPL) to protect it from inclusion into proprietary code.

2.2.3 Open Data

Open Data is the label given to movements focusing on the transparent release of data. A focus on data privacy and protection has been common over the past several decades. Academia and governments, however, have experienced recent and separate movements pushing for the free and transparent disclosure of data. These movements coincided with the advent of communications technology that made data sharing low-cost and low-effort.

The push for open data in academia focuses on the release of the raw research data behind published papers and research projects. Often these papers and projects publish methodologies and summaries. It is becoming more common for some journals to allow the inclusion of supplementary information, but there is still no precedent for releasing raw data through the journal or on a university website. It is difficult for others to verify results or duplicate an experiment without seeing details of the process and data collection. The reluctance to share can stem from the worry that the scrutiny of raw data may reveal weaknesses in study conclusions or outright mistakes. There may also be a mis-trust that the people outside of the academic community may manipulate data in misleading ways (Miller, 2007). However, the argument of open data proponents, especially in the scientific community, is that open data allows for the more rapid progress of knowledge, tools, and results. Researchers can more easily build on the contributions of others. The push is still in its infancy and, while there are some open data academic projects [for example, OpenTox (OpenTox.org, 2011)], there has, as of yet, not been much adoption.

The push for open government data has come from multiple stakeholders for multiple reasons. Governments are a veritable wealth of information, from census data, to geographical surveys, to urban planning data, to wildlife sampling. The main argument for open government data is that it should be made public because it is public data, paid for by citizens for citizens. There are many benefits to be had from this availability. Accountability is one; citizens and the media are able to perform their own analyses about explanatory or illustrative data. Another more novel argument, highlighted by Clive Thompson in Wired Magazine, is that open government data can feed and grow businesses based on turning the data into valuable services (Thompson, 2011). He cites the example of Brightscope, a company that arose from the release of US government data by offering comparisons of 401K performance between portfolios. Open data can also result in the production of public goods. Open data releases by the UK government almost immediately resulted in the development of online maps and tools by the technically-skilled public, including maps pin-pointing hotspots for bike-related injuries (Berners-Lee, 2009).

The advent of the open government data movement came out of efforts by the public to enforce more transparency. One paper cites early projects like TheyWorkForYou.org and

PublicWhip.co.uk, both UK civil engagement projects, and Govtracker.us, an American project, as some of the first tools using government data. At the time of these projects (2004 – 2006) government data was not readily made available and there were no clear terms of use (Hogge, 2010). Datasets was often difficult to find, locked in formats that made it difficult to use, and was not expressed ways that were friendly and accessible to the public. Subsequent projects focused on addressing these issues, parsing datasets and visualizing them in useful ways (Hogge, 2010). Public pressure and new organizations such as the Open Knowledge foundation began to put pressure on various governments to put effort into organizing and releasing their data. Interestingly, the focus was on the release of data rather than the development of better services based on the data. This is because the release of this data was a relatively easy request to make of governments, who only needed to gather and release it. In 2009 the US federal government created a website to serve as a data portal, called data.gov. The UK and Canadian federal governments have since followed suit, as well as the Canadian province of British Columbia and the cities of Toronto, Vancouver, New York City Washington D.C., Boston, and others. These portals have very quickly sparked creation of both public and private goods and services. It is worth noting that, in that case, the release in open data resulted in completely-unprompted voluntary contributions to the social good. Whether the UK government had the will and resources to create a map for reported bike-route injuries is less relevant because they allowed the public to help participate in creating social good. This is also the argument for academia. Within academia, though, opening study data is often framed as inviting the opportunity to be “scooped” rather than thinking of it as a more efficient and collaborative way of doing research and achieving results (Miller, 2007).

2.2.4 Open Access

The open access (OA) movement pertains to the "free-as-in-beer" as well as the "free-as-in-speech" access to literature and publications. This access would mean that published papers would not sit behind a digital pay-wall (used by many existing publishers like JSTOR and Springerlink). They would be directly accessible and would include a license that allows for the use of publications for scholarly purposes (Suber, 2008). This has been discussed at length in the scientific community as both desirable (improving access to research) and corrosive (changing existing academic processes) to academic progress (Miller, 2007). The argument of proponents is that scientific progress is accelerated by the free (as in gratis as well as in libre) academic publications.

One of the greater pushes in OA has come from government and foundation funders of research, many of which have adopted OA mandates since 2005 to assure research associated with the funding was mandated to be Open Access, including research coming out of the US National Institutes of Health, the Howard Hughes Foundation, and the UK Medical Research Council (Suber, 2008). The push within academia to adopt OA mandates has resulted in efforts by Universities to self-publish journals as OA. The Directory of Open Access Journals (DOAJ) has a list of 121 journals (as of July 20, 2011).

2.2.5 Semantic Web/Linked Data

The "A lot of information is merely *On the We*b when it would be more useful *In the Web*."

- Danny Ayers

The semantic web is the name for a different approach to digital information in the web. While people can understand digital information by reading paragraphs of text, it is difficult to design software that can derive some meaning from text and understand how concepts relate. The semantic web stores digital information in pieces and gives them meaning by denoting their relationships to other pieces of digital information.

The term “linked data” was initially coined by Tim Berners-Lee, one of the inventors of the World Wide Web, to describe his proposition on how to create a semantic web by storing, sharing and linking data. The goal of linked data is to make data open, accessible by the public and connected to other data. This can be done by storing information and data in a semantic, machine-readable form. Berners-Lee proposed four key concepts required to do this: the use of Uniform Resource Identifiers (URIs), the use of HTTP URIs, the generous use of metadata, and the inclusion of links to other related URIs. URIs are used to universally identify something (whether it is a concept, a geological feature, a person, a book, or a product). The use of HTTP URIs allow for public online access to the data. Figure 2-1 Explanation of a Linked Data Triple shows one form that linked data may take in order to show how URIs are used to describe something. The inclusion of metadata informs search engines of the contents and context of the data. The inclusion of links functions in multiple ways. It allows people to assign relationships between data. It also eliminates a considerable amount in redundancy by eliminating the need to repeatedly redefine the same things, concepts, and people.

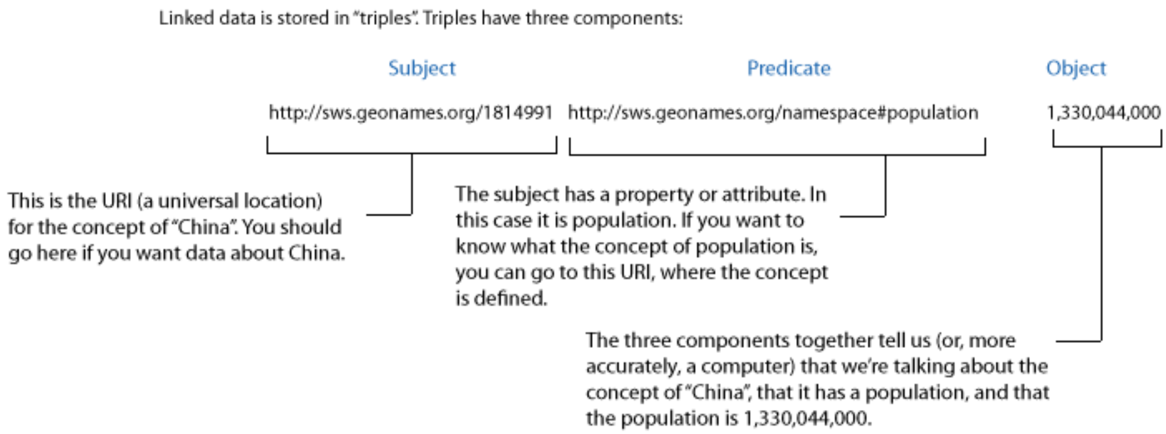


Figure 2-1 Explanation of a Linked Data Triple

Representing data as linked data is a possible solution to the existing limitations of digital data; Existing digital forms of data can only be searched by humans. For instance, a Wikipedia document on "China" may contain data on the country's population, but a human must scan the whole document to extract that information. Search engines have an extremely limited ability to do similar tasks like scanning documents for specific pieces of data and "understanding" the content. The concept of linked data requires data to be presented in a markup format. This means that information should be presented with machine-readable "grammar" which denotes the meaning of each part of the document. A document on the subject of China presented in a linked data format may contain "tags" around the population that describe the data as "type" population.

The other key concept of linked data is the connectivity between each different datum. When one defines China as a country, normally one reads the words "country" and "China" and connects them together as a concept. With linked data, instead, the URI for China is linked to the URI for "country". Each URI may contain links to other URIs. For instance, a subclass of China may refer to the URI describing communist governments. When a user queries the body of linked data for countries with communist governments, the search can semantically “understand” what is being searched for and retrieve that information from the links across multiple URIs and domains.

The semantic web is defined by ontologies. An ontology is a taxonomy or vocabulary for describing something. In linked data it provides a similar purpose as a schema, explaining the names and URIs that should be used to define data and the structure the data should take. These ontologies are meant to be universal, a machine or person can refer back to these ontologies in order to know how to look for data. For instance, the Geonames ontology provides a way of describing geography data. The ontology informs machine code that it can search for something that is a “Feature” and that it has a “name”, “alternative names” and “latitude” (which are actually URIs that look like “http://www.geonames.org/ontology#name”, “http://www.geonames.org/ontology#alternativeName”, and “http://www.w3.org/2003/01/geo/wgs84_pos#lat” respectively).

When one wants to make data available in the semantic web, they can either pick vocabulary from ontologies that already exist or create their own. They simply have to provide a linked data document that describes something with this vocabulary linking back to the ontology. A simple example is provided below using the Geonames ontology for describing the geographic feature of the city of Waterloo, ON.

```
<gn:Feature>
  <gn:name>Waterloo</gn:name>
  <gn:alternateName >The Loo</gn:alternateName>
  <wgs84_pos:lat>43.45</wgs84_pos:lat>
  <wgs84_poslong>-80.483333</wgs84_pos:long>
```

</gn:Feature>

2.2.6 Issues with the Success of Projects Using Open Frameworks

There are many issues or barriers that need to be overcome in order to implement open frameworks.

1. Attracting Collaboration/Participation

Open projects are dependent on voluntary participation; however, volunteers can be difficult to attract. Tovey (2008) points out that these projects and communities are asking for a volunteer's personal time, and must make a convincing case for that precious time to be spent contributing to a particular project. The success of the strategy of creating software from free and voluntary contributions can be seen in the robust and growing communities around open source software development like Linux and Open Office. However, there are also plenty of examples of open projects that failed to attract enough community members. Nokia’s Symbian (an operating system for mobile devices), Xara (an engine for generating vector graphics), Chandler (task management software), and Apple’s Darwin (Apple’s open source release of their operating system for Macintosh computers) project are only a few of the OSS projects that failed to build a lasting community (Nokia, 2011; Asay, 2007; Braun, 2006). Smaller open source failures cannot easily be measured because a side-effect of failure is virtual invisibility.

2. Maintaining Collaboration/Participation

Attracting volunteers to initially contribute to a community may not assure their continued participation in a community. For some projects, which may depend heavily on repeated contributions, improvements, and ongoing conversations and feedback, this continual participation is essential. This is in sharp contrast to proprietary projects, which are directly motivated by continued income.

Existing literature on sustaining participation in open source communities has pointed to the usefulness of resulting software, status, learning, personal enjoyment, ownership and control, career motivations, strong moral beliefs about free software as motivators for joining a development community (Fang and Neufeld). Fang and Neufeld (2009) note that the motivators for joining a community may not be the same as the motivators in continuing to actively participate within that community.

They identified “situated learning” (using knowledge and contributing purposefully) and “identity construction” (identifying with the open source development community) as factors that would sustain participation. Lakhani and Wolf’s (2005) survey of open source development participants found that the need for the software was a dominant motivator, followed by enjoyment and ideology. Reputation and status and competing with closed source software were not common motivators (Lakhani and Wolf, 2005).

3. Educating Collaborators/Participants

In order for contributors to contribute meaningfully they should be familiar with the processes and content relevant to the field. This can be difficult for more technical fields. However, it is essential for assuring meaningful contributions. Education material must assure that people that have the basic skill set needed to participate are easily able to do so.

4. Filtering Contributions

By opening up the development and discussion of data, an open project will likely yield more data than a similar project developed in a proprietary or inclusive environment. This can be very desirable, but it also creates the need for sorting through contributions and organizing them in a useful way.

The difficulty arises in the (1) definition of "useful" or "good" contributions and (2) the mechanism(s) for organizing them according to the value of the contributions. The definition of "useful" or "good" contributions may or may not be easily agreed-upon. Software code can be measured for efficient execution times, which provides a yardstick by which to measure contributions. Fields with more qualitative contributions require the development of more complex criteria for sorting and ranking contributions. For most fields, these criteria are yet to be determined and require careful consideration.

The mechanisms for filtering and organizing contributions can be done in multiple ways. Algorithms may be designed to automatically organize contributions according to a set of criteria. Alternatively, a committee of moderators or experts could be asked to filter contributions. Another solution harnesses the collective intelligence of community participants to vote or rank contributions in terms of subjective value.

2.3 The Intersection between LCA and Open Frameworks

The LCA field has previously had non-proprietary projects (for instance, the products from SETAC workshops), due to the need for some standardization (e.g., the results of SETAC's Society for the Promotion of Lifecycle Development, including a LCA data format), the value of collaboration over regions and industries (the ELCD/ILCD project), and the effort put in by LCA academics. However, until the development of OpenLCA in 2007, there had been no self-described "open source" efforts. More recently (in 2010) Delft University and Earthster have also developed open projects (described below).

The use of open frameworks is supported in the environmental information and LCA communities in the literature. Ciroth et al (2007) have pushed the use through their introduction of OpenLCA. Davis et al (2010) made a very detailed and convincing case for the use of linked data in industrial ecology by describing the alternate models for research available to an industrial ecologist with access to tools like a linked LCA database. The following projects have demonstrated use of open frameworks in LCA.

OpenLCA is an ongoing open source project to develop LCA software. It is a software program meant to be installed on a computer to allow users to build an LCA study. The functionality of the program is anticipated to be comparable to or competitive with software products like SimaPro and GaBi. The project started in 2007 and was the first open source project in the LCA field. Currently, it is spearheaded and guided by members of the private sector (such as GreenDelta), with support from Pre Consulting (the makers of SimaPro). The project clearly provides direction for potential contributors, and the software and source code is free. OpenLCA initially released a skeleton application with the ability to import and examine LCA studies of multiple formats, and convert between Ecospold v1, Ecospold v2, and ILCD (OpenLCA, 2010). The conversion functionality is available in a separate application. A new version was recently released in January 2011 that has the ability to create LCA data sets and build LCIA with multiple LCIA models.

Delft University is developing an open database for industrial LCA processes in the form of a wiki called Enipedia (a set of interlinked webpages that can be easily edited) (Enipedia, 2011). The wiki has begun with an inventory of elementary flows (e.g., process chemicals), with multiple processes that can refer to these flows. The wiki format offers a simple interface through which users submit information, making it easy for various to contribute.

Earthster is a private sector consulting company in the US. They have developed proprietary LCA and supply chain software. They have also developed a way of describing LCA and supply chain data in a database format. This project is called the Earthster Core Ontology (ECO). ECO is a description of how supply chain and LCA data can be described as linked data in the semantic web. Earthster also has also produced software for visualizing LCA data, but it is not open source.

Sourcemap is a web application that allows anyone to create interactive maps of supply chains of products, services, or events. These maps also have useful modules for telling multimedia stories about the supply chains and calculating various environmental impacts akin to LCA. The application is an implementation of the Open Supply Chains open source software.

2.3.1 Justification for the use of Open Frameworks to Address Gaps in LCA

The application of open frameworks to LCA has been extremely limited, unproven in effectiveness, and applied in the few cases mentioned in the previous section. There are several gaps in LCA tools that require further development of open frameworks, which are addressed in this research project.

One of the major advantages of an open project (in general) is the access to the vast numbers of potential contributors, as any interested party can contribute. As stated previously, LCA suffers from a lack of accessibility beyond a small subset of possible contributors, LCA experts. However, even though projects like OpenLCA are open source, they are still open to members of the LCA (by necessity of the education required to contribute code to a LCA software project). Opening LCA will allow for the engagement of a wider group of people that are not considered LCA experts, including programmers, environmental metrics experts, and other interested parties. This engagement could lead to a greater public understanding of and receptivity to LCA. Therefore, it could also lead to a greater number of applications of LCA methods.

The increased availability of LCA data may help to make future LCAs and other environmental reports more accurate by allowing for better choice in secondary data. Users will be able to easily search for data that is specific to their region, has similar system characteristics, or is considered more reliable. Cooper and Fava (2006) found in their survey of LCA practitioners that one of the major difficulties in conducting a LCA study was the time and resource requirements in the collection of data.

Opening up data and software offers transparency that allows them to be open for criticism, which can result in more critical feedback and changes that results in better data and software. Open projects also allow users to assess their level of trust in the results, and can result in increased (or decreased) reputation and confidence. Nevertheless, some academics and companies find opening work to criticism threatening (Miller, 2007). However, from a constructive perspective, this criticism can take the form of peer validation and verification. This project aims to demonstrate that opening data allows participants to assess the data and metadata and flag any concerns for the community, resulting in better data and better descriptions of data. A popular example of peer verification is Wikipedia.org, an open encyclopedia. Initial submissions are reviewed by other members of the public; unverified submissions are flagged, and, in the case of documented contrary information, deleted.

Opening up LCA data will give interested parties the opportunity to re-express data in new ways. Existing ways of expressing LCA data is limited and has not taken advantage of modern design principles, experiments in visualization, animations and dynamic applications. The push to experiment and modernize does not have to come from within the LCA community; as previously discussed, Berners-Lee cites (2009) many examples of external contributors to the expression of municipal and other government data. Either way, this allows the opportunity for more experimentation in data expression.

While it is useful to have a significant amount of data gathered together, it increases the difficulty of finding desirable data within the veritable flood of information. The use of semantic data frameworks makes it much easier to search through large bodies of information. Community participants can also be encouraged to qualitatively rank data quality, which can also contribute to increasing the signal-to-noise ratio. An example of this is the news website, reddit.com. While the quality of news articles are difficult to evaluate, the website has harnessed the collective opinion of community members to rank articles in order to bring the “best” to attention. This is an innovative way of harnessing the power of an open community to sort and filter through some forms of information.

While the LCA process is considered iterative, it is difficult to facilitate such a process through bi-monthly journal publications and company websites. The storage of data as linked data allows both the authors and the public to update data easily and quickly, link it to previous iterations, and compare the differences and changes.

By developing an open community, the field benefits from the collective knowledge and input of all participants. If similar models from the open source software community can be applied, LCA could also benefit from this. Users could fill in missing data or metadata, comment on and rank existing data, and provide other contributions that improve the quality and understanding of the data beyond the original submission.

Cooper and Fava (2006) also noted in their survey of LCA practitioners that 45% of their respondents participated in LCA studies with no peer review, some further commenting on acquiring peer review as an added cost. This project will help to increase feedback initially by creating the ability for members to comment on the submissions of others. This feedback can be funneled into useful validation and verification of data that can help filter through multiple data submissions for similar entries.

Data quality can be divided into two approaches. The first is the metadata, because the quality of data can be inferred from metadata. Metadata, such as the source of the data, must be complete. If metadata are incomplete, it casts doubt upon the validity of the data itself.

The use of a linked data format can cope with capacity issues that may plague a single database while still linking data together. An advantage of linked data is that the data can be hosted across many servers and websites while still being linked together as if part of a single database. The number of LCAs is expected to increase, far beyond the capacity of a single database. Additionally, companies and other organizations may prefer to host their own LCA results.

Preferences for data storage aren't easily predictable, but this strategy allows for flexibility without any trade-offs. For these reasons, this research project was designed to address the gaps of transparency, iterative development, appeal to wider audiences (using and including visualization), reporting and communication, the desire for standardization with multiple formats, and variation of LCA data.

3 RESEARCH STATEMENT

These two fields of LCA and open frameworks have prompted the research of interest here, which is ***to show how open frameworks can make life-cycle assessment more relevant to more stakeholders and more uses by providing a framework for sharing LCA data that addresses existing LCA issues.***

Relevant, in this context, means the use of LCA data for the fundamental purpose of the environmental field, which is to improve environmental sustainability. In order for LCA data to contribute to environmental sustainability, it must play a role to influence policy and decision-making. This requires increased understanding and use of LCA by stakeholders, which entails remedying existing gaps in the field. The LCA field intends to support decisions that reduce environmental impacts. Advances that improve access LCA information and/or increase comprehension and uptake of LCA information will make the field more relevant because improved comprehension and greater use of LCA data and results should lead to better decisions towards the improvement of environmental sustainability.

This research question will be explored through three projects: 1) an exploratory survey of the LCA community to establish current use and potential need for new and open frameworks; 2) the development of an online open source data-sharing platform for LCA data, or LCA database; and 3) the implementation of this open source application as a proof of concept. The rationale for including these projects is discussed in the following section.

The existing LCA issues that will be addressed are ones of increasing use to wider audiences, validation and review, visualization, sharing/reporting/publishing, and transparency (all discussed at length in section 2.1.3).

3.1 Exploratory Survey

Two components that are key to the use of open frameworks in Life Cycle Assessment are: 1) the provision of the necessary and faciliatory technology and literature and 2) the development of a participatory community. The goal of this thesis is to show how open frameworks can make LCA more relevant, which requires, first and foremost, participation from the LCA community. This means that the technology and literature must be developed with a deep understanding of the target community and how OSS communities function.

Faciliatory technology and literature are an essential part of OSS communities (Crowston, 2003). These communities must have information technology (IT) tools for communicating (e.g., a forum) and managing contributions (e.g., code and data). Literature is used to provide clear documentation of the goals of the community so that potential and current community members have a clear understanding of what they are contributing towards. This understanding is important to establishing a personal commitment that leads to participating and contributing to a community. Literature in the form of documentation of the software and data also assures that community members can educate themselves in order to use the software or write their own code.

The development of a participatory community, however, is the real goal; IT and literature are tools that assist in this goal. In any open project its utility and success lies in the continued participation of community members. Since members provide the content (e.g., software, documents, data, feedback, etc.). Key people with expertise in the field must participate in a sustained manner, otherwise there is no generation of a public good. Fang and Neufeld (2009: 10) explain this succinctly:

> *Despite the notable success stories, many more OSS projects have failed, frequently due to insufficient volunteer participation … OSS communities cannot survive or thrive without individual developer contributions. Because respondents are often self-employed freelancers and volunteers rather than traditional employees, it is impossible to rely on standard employment contracts and incentives to motivate and retain them.*

Even if a small group is convinced of the value and necessity of an open community and they construct tools and frameworks for a community, it may not assure the participation necessary to make it a success.

There are plenty of instances of failed or stalling open source and open data projects due to insufficient community participation. As mentioned before, Nokia's Symbian, Chandler, and Apple's Darwin project are only a few of the OSS projects that failed to maintain a community. There are many factors that contributed to these failures, from the partial release of source code (which significantly impedes the ability to build on top of software) (Asay, 2007) to the poor quality of the initial offering of software or source code (Asay, 2010). A commonly mentioned measurement of OSS failure are the vast number of "abandonware" projects untouched on code hosting sites like Sourceforge and Google Code.

Therefore, it is vital to understand what motivates people to participate in open communities and how they do so. Analyzing existing literature regarding successful open communities will help to understand to understand key factors that elicit sustained participation of community members, which may be transferrable to the LCA community.

3.2 Open Source Database Application

Open frameworks could contribute to many aspects of LCA, but the case for concentrating on LCA databases (including any LCI, LCIA, or results database) is strong. LCA databases have a few parallels with the condition of government data before the advent of data portals. The data are difficult to find, difficult to access, and, in some cases, difficult to parse through and use (see section 2.1.4.4 for a discussion of these issues in current LCA databases). The organizations responsible for providing the data are either disinterested in others using the data or unaware in the potential wider interest. The LCA community may not realize that hackers could be very willing to build applications and educational tools based on this data much like others have taken open government data and essentially become a provider of social goods through their efforts. There are several successful examples in other areas. The city of Portland, Oregon showcases voluntary projects based on city open data through their Civicapps.org site (Civicapps, 2011). UK hackers contributed many civic projects through a hack-day hosted by Google India (McCandless, 2010).

An open source LCA database project could be beneficial in the following ways. 1)This could help build an academic community that works together, 2) it could create a body of work and tools that evolves cumulatively by really building on previous efforts and benefits from the best contributions that the community has to offer; 3) provide an accelerated workflow due to ease of use; 4) greater recognition and interest with the public; and 5) more widely-used and applicable implementations of LCA databases.

3.3 Open Data Website

The implementation of an open data instance of the database application will serve as a proof of concept and also give the LCA community a ready platform by which to begin contributing content. It also gives an example by which the community may develop criticisms or improvements.

4 THE LCA COMMUNITY – A SURVEY

4.1 Literature Review

To determine how best to implement OSI, it was important to develop an understanding of the receptivity of stakeholders in the LCA community to sharing and linking together LCA data. The OSS community is not rooted in academic theory. There is some literature on measuring the feasibility of applying open frameworks to a community, and they have pointed to some key factors in determining the success of open projects, including: software use value, learning and personal enjoyment, recognition and reputation, personal ownership and control, career advancement opportunities, the free software ideology, and desire for social identity as motivators to participate (Fang and Neufeld, 2009).

Fang and Neufeld (2009) divide these factors between those that contribute to initial entry into the community and participation and those that contribute to sustained participation. Previous studies have pinpointed software use value as one motivator for initial entry into an OSS project. Fang and Neufeld (2009) studied the positive contributors to sustained participation and concluded that situated learning (which is active and purposeful participation in a community and project) and identity construction (the formulation of identity through participation in a community) can change respondents' initial motivations to join into a community and to continue participating. It built on the conclusions of Bagozzi and Dholakia (2006) that sustained participation related to personal investment in the community and was motivated more by personal enjoyment than need.

However, this does not necessarily provide guidance on community indicators for the success of a project that encompasses open data as well. For this we can turn to a body of academic work on "information sharing" in the field of organizational behaviour, which has established some indicators. Davenport (1995: 5) defines information sharing as a "*voluntary act of making information available to others*". This research on "information sharing" is often conducted within the context of an organization, as it is a major component in improving and disseminating knowledge assets within organizations.

Constant et al. (1994) found that **information ownership, pro-social attitudes, and organizational norms** as attitudes that encourage information sharing. They studied the use of information sharing through experiments in which respondents express their attitude as a response to a scenario.

Information ownership pertains to both the perceived and actual ownership of information. Members of an organization that believe that information belongs to the wider community or the organization they belong to are more likely to share their information than members that believe that they have personal ownership over the information.

Constant et al. (1994: p. 4) defines pro-social attitudes as, "*wish for good outcomes not only for themselves but also for other employees or for the organization more generally*." Their research on pro-social attitudes is focused on the direct interaction between two parties and their response to interactions in which they will not benefit. They asked people to consider whether they would share information with a co-worker that had previously refused to share information with them. In the case of the LCA community the focus is not on direct interactions. Rather, the interest in the LCA community is in general pro-social attitudes, like how likely it is for researchers to post their data automatically before any personal requests are made. This also reduces the need for a strong pro-social attitude to surmount previous anti-social interactions. What we would like to assess, then, is a more general pro-social attitude.

Organizational norms are shared beliefs with a community or organization on which types of behavior are usual and good and which are not (Constant et al, 1994). An organization that has established organizational norms for sharing information will cause much less friction when members participate, while organizations that do not establish norms may experience considerable, preventative friction toward the act sharing.

Tohidinia and Mosakhani (2010) gauged the receptivity to information sharing within oil companies in Iran. They concluded that self-efficacy and anticipated relationships contributed toward a positive attitude on knowledge sharing. They also concluded *"attitudes, subjective norms about knowledge sharing and perceived behavioural control had a positive effect on intention to share knowledge" (Tohidinia and Mosakhani, 2010: p. 9).* Therefore, we should measure attitudes toward and intention to share. However, applying the findings of this study may be limited as it focused on internal information sharing within a company. Information sharing within a company can be partially motivated by the belief by the employees that the greater organizational needs supersede their personal convictions, and this attitude may not necessarily extend to an academic community. Tohidinia and Mosakhani (2010) also note that their results were drawn from a collectivist society, which may not translate well to the LCA community.

These findings can be extended to the LCA community, as it is reasonable to assume that the same attitudes will affect sharing in the same ways. If these attitudes can be measured within the LCA community, the results provide an indication of a community's receptivity to sharing.

In addition to those attitudes, Jarvenpaa and Staples (2000) found that propensity to share (a personal pro-social attitude) was related to the use of media for information sharing. So it is useful to gauge whether some propensity to share already exists within the LCA community. Additionally, they studied the use of collaborative electronic media (e.g., email, websites, etc.) in an academic setting and concluded that **task interdependence, comfort with computers, and perceived information usefulness** were key to encouraging the use of collaborative media for the purpose of information sharing. They found that task interdependence, which is the potential for required tasks to be completed using results of collaborative electronic media, showed increased use of collaborative electronic media. Comfort with computers did correlate with an increased use of collaborative electronic media. However, the relevance of that finding is minimal, as this project holds little influence over individual comfort with computers beyond assuring that the project is usable and accessible. Perceived information usefulness, the extent to which someone believes that information is useful to themselves and/or others, also correlates with an increased use of collaborative media.

Constant et al. (1994) give a brief discussion regarding the fact that the personal cost of sharing is very relevant to this case. Personal cost of sharing may involve time, an intangible emotional disadvantage (such as working with uncooperative members of the community), or other resources. Given that there may be no superseding sense of community or purpose to propel someone in the LCA community to participate (as there may be for someone that is sharing information within a company), this factor is even more relevant.

Therefore, by gauging information ownership, pro-social attitudes, organizational norms, self-efficacy, anticipated relationships, and personal cost to share within the LCA community there will be a good indication of the potential for information sharing within the LCA community.

The goal of the inquiry is to understand the needs and gaps in LCA data and database tools as well as the attitudes and opinions of members of the LCA community. While there have been surveys of LCA practitioners, none have addressed these areas of inquiry (Bjorklund, 2002; Baumann, 1996).

4.2 Summary of Survey

This research was conducted in order to establish some attitudes and norms in the LCA community regarding the tools available to them. The research focus related more specifically to LCA data and sharing, access, and use of that data. The LCA community yielded 30 respondents that made up the sample for the survey. Participants answered an online survey that included open, semi-open, and closed questions focused on attitudes, opinions, norms, and awareness of tools and data. The responses were then tabulated and themed to gauge a baseline for community opinion.

4.3 Method

4.3.1 Participants

The target population consisted of members of the LCA community, which is made up of academics, consultants, public employees and researchers, and company and Non-Governmental Organization (NGO) employees that make and use LCA studies. Members of the LCA community have access to IT, as LCA studies are almost always performed with access to computer software and online information resources. In addition, the actual population is very small, numbering in the thousands.

The first step to acquiring respondents is finding a point of entry (Gillham, 2008). This can be done by finding a list of a sub-set of the target population. There are few opportunities for accessing a suitable pool of LCA experts. In order to gain access, a partnership was formed with the facilitators of the upcoming UNEP (United Nations Environment Program)/SETAC (Society for Environmental Toxicologists and Chemists) workshop on data and database standardization. While this created some bias in the pool toward respondents that were especially interested in or had some expertise with data, it also was hoped that this would improve the study due to the high level of expertise in the respondents. This approach does not attempt to assure a representative sample of the community and does rely on the convenience of a readily available number of people known to be interested in the subject, which means that it is a case of convenience sampling. Convenience sampling, in which participants are picked by their ease of availability, is often not considered a viable approach for surveys, as it is not a random sample (Gillham, 2008). However, it is recommended for inquiries that are attempting to “get a feel for the issues involved” (Robson, 2000, p14). Also, because potential respondents were experienced with LCA data (as the list consisted of LCA data and database experts), it was assumed that they would find the topic of information sharing relevant to their interests in the field. UNEP/SETAC consented to partner on a collaborative online survey. An invitation to participate was emailed out to 161 people on the UNEP/SETAC database workshop mailing list on November 23, 2010. The invitation was carefully considered to: 1) comply with the University of Waterloo's ethical research requirements; 2) be clear about the intent and goals of the survey; and 3) convince recipients to participate.

Between November 23 and January 13, 2011, 30 anonymous responses were received (18.6% response rate). This is a lower-than average success rate, as response rates for online surveys range from 24-76% (Sue & Ritter, 2007). While this number of responses is not statistically significant, it does create a foundation for discussion. Furthermore, the open-ended questions provided some insight and direction for future inquiry into the LCA community.

4.3.2 Instruments

As members of the LCA community are geographically scattered, and the community is sporadic and spans the globe, in-person inquiry methods would be impractical. Therefore, a survey was used as the method of inquiry. Surveys are also appropriate for simple, well-structured questions. Surveys can be done by mail, phone, in-person, and online. While in-person interviews can have good response rates and complex, adaptive questions can be posed, they are not ideal for a geographically dispersed population. Telephone surveys also have the benefit of being able to ask complex questions, however, they are intrusive. Mail surveys currently do not provide advantages over online surveys (they cost more to implement and require more time to receive responses), so an online survey was chosen as the method of inquiry. They work well for potential respondents that are geographically scattered (Sue and Ritter, 2007) and may require flexibility in the time and decision to participate. They are also a good choice for inquiring about straightforward factors like preferences (Gillham, 2008). Online surveys are also advantageous when the collection of responses must be done in a short window of time (Sue and Ritter, 2007). While they can limit an investigation to respondents with computer and internet access, the LCA community has that access.

These questions were developed based on previous research in organizational behavior (see Section 4.1); however, the questions themselves were constructed for this inquiry. The questions were a combination of close-ended, semi-closed-ended, and open-ended questions. Close-ended questions "*are easy to answer, familiar to respondents, and provides reliable measurements*" (Sue and Ritter, 2007: p17). Close-ended questions were used as much as possible when appropriate, because the analysis of results is easier to perform on close-ended questions than open-ended questions. Semi-closed questions allow most users to respond easily and gives the opportunity for easy analysis of results while allowing for alternative responses. As this is an initial exploration, close-ended questions (made up of multiple choice and yes-or-no questions) may limit the opportunity to gain from the expertise from respondents in understanding all issues they may be aware of. Therefore, some open-ended questions were included to allow respondents to contribute freely. It is hoped that respondents will contribute freely about their attitudes and individual and personal experience with the LCA process. Open-ended questions are useful for initial explorations because they "*allow the survey group to tell you what issues are important*" (Gillham, 2008: p29).

Many analytical methods are available to examine a survey. For this study, most open-ended questions were themed to make patterns across responses evident. Close-ended questions and themed open-ended questions were summarized using frequency distribution tables. These simple methods were chosen because more complex analytical methods are not as useful for exploratory surveys with a small number of responses such as this study. The survey questions also do not contain quantitative information, limiting the analytical tools that can usefully be applied. Demographic questions are provided to indicate the origin of respondents, and will be discussed briefly as well as used to divide other questions by demographic categories. The survey and rationale is provided in Formulation of the Survey Questionswith further explanation.

4.4 Response Analysis

The following discusses the results of the survey from the respondents and how they contribute to the research statement of this thesis. The results will also inform decisions made in the research project (described in Chapter 0).

Table 4-1 shows the themed results of gaps in existing LCA data, databases, and tools. This was presented as an open-ended question, so respondents discussed their concerns in their own words. The comments were associated with general themes and then tallied for frequency. As noted in Table 4-1 Survey Results – , most respondents were not satisfied with existing software and datasets. they mentioned a lack of data, particularly a need for more local and region-related data, newer data, and data for particular sectors. This finding identifies a need within the LCA community and a potential endeavor for future projects.

Table 4-1 Survey Results – Satisfaction with LCA Datasets and Software

Q8. Are you satisfied with existing LCA software and datasets? What do you feel are the gaps in existing software and datasets?

Themed Responses	Frequency Bars	Percentage	Frequency
Yes	||||||	23%	6
No (or implied)	|||||||	27%	7
No (discussed gaps)	|||||||	27%	7
Insufficient data	|||||||||	35%	9
Insufficient quality in existing data	|||	12%	3
Insufficient transparency in existing data	||||	15%	4
Insufficient quality in meta-data	|	4%	1
Missing functionality in software	|||	12%	3
Not User-friendly software	|||	12%	3
Not user-friendly data/esoteric	|	4%	1
answered question			**26**
skipped question			**4**

Respondents were divided between satisfaction or dissatisfaction with file formats (shown in Table 4-2). Respondents that did state dissatisfaction desired better representation of metadata (such as quality indicators), more statistical information (such as standard deviation, minimum, and maximum values for inputs or outputs), and more transparency.

Table 4-2 Survey Results of LCA file formats used- Question #9

Q9. Are you satisfied with the LCA file formats you use? What features drove your choice of data format? Do they contain sufficient indicators of data quality, detail, and meta-data? Please explain.

Themed Responses	Frequency Bars	Percentage	Frequency
Yes	||||||||	33%	8
No	|||||||||||	46%	11
answered question			**24**
skipped question			**5**

Respondents were asked to select all publishing avenues that they have used, resulting in multiple selections for some respondents. Survey results showed that the majority of the respondents publish their studies in journals (60%) as well as databases (48%) (Table 4-3). Journals and reports are well-established mediums for publishing and sharing LCA data, supported by the survey results which show strong use of journal articles and reports. Of the respondents, 48% have submitted to LCI/LCA databases, indicating a willingness on their part to share information. Coupled with the number of respondents that have submitted online supplements, there is an indication that some in the LCA community are willing to put in extra effort to share rich data.

Table 4-3 Survey Results - Publishing

Q10. Do you publish your data anywhere? If so, where? If not, why?

Responses	Frequency Bars	Percentage	Frequency
In journal articles	|||||||||||||||	**60%**	**15**
As online supplements to journal articles	|||||||	28%	7
Corporate reports	|||||||||	36%	9
In LCA/LCI databases	||||||||||||	48%	12
Other (please specify)	|||||	16%	5
answered question			**25**
skipped question			**5**

Of particular note is that 16% of respondents do not publish data at all; this group is not large, but their existence indicates a group that already has little motivation or significant barriers to publishing in common mediums (which may persist to new mediums, or may be mitigated if internalized into the approach of a new medium).

Table 4-4 describes the norms of existing data sharing. As expected, most respondents share results and LCA summaries, the smallest fractions of a full LCA that remain useful. However, 39% of respondents also do publish full unit process data, indicating some comfort for sharing full datasets.

Table 4-4 Survey Results – Availability of LCA Data in Published Studies

Q11. If you've published your LCA, in what form did you publish it? (Please select all that apply.)

Options	Frequency Bars	Percentage	Frequency
Full LCA – Unit Process		39%	**9**
LCA Summary		44%	**10**
Results		**83%**	**19**
answered question			**23**
skipped question			**7**

Table 4-5 shows how respondents store data. Unexpectedly, spreadsheets are used more than LCA specific file formats. This begs further inquiry into the functionality that spreadsheets provide over supposedly optimal LCA file formats, including perhaps a more common format among the community. Sixty percent of respondents use LCA-specific file formats, which will require further research on what the disadvantages these field-specific file formats are posing.

Table 4-5 Survey Results – Data Formats

Q12. How do you store your LCA data?

Options	Frequency Bars	Percentage	Frequency
Spreadsheet File		82%	22
LCA-specific file format		59%	16
PDF		11%	3
a text or rich-text format		19%	5
Other (please specify)			2
answered question			**27**
skipped question			**3**

Table 4-6 displays responses on data ownership, which allowed for multiple responses per respondent to indicate which types of ownership they had experience with. It was asked to establish what stakeholder owned data in previous studies, and also, more importantly, feelings of ownership toward a person's work in LCA. The most stakeholder ownership (33%) is attributed to the company. Company ownership is a practical barrier to sharing. This might be attributable to confidentiality agreements and intellectual property issues but will require further investigation. A few respondents felt that it was worthy to note that they were fine with not having ownership over the LCA data they yielded, which may also indicate a barrier or opportunity to sharing, depending on whether this could be coupled with a public license.

Table 4-6 Survey Results - Ownership

Q13. In your current and previous LCA research, who owns (owned) the resulting data? Do you feel that you have some ownership over your data?

<table>
<tr><th>Options</th><th>Frequency Bars</th><th>Percentage</th><th>Frequency</th></tr>
<tr><td>Company</td><td>|||||||||</td><td>33%</td><td>9</td></tr>
<tr><td>Government Entity</td><td>||</td><td>7%</td><td>2</td></tr>
<tr><td>No one/Public Domain</td><td>||||</td><td>15%</td><td>4</td></tr>
<tr><td>Self</td><td>|||||</td><td>19%</td><td>5</td></tr>
<tr><td>Client/Commissioner</td><td>||||</td><td>15%</td><td>4</td></tr>
<tr><td>Contractors</td><td>|</td><td>4%</td><td>1</td></tr>
<tr><td>Institute</td><td>|</td><td>4%</td><td>1</td></tr>
<tr><td>Mixed/Multiple/Depends</td><td></td><td>0%</td><td>0</td></tr>
<tr><td>answered question</td><td></td><td></td><td>27</td></tr>
<tr><td>skip question</td><td></td><td></td><td>2</td></tr>
</table>

To determine the role of review of LCA data studies, respondents were asked about current avenues for feedback and peer review. Responses reveal that 1) critiques and comments through the prescribed critical review process is a type of collaboration and communication within the community that is very relevant; 2) assessing whether better communication is perceived as a need which could point to potential use value for an information sharing system; and 3) improving the ability for community members to comment or critique data could be a potential basis for identity construction and situated learning. Some (but not all) respondents are satisfied with existing avenues for commenting and critiquing.

Table 4-7 Survey Results – Feedback on Studies

Q14. Do you receive critiques and comments on your LCA data/studies? How? Are you satisfied with existing avenues for soliciting feedback?

Themed Responses	Frequency Bars	Percentage	Frequency
Yes (Satisfied)	||||||||||||	44%	12
Yes (but not satisfied)	|||||	19%	**5**
Yes (No further comment or neutral)	|||||	19%	**5**
Never received critiques	|||	11%	3
By Email	||	7%	2
By Critical/Peer/Journal Review	|||||||	26%	7
Internal Feedback			4%	1
Published Responses			4%	1
answered question			**27**
skipped question			**3**

Table 4-8 shows that the majority of respondents do conduct LCA studies in collaboration with others, indicating collaboration is already a widespread norm. This means that the community does not needs to be convinced to collaborate and may be receptive to faciliatory tools.

Table 4-8 Survey Results - Collaboration

Q15. Do you conduct LCA studies in collaboration with others in the LCA field?

Themed Responses	Frequency Bars	Percentage	Frequency
Yes – Often	||||||||||||	**44%**	**12**
Yes – Sometimes	||||||||||||||	**52%**	**14**
No			4%	1
answered question			**27**
skipped question			**3**

Iterative development is a prescribed, but not required part of the LCA process. Table 4-9 shows that iterative development (revising and republishing LCA studies through incremental improvement) is practiced by the LCA community. Respondents appear to be divided on the practice of iterative development of LCA data with 12 positive responses and 12 negative responses, though several respondents who do not currently publish iteratively indicated intention to do so in the future.

Table 4-9 Survey Results – Iterative Development of LCA Studies

Q16. Do you revise and re-release LCA studies (akin to iterative research proscribed by ISO 14044)? Why or why not?

Themed Responses	Frequency Bars	Percentage	Frequency
Yes	|||||||||||||	46%	12
No (or implied)	|||||||||||||	46%	12
answered question			**26**
skipped question			**4**

In order to understand the most dominant barriers to sharing data, respondents were asked whether they identified with common barriers. Multiple responses were allowed in order to understand all the concerns that may influence the community. Table 4-10 shows respondent attitudes that may form barriers to sharing LCA data. These questions also help respond to the perceived cost of sharing and perceived use value. The predominant concern (in 53% of respondents) appears to be the potential mis-use of data that is made available. Concerns with the general risks of sharing data, the time commitment required, the perceived minimal benefit to the sharer of data, the possibility of being "scooped" by others, and a lack of understanding how to share data also all were selected by a few respondents. The presence of respondents that do not know how to share their data (18% of people that responded to the question) may indicate a need for friendlier tools that facilitate sharing. No respondents seemed to believe that others did not benefit from sharing data or were worried about opening their data to criticism.

Table 4-10 Survey Results – Barriers to Information Sharing

Q17. Select barriers to information sharing that applies to you (select all that apply):

Options	Frequency Bars	Percentage	Frequency
I worry about the risks of sharing my data	|||||	29%	5
I think the time-commitment for sharing data is probably too high	||||	24%	4
I worry that my data may be mis-used	|||||||||	**53%**	**9**
I don't really see a benefit to me in sharing my data	|||	18%	3
I don't really see a benefit to others and to the field if I share my data		0%	0
If I release my data, I may be scooped on future publications based on this data.	||||	24%	4
I don't want to open my data to criticism		0%	0
I don't know how to share my data	|||	18%	3
Other (please specify)	||||||||||		10
answered question			**17**
skipped question			**13**

Attitudes toward information sharing have significant implications for the success of an open data LCA project in establishing demand and anticipating participation rates. Table 4-11 shows responses to information sharing related to their LCA data. Most respondents state that they would like greater access to LCA data. The majority of respondents also agree that access to LCA data “contributes to the quality and relevancy of academic data”. However, fewer respondents stated that they were willing to share their data.

Overall, respondents see a benefit in sharing information, but are less inclined to share their own data. It is interesting to know a similar number of respondents considered their data to be valuable to others as the number that indicated a willingness to share their own data.

Table 4-11 Survey Results – Attitudes about Information Sharing

Q18. Select attitudes toward information sharing that apply to you (Select all that apply):

Options	Frequency Bars	Percentage	Frequency
I think others will benefit from having access to my my data	||||||||||||||||	59%	16
I think access to LCA data contributes to the quality and relevancy of academic data	|||||||||||||||||||||	**78%**	**21**
I would like greater access to LCA data	|||||||||||||||||||||	**78%**	**21**
I want to share my LCA data	||||||||||||||	52%	14
I would like to learn about how to share my LCA data	||||||||||	37%	10
Other (please specify)	||		2
answered question			**27**
skipped question			**3**

Respondents seem to already be sharing data predominantly (79%) as seen in Table 4-12, possibly indicating an established norm for data sharing in the LCA community.

Table 4-12 Survey Results – Data Sharing

Q19. Do you share your data with others? Have others shared their data with you? Why or why not?

Themed Responses	Frequency Bars	Percentage	Frequency
Yes	|||||||||||||||||||	79%	19
No	||||	17%	4
answered question			**24**
skipped question			**5**

Establishing awareness of open LCA projects can indicate whether lack of awareness is a barrier. Table 4-13 shows whether respondents were aware of the existing opportunities for the use of open LCA databases and tools. It appears that most respondents are aware of the existence and availability of open databases and software. While most also apparently use open databases, open source software is not widely adopted. Only one respondent claimed use of open source software. This begs further investigation into why open source software is not used, which may be partially answered in question 17. The high use of open databases is significant and reassuring, as they may be very receptive to open databases and don't seem to be inclined to use only proprietary databases (though, in hindsight, an inquiry into attitudes of proprietary databases may have made this clearer). It appears that members of the LCA community are aware of the options available to them with respect to free and open software and databases, so they are electing to use or not use them. This also suggests that they are receptive to open tools.

Table 4-13 Survey Results – Use of Free and Open LCA Tools

Q20. Check which apply:

Options	Frequency Bars	Percentage	Frequency
I use LCA databases that are free & open	|||||||||||||||||	68%	17
I use open source LCA software			4%	1
I am aware of open LCA databases	|||||||||||||||||||||	**84%**	**21**
I am aware of open source LCA software	||||||||||||||||||	72%	18
I am not aware of open LCA databases		0%	0
I am not aware of open source LCA software		0%	0
answered question			**25**
skipped question			**5**

Establishing awareness of open frameworks can indicate whether open frameworks will be an alien approach to the LCA community. Table 4-14 shows respondent awareness in the field of new technological and collaborative concepts. These are all approaches that may (or may not) become incorporated into the next generation of LCA tools and software. Recognition of these concepts among members of the field can shape receptivity to participating in online LCA communities and to using tools that are built with these in mind.

It appears that all respondents who answered the question are aware of open source, and most (23 respondents or 88%) are aware of open data. The concepts of "linked data", collective intelligence, and crowd sourcing were less well known, with only 31%, 23%, and 31% of respondents selecting the terms respectively.

Table 4-14 Survey Results – Familiarity with Open Frameworks Concepts

Q21. Which of these concepts are you familiar with (check all that apply):

Options	Frequency Bars	Percentage	Frequency
Open Source	||||||||||||||||||||||||||	**100%**	**26**
Open Data	|||||||||||||||||||||||	89%	23
Linked Data	||||||||	31%	8
Collective Intelligence	||||||	23%	6
Crowd Sourcing	||||||||	31%	8
answered question			**26**
skipped question			**4**

Table 4-15 shows respondent receptivity to the idea of open source, open data, and linked data. Overall, respondents were positive and seemed to be very receptive to the use of open frameworks within the community. Many included caveats for success and applicability, including sufficient documentation, possibility for validation, sufficient transparency, quality control, and a more structured initiative.

This is a more direct, simplistic way of assessing receptivity, but it is also important because while these concepts/movements indicate information sharing, they also carry their own meaning which may influence respondents one way or another. For instance, open source is a well-recognized movement, which may spark both recognition and interest. However, open source is also associated with an organically built community with no structure, accountability, or reputation. This is a concern, as one respondent indicated that they believed the field could benefit from these frameworks if "*they are part of national initiatives*" (Respondent #12).

Table 4-15 Survey Results – Benefits of Open Frameworks to LCA

Q22. Do you perceive benefits from the use of open source, open data, and linked data approaches to LCA?

Themed Responses	Frequency Bars	Percentage	Frequency
Yes (with or without conditions)	|||||||||||||||||||	76%	19
No	||	8%	2
Mixed			4%	1
Not Sure	||	8%	2
answered question			**25**
skipped question			**5**

Finally, respondents were posed with an open-ended question ("Are there any issues on LCA data and data quality that you would like to discuss?") to identify important issues that may not have been included in previous questions. Quality and completeness were major concerns, with 9 of the 19 respondents addressing the topic. Respondents noted that specific descriptions were lacking in existing data. One respondent said, however *"…there are many problems with how LCA data are reported and quality is often a problem, but there is little that the LCA practitioner can do because there are so few choices for data. In the U.S. we really don't have enough data to even start talking about quality…"* (Respondent #4) which indicates a different root problem than simply a lack of detail.

Interestingly, several of the issues are already incorporated into ISO LCA recommendations as well as many of the LCA-specific file formats. Temporal and geographic specific information as well as metadata were stressed as lacking, even though ISO, Ecospold, and ILCD all allow for the provision of that information. Even more generally, one respondent stated that "Standardization of structure and nomenclature" was needed, which is of concern considering the multiple continual efforts dedicated to achieving just that. Of course, this leads to the question of why these efforts, which do provide some consistent ways of structuring LCA studies, still do not seem to satisfy LCA practitioners.

Several respondents brought up the issues of transparency, noting that it was “crucial”. However, one respondent also noted that “*The question of what transparency really means has not been truly answered*”(Respondent #11), perhaps requiring a general agreement first.

Uncertainty was also a focus. One respondent mentioned that the focus should be on reducing uncertainty, not trying to create new methods to quantify uncertainty. Confidentiality was also noted. One respondent noted that “*LCA data contains business critical information ... that companies usually cannot share with the public*.” (Respondent #26) This could be an important consideration for new formats and databases that are interested in corporate adoption or participation. Of course, this is desirable as they will be the dominant generators of LCA data that is specific to a product and a process, unlike the national databases that are based on abstract commodities.

The dissatisfaction with databases confirms the need for new approaches that can help increase the number of data sets, data quality, and metadata as well as increasing transparency. Other indications, including the preference for storing data in spreadsheets over LCA-specific formats, point to a need to improve upon the tools available to LCA practitioners.

4.4.1 Summary

The analysis is summarized below to focus on key results with useful implications to this research project.

4.4.1.1 Publishing, Sharing, and Ownership

The majority of respondents reported that access to data is important and would like to have greater access to data, although fewer respondents seem interested in sharing their own data and learning how to do so. While the desire for better access could be a driver for improving access, a lack of desire to be part of the solution by contributing data could be a significant barrier. Respondents indicated that information ownership was distributed among stakeholders (see Table 4-6), creating complications in getting LCA practitioners to share their data. However, the demonstration of some pro-social attitudes and organizational norms (both characteristics important for information sharing mentioned in section 4.1) for collaborating and publishing (see Table 4-3 and Table 4-8) indicate that the community may be receptive to data sharing. Perceived value of data, another indicator, appears to be contingent on the thoroughness, transparency, and detail of LCA datasets. This supports other statements indicating that increased quality of LCA data is a priority to encourage data sharing. The personal cost to share, another indicator, does not appear to be a strong barrier to sharing (see Table 4-10 Survey Results – Barriers to Information Sharing), though misuse of data does appear to be a significant concern. This could be a big disincentive to share, but potential misuse is already a problem for academic research or any research figures, open or not. Overall, the indicators outlined in section 4.1 (**information ownership, organizational norms, propensity to share, perceived value of information, and personal cost to share**) indicate potential for information sharing.

Some of the questions in the survey can also contribute to our understanding of motivators for participation. Access to transparent and region-specific data is likely to provide use value for the community, creating an opportunity to motivate initial participation in an open community. A few respondents mentioned a desire for peer validation and the ability to comment and criticize on LCAs, creating an opportunity for identity construction and situated learning. This is in addition to the existing potential to simply contribute data to an open community, which would also contribute to both factors.

4.4.1.2 Collaborative IT

Most respondents considered there to be some benefit from incorporating open approaches into LCA. Transparency was considered as a very positive outcome of open approaches, as commented on by several respondents. However, one respondent also noted that defining transparency within the field was needed first. Respondents also noted that the quality of data may degrade or be called into question with an open approach. They also cautioned focusing on key areas and the importance of assuring better data with an effective approach. The focus should be on good data and software (no matter the approach), standards, documentation, reducing uncertainty, and validation.

Respondents also described their concerns with respect to the risks of participation. Good representation of owners and users of data would be necessary for success. A respondent noted that such an endeavor would have more likelihood of success as a national initiative. Issues of confidentiality were noted by respondents as a potential barrier (Respondent #11, Question 23). This could be an important consideration for projects that are interested in corporate adoption or participation. There are multiple complications that arise with courting corporate and industrial buy-in into an open project. The project must be perfectly dependable and have sturdy, clear licenses and use conditions for legal reasons. This survey helps to answer questions about information sharing and the required tools (data, data formats, and databases).

The survey can help us made a few conclusions. If open frameworks are introduced into the LCA community the projects will not have to combat active resistance, but will likely have to put much thought into surmounting ambivalence to active participation. Open projects will have to maintain high levels of quality control in order to be considered relevant.

4.5 Survey Conclusions

This survey has pointed out several gaps that could be remedied with new tools. First, respondents have brought up a need for deeper descriptions of LCA studies through LCA data formats, including more data, meta-data, statistical figures, and quality indicators. The answer may not be another standardized format project for several reasons. Previous standardization projects do not seem to serve the needs of respondents. Improvement of existing formats is managed through governing groups (such as SETAC), and improvements are made through long consensus-based processes. A more flexible format that does not require a strictly consistent data structure, like those achieved through linked data structures, could allow all contributors to append their own additions to a core LCA structure, allowing them to implement changes to LCA data structures they think are important.

Issues of transparency and insufficient datasets could be remedied with open projects. Ambivalence to participation will likely be an issue and will likely have to be mitigated with many of the factors that are attributed to successful open projects.

4.6 Areas For Further Questioning

The survey results prompted further questions. The ambivalence to sharing could be further explained by exploring which audiences (including the general public, competitors, etc) respondents would consider sharing data with, revealing whether they are interested, ambivalent, or threatened by certain groups (and why). Respondents could have been queried on their database use: which databases are known to respondents, which are used, why they are or are not used, and what the experience of searching for data has been like for users. It would also be interesting to know how many respondents develop their own software and whether they have encountered barriers in doing so. As the survey revealed a preference for using spreadsheet files for LCA data, it would be interesting to note which features make it the most attractive.

The initial survey revealed that members of the community desire advances in areas such as quality and transparency, but what must be explored now is which indicators are needed. Furthermore, these areas must be then represented satisfactorily in terms of data fields or calculated indicators derived from existing datasets.

Further inquiry into data ownership might also make it clear how it affects behavior and whether ownership arrangements could change. Specifically, could the provision of information and tools for publishing open data and open access studies result in more contributions to that end? There are likely some valid binding issues in ownership, but it is unclear which issues are flexible and which must remain closed.

5 OPEN SUSTAINABILITY INFO AND FOOTPRINTED.ORG – IMPLEMENTATION OF AN OPEN LINKED LCA PLATFORM

5.1 Introduction/Goals

In order to examine the potential for the open source movement that could make LCA more relevant, this research focused on the development of an open source/open access/open data/linked data web application to gather, organize, and present open LCA data. It was developed as a response to gaps explored in the literature (see section 2.1.3) and shared by the LCA community in the survey (see Chapter 4). This application was released as an open source project under the name Open Sustainability Info (OSI).

The source code project allows others to both launch their own database easily and contribute code and modifications for others to use. If one wanted to implement ones own LCA database they would download the OSI source code, make changes, and launch the modified application under a different name, but give attribution to OSI for the use of the code.

The open source application, known as Open Sustainability Info, will allow others to quickly and easily implement their own LCA databases. The focus of the original source code will be not be on the process of building a LCA study, rather it will be dedicated to facilitating the presentation, sharing, discussion and organization of LCA results. The main function will be to allow people to post the environmental impact of a material based on the functional unit of an LCA. It is hoped that the source code can provide a platform for iterative, improving, collective sharing and presentation of LCA data.

The code, along with a copy of the database that includes supporting data, is uploaded to github, a code repository. Interested parties can go to http://github.com/zapico/Footprinted to download the source code, read documentation, and make their own contributions.

5.2 Approach

5.2.1 Criteria

In order to validate the project, criteria for software success were used. Criteria for success for software projects vary. Most of them are based on a common private sector project delivery models, which focus on cost, time, and client satisfaction (Rook, 1987; Redmill, 1990). However, some project management literature proposes alternative methods for measuring success that are more appropriate for this project, including criteria such as fulfilling client requirements and production to specification (Turner, 1993; Morris and Hough, 1987; Wateridge, 1997). Wateridge (1997) proposes that, in order for IT projects to be successful, they must meet six criteria: meets user requirements, achieves purpose, meets timescale, meets budget, happy users, and meets quality. Some of these criteria were not applicable to this project. There was no budget, making that criterion irrelevant. Quality was deemed too subjective to measure or meet. The rest, however, were relevant and were used to judge the success of the project, as follows.

User requirements for this project were not defined by a client. However, survey feedback (discussed in Chapter 4) outlined gaps in LCA tools. That feedback essentially outlined community user requirements in the next generation of LCA tools. Of the feedback, increased transparency, the expansion of possible information in LCA data (e.g., statistical figures, quality indicators), and the ability to offer comments and feedback were important gaps that were used as criteria for meeting user requirements. The purpose of the software was to create a database that can share LCA data. In order fulfill this criterion, the case should be made for how the software fulfills the purpose. The timescale for the project was set for May 2010 to August 2011 to be completed as a master's thesis project. User satisfaction would ideally be measured by a survey of stakeholder use. Time was a limitation, so a survey of usability was not performed. A case study was used instead to show user satisfaction with accessing data on the site.

5.2.2 Description of OSI

Figure 5-1 illustrates how existing technologies, formats, and code libraries were used and connected together to create the OSI software. The choice to use linked data and the desire to have an online database were factors in choosing software libraries and technologies. The project was written predominantly in PHP (PHP Hypertext Preprocessor), a programming language used to develop web applications. The ECO ontology was chosen as the structure LCA data would take, explained in detail in section 5.3.1. MYSQL database software was chosen to manage the databases that would store the data. MYSQL is a program for storing data in relational database structures, not linked data. However, there is a PHP code library (ARC) that can take linked data and represent it in MYSQL tables that simulate a linked data structure. The ARC code was implemented to do so. Customized PHP code was written to create interfaces and convert user interactions from the interface into queries that find, add, or change data. The Graphical User Interface (GUI), Application Programming Interface (API), and file-based representation of data are the three existing interfaces for users to interact with the data, explained in detail in 5.3.3. The GUI allows users to access the data through web pages. The API allows programmers to write programs that can easily search through and retrieve data. Accessing the data through file-based representations allows users to examine the data in machine-readable form.

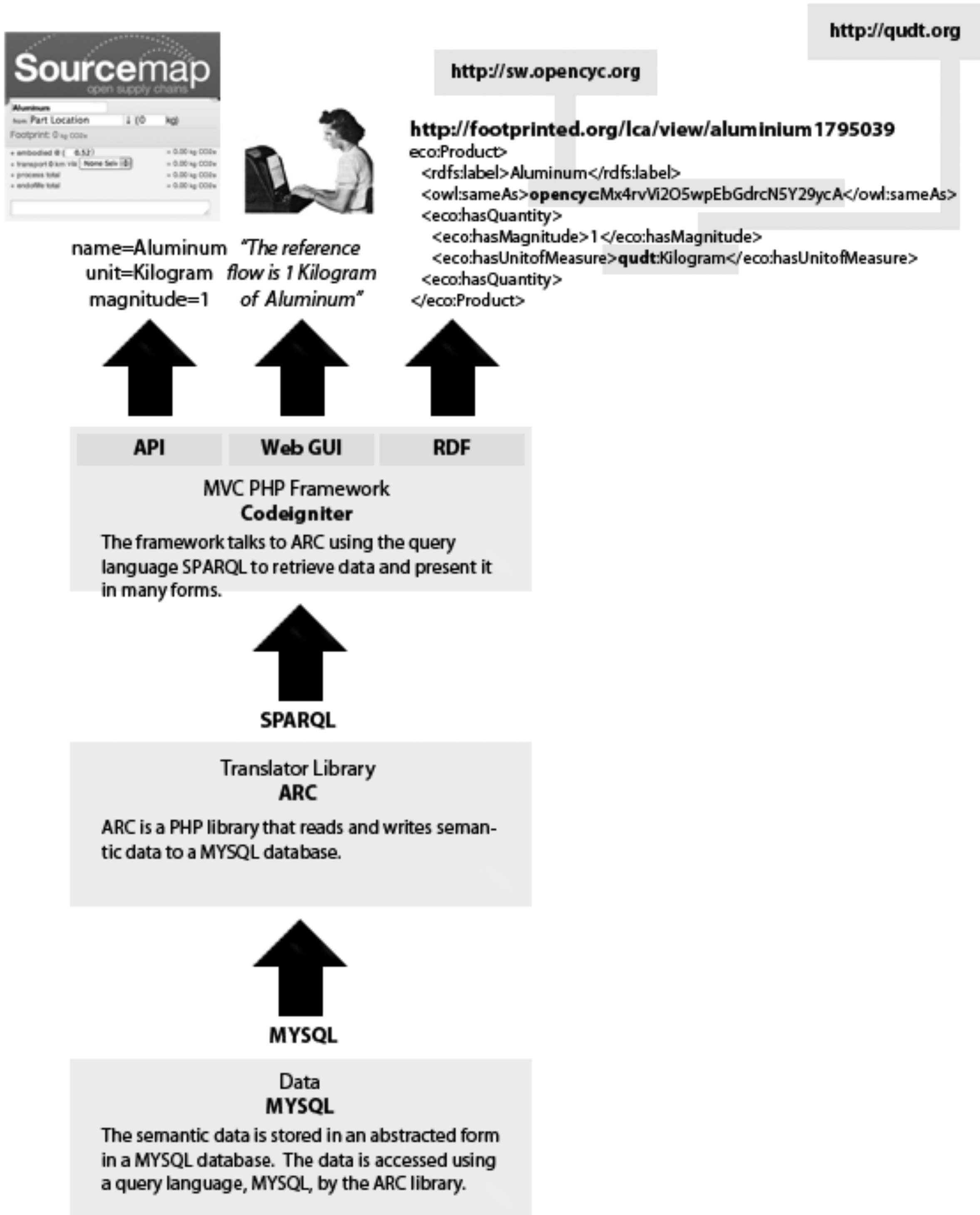


Figure 5-1 OSI Software Framework

This chapter explains the decision-making process for developing the code (section 5.3), the implementation of the code as Footprinted (section 5.4), the validation against criteria for measuring project success (section 5.5), and the discussion of outcomes (section 5.6).

5.3 Decision-Making Process for OSI Code

OSI was built through the combination of different technologies, formats, and code libraries, requiring careful consideration as to which would be used. The choices of which to use were important because they would affect how easy it would be for one to implement the code into a working database, how fast the code runs (and, in turn, how fast the site works), and how straightforward modifications to the code can be made. The first step was a choice in data format. The method of data storage was then chosen, followed by seeding the database with existing datasets. Finally, three interfaces were developed, including search and submission features.

5.3.1 Choosing a Data Format

The first step was the choice of format representing LCA data. There are many existing data structures for LCA data, which were discussed previously in section 2.1.4.1. Constructing a new data format is an extensive undertaking. Previously, such endeavors of that magnitude have been conducted by committees and through large academic projects. There were four data formats that could easily have been used. However, ECO is the only data format that is designed for linked data. While the ECO ontology is in its first version and is not complete, it provides an extensive vocabulary with which to describe LCA data and open for development by others. It was also designed to be compatible with Ecospold v2. Since Ecospold v2 is compatible with Ecospold v1 and ILCD, ECO would also be compatible with them. There is also the possibility of creating a new format/ontology. However, this would be redundant; ECO has the potential to be sufficiently compatible and, as linked data, could likely be built on with new, supplemental ontologies to create a multitude of hybrid-LCAs. Another argument against creating another format is the previous success of creating a new format to fix LCA data issues. There are now four formats that have only further fragmented LCA data. A more appropriate strategy may be to allow for the import or export of any LCA format, but store the data in ECO.

Since the ECO ontology was incomplete, other ontologies (BIBO, FOAF) were referenced in order to be able to describe bibliographic references and people (respectively) in a LCA dataset.

Software development required sample data. Four free, widely-available environmental impact datasets were used to test and seed the site's database. This includes Foodprint (the carbon dioxide equivalent for 90+ food items), Okala (300+ design materials and processes), ICE (a UK construction materials database), and the Canadian Raw Materials Database (15 commodities sourced from Canada). Copies of the datasets were made and translated into the ECO format.

5.3.2 Data Storage

Linked data can be stored in multiple ways. Most simply, linked data can be stored in a file with semantic markup. Notation 3 (N3), Terse RDF Triple Language (TTL), and Resource Description Framework (RDF) are three common formats for storing linked data in a file. These files can be directly accessed and viewed. These files are capable of being read by humans, but not optimized to do so. XLT can be added to style and format these documents to hide the markup and make the file appear more human-readable. In order to make linked data in this form truly useful, some software must be included to allow users and computers to query the files and retrieve particular pieces of data using the query language SPARQL.

Other projects utilize software libraries that simulate static linked data pages, storing semantic data in triples in a database. These libraries query and parse the linked data.

LCA data is fairly complex, especially compared to most existing implementations of linked data like dbpedia, various bibliography indexes, and friend-of-a-friend (foaf) applications (that link data on people and other agents). Because of their relative simplicity, many of these existing implementations provide a simple interface of attributes and their values, a simple dump of the data. For instance, a Dbpedia.org entry simply lists the label, description, and the location of other pages online that have the same topical subject (for instance, a dbpedia.org entry on aluminum may have included the URL of the entry on aluminum on Opentox). This is almost sufficient for simpler linked data, yet does not explain the data or present it in a friendlier format, which means the pages retain an intimidating aspect. This project needs an option that both deals well with complex data like LCA data and allows for reformation and interpretation of the data that makes it understandable.

The ARC library is an open source PHP library that allows for relatively simple implementation and is easy to incorporate with frameworks and libraries. This library stores linked data in a MYSQL and offers many useful features for managing linked data.

The sample data was placed into a MYSQL database using ARC. Code was then written to pull data from the database and write new LCA datasets to the database.

5.3.3 Interfaces

Three interfaces were developed to provide for different needs. The GUI was developed for users to access the data through a web browser to browse and search through data. The API was developed for web applications to directly access LCA data. Offering datasets as raw files allows for transparency and for users to acquire their own copies of the data to use for their own purposes.

The GUI is important because it is the "face" that the LCA community (and wider communities) will interact with in order to access tools and data and to participate. Users can access the database through a web browser through a web interface, a very common and widely used avenue. A web interface was developed to allow users to submit, search, compare, and graph LCA data. While RDF data can be created and made accessible without a web interface, it is more useful to have one because 1) a web interface is familiar and friendly to all internet users, and 2) people are still unfamiliar with RDF data and the existing interfaces for searching and managing it.

This GUI must be accessible in a usability sense; it must somehow introduce the user to unfamiliar concepts like linked data and complex LCA datasets in a useful manner. There are many linked data projects which can be potentially modeled. One high-profile example is Dbpedia.org, which shows simple semantic data with few filters. However, it, like similar projects (Opencyc.org, etc.) do not have interfaces that would be familiar to the average computer user. While the concentration of linked data is to make it machine-readable, this should not be done at the expense of making the same data human-friendly. This issue would be compounded with the complexity of LCA data, potentially making it unusable. Therefore, there should ideally be two faces, one for human reading and one for machine reading. The human-readable face can be developed by filtering LCA data through a friendly web application.

A second audience to consider is that of web application and software developers that may want to incorporate references to LCA data in their products and projects. While technically they can be expected to interact with the machine-readable linked data interface, linked data is not currently being incorporated into many web applications or static software. Current common ways of acquiring remote data include: 1) scraping a static, human-readable site using a script; and 2) sending a request to an Application Programming Interface (API) and receiving the data in a recognized format like XML or JSON. Therefore, if footprinted.org is expected to provide a service of accessible LCA data for the purpose of software development, an API would be required.

An Application Programming Interface (API) is an interface that allows a piece of software to interact with other software, for instance allowing one application to make use of the features of another application. This means (generically for all APIs) that a user should be able to use the API according to some input parameters (the user may have to provide a query or other information) and receive output in a usable format over multiple platforms.

The provision of LCA datasets as files allows for users to save the files and use them for their own ends. While it is not expected that the files will be used as much as the other two interfaces, provision of the files offers complete data transparency.

5.4 Implementation - Footprinted.org

Footprinted.org is the first implementation of the OSI code. Users of footprinted.org can begin to submit data, build applications that communicate with the footprinted.org database, and comment on LCA datasets. It has been developed with multiple partners and can be found at http://footprinted.org.

The design, public-facing API, and branding were done in conjunction with Jorge Zapico (of Royal Technical University, Stockholm, Sweden), Leonardo Bonanni, and Matthew Hockenberry (New York University). The implementation is a global open data project for carbon footprints and LCA data.

Figure 5-2 is a sitemap that represents the pages and the navigation hierarchy through the site.

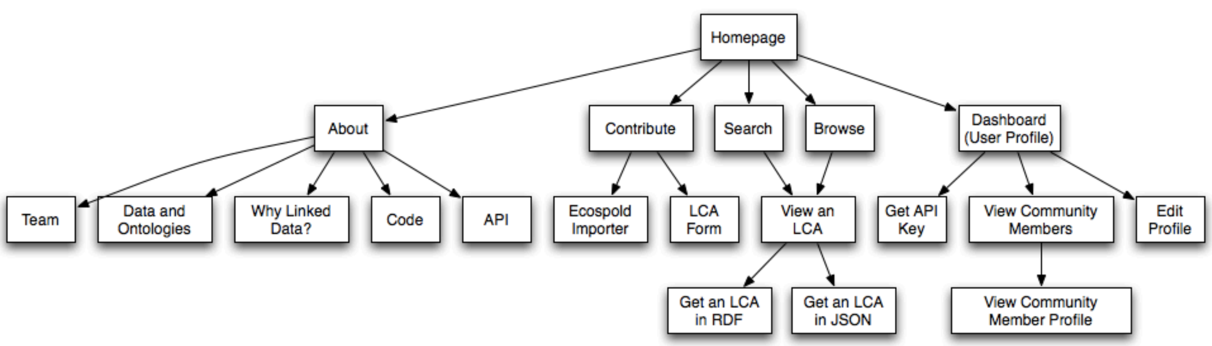


Figure 5-2 Footprinted.org Sitemap

5.4.1 Site Components

1) Welcome

The home page is designed as a gateway to the rest of the site to welcome users by sparking interest in the wealth of environmental impact assessment figures available without overwhelming them.

2) Search

Search capability is a common feature in web applications, and are especially important when the website is presenting large datasets. Users might often be searching for information about a particular substance or material, or a particular collection of LCA studies. The search capability on footprinted.org currently includes search by reference/source, year, country, category, and by the use of keywords that can be matched to LCA study titles.

Figure 5-3 is a screen capture of the search graphical user interface (GUI). This is what users will interact with to find LCA data.

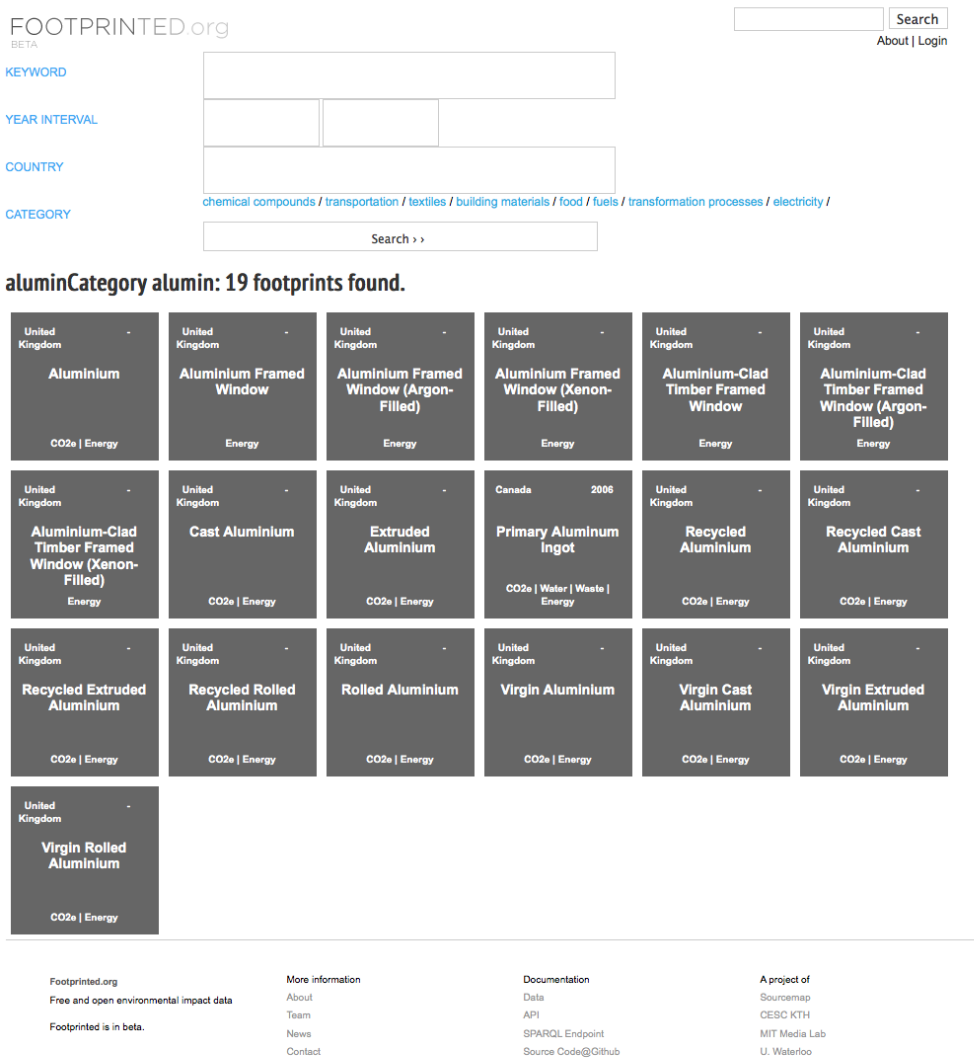


Figure 5-3 Footprinted.org Search Graphical User Interface

3) View

Each LCA data set should be able to be viewed by users in a comprehensible way. The site has a view which interprets a single LCA dataset into a summary that shares all information from the data set augmented by simple charts and infographics. LCA concepts including flows, impact assessment, functional unit, reference flow, bibliographic reference/source, years for which the data set is relevant, and country/region to which the study is relevant is presented in a simplified, easy to read format. Linked data from external datasets are also included to enhance information. These external datasets currently include Dbpedia.org (a linked data mirror of Wikipedia.org) and Opencyc.org, which provide encyclopedic and hierarchical categorization respectively. Figure 5-4 View of a LCA dataset is a screen capture for the visual representation of a LCA dataset.

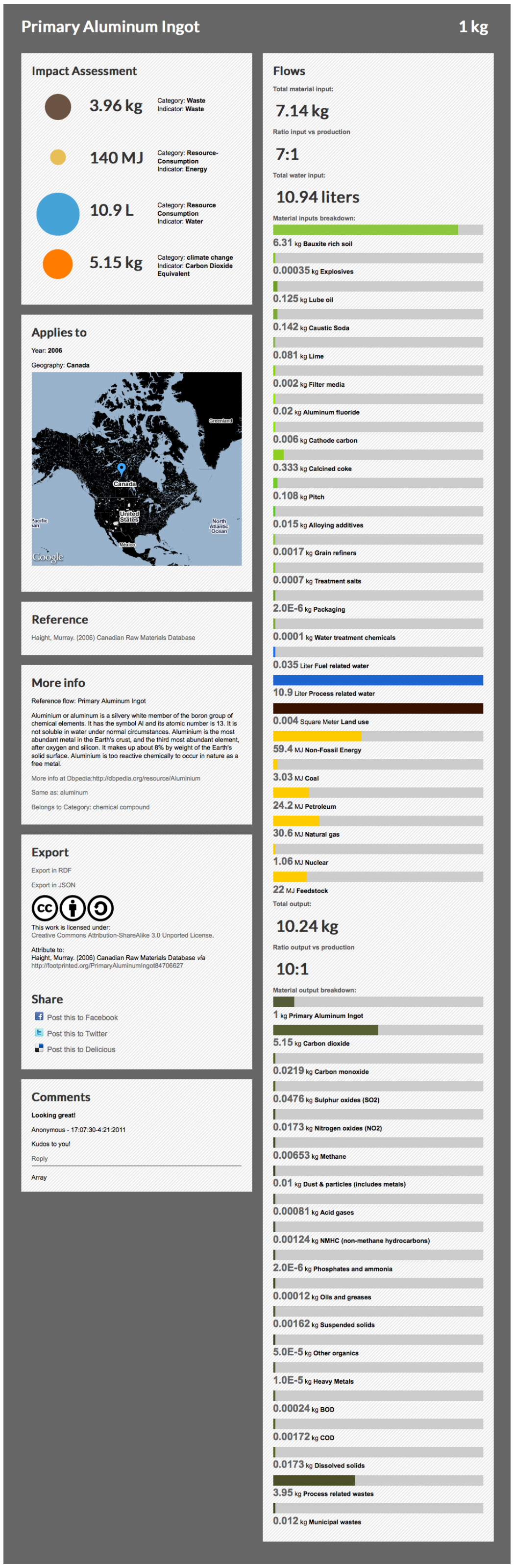


Figure 5-4 View of a LCA dataset

4) API

The API was created to give access to LCA data. An API was implemented for two reasons: 1) APIs are a common, controlled way to provide outside applications with data, and 2) most developers would find the complex structure of LCA and the use of linked data daunting to work with.

The API currently works as follows:

The root is http://footprinted.org/API

This will return the first 20 records in the DB. Currently three variables can be passed through the URL:

Name - This is the name of the product or process.

Offset - This allows for retrieving records from the results starting at any point in the results. An offset that is set to ten, so that it will skip the first ten records and return all records after that value. The value can be any positive integer.

Limit - This allows for limiting the number of results. The value should be any positive integer.

Using offset and limit together allows users to paginate through large datasets in manageable chunks.

A typical search may look like this:

http://footprinted.org/API?name=apple&limit=10&offset=10

The API currently only returns the functional unit/reference flow and impact assessment, as they are likely the most desirables feature for most potential API users.

5) Contribution

Users currently have three ways of submitting LCA studies: submission via email, submission by an ecospold translator tool, and submission via online form. The online form currently allows for the submission of simplified LCA data, including inputs and outputs, impact assessment indicators, and the bibliographic reference.

6) Comments

The capability for registered users to comment on LCA datasheets provides an initial attempt to provide a method of feedback and interaction.

5.5 Software Validation

The code had to meet four criteria in order for it to be considered a success. The first was that the code met user requirements, which were derived from the LCA community survey (discussed in Chapter 4). The project was developed to include features that addressed many of the concerns shared in the survey, such as transparency. This concern was addressed in the project by: 1) providing the LCA datasets in raw formats so all the information and how it was structured can be viewed; 2) providing the source code for the project for interested parties to examine; 3) providing documentation on the project; and 4) allowing for multiple sources of supplementary data in detailed bibliographic form. The provision for feedback was executed through a function for including comments. The expansion of information facilitated through the use of linked data, a flexible, easily expandable storage method. While most types of databases (such as relational databases) require changes to the schema (description of the database structure) to include new information, linked databases do not and can readily accept datasets of varying structures. While few expansions have been implemented in OSI and footprinted, they can be included as they are requested by members of the LCA community.

The second criterion was for the project to fulfill a purpose, which was to provide code for a new and better LCA database than what is currently available. The project has fulfilled the purpose of providing code for a functioning LCA database. Proof of this can be seen in the first implementation of the code, footprinted.org. The code also offers several features available to potential users that were not available in other databases, including the ability to submit data to the database through a form and ecospold importer and the ability to interact with the database through an API.

The third criterion, adhering to a timeline, was met. The first release of the project was completed by August 2011.

The fourth was user satisfaction. The case study below shows that the implementation has already been used to support Sourcemap, a supply chain service. Sourcemap (found at http://sourcemap.com) is the first user of footprinted.org data and services. The site, which allows others to map out supply chains and estimate their environmental impact has a “footprint catalogue” based on the footprinted.org database and API. The site pulls data from the API to allow sourcemap.com users to select materials and calculate their carbon footprint. Figure 5-5 shows a screenshot of a user creating a sourcemap (a supply chain expressed over a map) using data taken from footprinted.org.

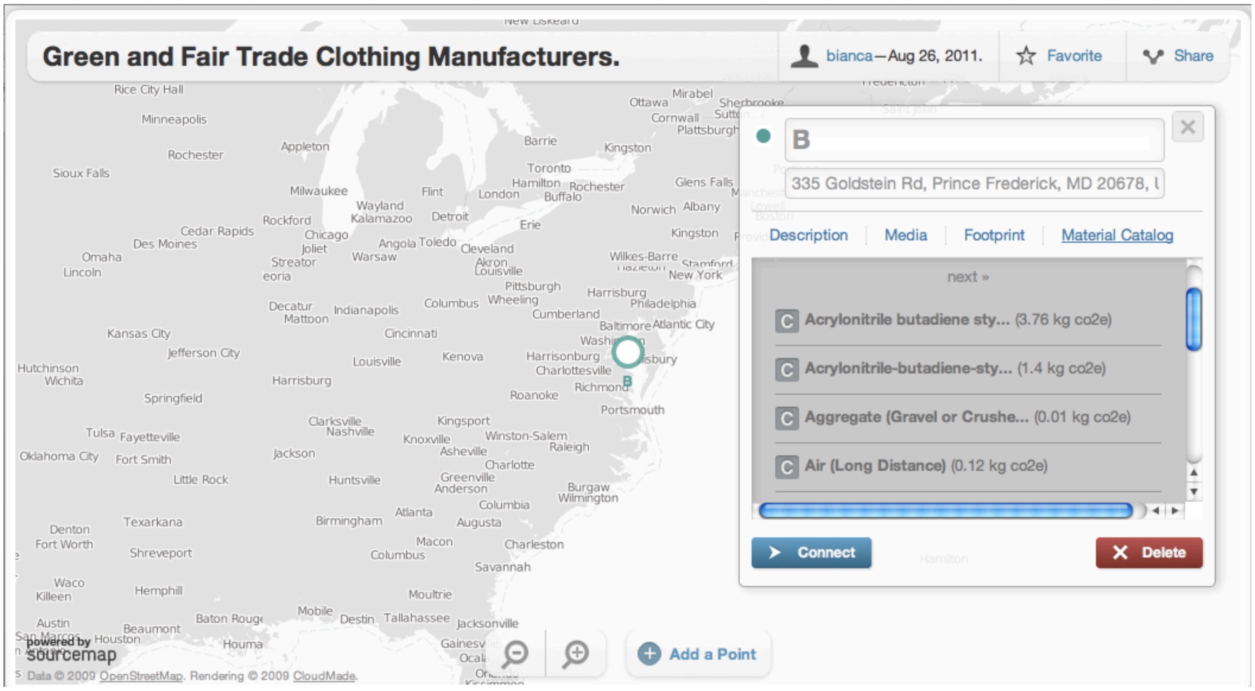


Figure 5-5 Screenshot of Sourcemap.com Using Footprinted.org Data

5.6 Outcomes

The result of this project is a software tool that increases relevancy of LCA data by widening its accessibility to stakeholders and addressing some of the gaps in existing LCA tools. The tool attempts to resolve issues shared in both LCA literature (discussed in Chapter 2) and a survey conducted on the LCA community (discussed in Chapter 4). The implementation strives to make LCA data more appealing and usable to wider audiences through their re-interpretation and re-presentation through the software website (see section 2 for discussion). Challenges with reporting and iterative improvement (discussed in Chapter 4) were addressed through tools embedded in the software that publish study data easily. The implementation attempted to surmount issues of standardization by hosting data in a flexible format (the ECO ontology) and offering that data in multiple formats and file types (this approach is incomplete, as not all formats were made available in the initial implementation). The desire for transparency expressed in the survey was addressed through the provision of each data set in formats that show all information available. It also addressed the need for an expanded format and vocabulary (expressed as desired in the survey) to describe more information in a study through the use of linked data ontologies, which allow for the flexible description of LCA information. Simple functionality for feedback (as identified also in the survey) was offered through comments that become documented as part of the data set. It not only resolves these issues, but also allows members of the LCA community and interested programmers to easily build upon the code of others, adapting it to suite individual needs, and offering their contributions back into the LCA community. However, some of the challenges outlined in section 2.1.3 were not addressed and require further research in order to develop potential solutions.

5.6.1 Challenges

The following challenges have been identified through the project process.

5.6.1.1 Semantic Query Speed

Queries for semantic data take much longer than queries in common relational databases. Coupled with the complex data structure, which requires complex queries, the load time for a LCA dataset or a search is almost prohibitively slow. This is especially important for the API, for which other sites are dependent on respectable load times of no more than 3 or 4 seconds.

In the long-term, the database will have to be optimized for speed. In the short term, a simple subset of information useful for search queries was placed into a relational database for search purposes. This modification does not reflect any other changes to the use of the website. It is ideally a temporary measure, as future changes to search capabilities will ideally include semantic searching.

5.6.1.2 Independently-developed Datasets

Using supporting datasets developed by others has both advantages and disadvantages. These datasets have a significant amount of data. However, users of these datasets are dependent on the quality and extent of that dataset with little ability to improve the dataset and little control over assured access to it. For the case of footprinted.org, this has proven to be a problem for the use of the Geonames and QUD datasets. The Geonames server does not offer reliable access to the Geonames linked dataset; the server is often overloaded with queries and may not respond. The QUD dataset has a fairly extensive dataset of units of measure, but does not include all units of measure, including SI derived units like megajoules, gigajoules, etc.

The current solution for footprinted.org is to mirror the datasets locally on the footprinted.org server. This assures reliable access and influence over the local copy when modifications are useful.

5.6.2 Next steps

The following describes areas that need further development in order to address some of the gaps and barriers remaining in LCA data.

1) Import from and Export to LCA formats

While there is a desire for standardization (see section 2.1.3 for discussion) of LCA data into one unified format, multiple efforts to do so (see section 2.1.4.1 for the discussion of these formats) have only resulted in multiple formats. One approach would be to make the format of LCA data irrelevant by allowing for the import and export of all formats. This would also avoid the necessity of the LCA community adopting and adapting to new formats. In order to make the site useful for LCA professionals LCA studies should be accessible in common LCA data formats like ILCD and Ecospold 1 and 2 so that users may be able to use them in software applications and other existing projects.

2) Quality indicators

Quality indicators were a consistent concern in survey responses (see section 4.5 for a discussion). The inclusion of quality indicators would provide the community with a desired feature. A simple initial assessment could be the implementation of Weidema and Wesnaes's (1996) method for indicating data quality based on five factors: geographic correlation, temporal correlation, technological correlation, reliability, completeness.

3) Expanded Ontology Features

The ECO ontology allows for the ability to describe LCA data in greater detail than other formats and approaches. For instance, while most LCA data formats allow for the description of only one process, the ECO ontology could be used to describe a set of interconnected processes and their respective flows and transfers. Another potential expansion is the representation in multiple languages. Providing tools to allow users to submit or modify datasets may beget data with more detail than any other database.

4) More linked datasets

Footprinted.org has incorporated multiple useful datasets. There are several untapped sources that could provide useful enhancements. There are multiple government datasets for the US and EU released in linked data format through the data.gov and publicdata.eu portals. Another opportunity is the Opentox toxicology dataset. The site would also benefit greatly if datasets for climate, agricultural zones, geography of mineral deposits, regional electricity mixes, etc were offered to users.

5) Unit Conversion

Unit conversion is a common technical expectation of LCA databases. It is desirable to provide LCA data in any valid unit of measure. A conversion tool could be built in order to allow users to retrieve data in any unit of measure.

6) Visualization

Visualization was discussed in section 2.1.3 as a gap in LCA data, especially pertaining to the expression of LCA datasets to audiences that do not have LCA expertise. LCA datasets are complex and difficult to interpret. The offering of multiple visualizations could give all users (both in LCA and non-LCA communities) the tools to better understand LCA studies.

7) Enhanced Search

Searching is a common facility provided within large and/or data-rich internet software tools. Searches could be conducted by more detailed factors. Future search features could include searches by 1) input or output substance, 2) range quantity of input or output, 3) impact type, or 4) threshold of impact.

8) Nested LCA Datasets

LCA studies already depend on secondary datasets of LCA studies for basic materials. It would be useful to allow recursive nesting of LCA studies so that users viewing a LCA dataset can view the LCA studies that the inputs and outputs are based on. This could lead to layered LCA datasets. For instance, LCA studies of cars could be built on datasets of engines and car doors, which in turn can be based on datasets of materials. This could help others build their own LCAs with their own adjustments and deviations.

9) Multiple Language Support

Data for a dataset do not change across different languages. Additionally, some existing databases do share their datasets in more than one language. It would be helpful to allow data providers to share their data in as many languages as they care to provide.

6 DISCUSSION AND RECOMMENDATIONS

This thesis has identified high-level gaps in LCA tools and data, and has discussed approaches that can address these gaps and make LCA data and databases more relevant for stakeholders inside and outside the LCA community.

6.1 Contributions to LCA Research

The contribution of this work to the LCA field has been to develop and demonstrate the use of open frameworks in ICT applied to make LCA more relevant. The tangible contributions (the open source code and website) can be used by people both inside and outside the LCA community immediately to share, discuss, change, and use LCA data easily. The intangible contributions (the survey and discussion) can be used to pinpoint appropriate next steps for changes in LCA data formats and databases by outlining some of the major concerns in the field and developing possible ICT solutions. The subject matter is current and addresses issues that are less foundational to the LCA field and more faciliatory, but advancing these issues appropriately could assure better tools and processes and increased relevancy for the field overall.

This research also addressed issues of meta-research (i.e., research of research). This area often receives less interest, but in a field like LCA, in which the value of research is that it informs decision-making, how information is communicated and shared becomes as important as the information itself. From one perspective, the environmental research field is only as useful as its ability to reduce anthropogenic environmental impacts. Unlike other scientific fields, for which discovery is an end, environmental science fields are value and goal based. If the LCA field is limited to LCA studies that do not often reach stakeholders in need of LCA data or real-world decision-making processes (and, in turn, do not improve environmental sustainability), we would have to question the relevance. As shown in this thesis, there are new and exciting opportunities to improve upon the relevance of the field.

The LCA field is also diminished in its relevance when results and data are not open or useable. This thesis has contributed to improving relevance by developing real software that can be used to host an LCA database and providing an internet software site that can be used to host LCA data, making LCA data accessible and usable. This improves upon previous databases through several features. First, users can easily and freely implement databases of their own with no or small modifications to the code. Second, the use of a data storage type that does not require consistent definition of data structure allows for variable structures for representing LCA data. This means that input-output LCAs, economic LCAs, supply chains, and basic LCAs can all be stored easily in the same database. Third, web applications and software can be built to draw directly from these online databases using the API, retrieving new and updated datasets easily. Fourth, the data storage method allows for the direct connection to supplementary datasets already available in the web like units of measure and geographic information. Fifth, the database allows for search capabilities that existing LCA databases do not have, including nested categories (e.g., minerals -> metals -> aluminum), nested geography (e.g., North America -> United States -> Louisiana -> New Orleans) and dates. Sixth, no other LCA databases allow for or have the functionality to allow people to directly submit data to a database; this implementation has two ways of doing so.

6.2 Critiques, Improvements, and Next Steps

The survey provided only a short exploratory query into the attitudes of LCA community members. The breadth and response rate was not adequate to support statistical analysis. Nonetheless, their concerns and attitudes have much to offer in informing future developments of LCA tools and methods. Previous surveys on the LCA community have dealt with how LCA studies are performed (the use of software, impact indicators, cost of process, and peer reviews) (Cooper and Fava, 2006) and the collection and use of data (Vigon and Jensen, 1995). These previous studies concentrated on LCA practice, but no inquiry has yet been made into attitudes and opinions of members of the LCA community. While there is a wealth of in-depth questions yet to be asked, establishing some initial views within the community was an important first step and lead to the right in-depth questions. It was important to establish whether the community had concerns at all about the tools available to them before asking what they would change. It led to a wide range of questions (Discussed in *Section 5-7*). Further inquiries could inform the general direction for improvements, and also specific decisions on the representation and sharing of LCA data.

The footprinted.org site is an initial offering, but the analysis included is self-reflective. This is not sufficient, as it is biased and a very narrow perception of the project. The analysis would have benefited greatly from community feedback on the site. Time was a factor in not soliciting this feedback, but a solicitation of feedback will be conducted post-thesis. This feedback can be used to develop more targeted features (Like those mentioned in *Section 6.4.3* Next steps) that will benefit the community and attract them to the site.

The Open Sustainability Info code has been offered publically for any interested party to implement on their own (http://opensustainability.info). When code is developed with the hope or expectation that others will use and modify it, the code must be intuitive and well documented. While the Open Sustainability code works, it would benefit from extensive additions to documentation and the re-write of certain code libraries. There are also many changes that could be made to simplify development and allow others to implement changes with less knowledge and mastery of all of the code.

6.3 Recommendations

An anecdote from a recent conference illustrates some of the problems to be faced: when footprinted.org was presented to conference-goers, suggestions were made. But those who offered those suggestions offered them as a change for the original creators to make, not for the owner of the suggestion to do so themselves. Speculatively, this could be because they did not realize that they could contribute the changes themselves, they simply did not see an incentive to do so, or because they may have respectfully assigned informal "ownership" of the project to the presenters and did not want to impose. Whatever the reason for this reaction, this could pose a barrier to deeper collaboration between researchers.

Academic research is a macro-collaborative process; each researcher builds on previous research, collectively synthesizing a whole field of knowledge. Currently, this is accomplished through individual research projects or experiments that are shared through journal articles. This way of building a field of knowledge was appropriate for a pre-IT period, but the inclusion of ICT tools could not only improve the rate of research, but also the quality by helping to fully integrate previous research. Making models, code, and data available allows others to quickly and transparently access previous research and directly alter previous work, producing tools and supplementary data that is produced and used together. This is especially pertinent for the LCA field because the period of time and (technological and geographical) context for which a LCA study is relevant can be very short, making the period of time for which research results are relevant much shorter compared to other fields. This begs a method for sharing research results that can facilitate an appropriate pace.

Whether the process begins with the open sustainability project, the footprinted.org platform, or with other projects (including existing ones like OpenLCA or Enipedia), researchers should embrace this very collaborative and accelerated way of collective problem-solving and research through frequent feedback, small contributions to other projects and software, sharing and updating of raw results, and the release of LCA models.

Projects like footprinted.org can provide solutions to many problems by providing a platform on which to happen, but the perception of how researchers contribute will also have to change. This can be done through a change of language at conferences and in journals that invite collaboration and push accessibility and transparency, more use of ICT vectors like email, websites, and online message boards to communicate and the aggressive use of programs, data formats and files that are easy to use and modify (like those that are open source, free, transparent, and linked).

The use of a tool like footprinted.org can also enhance journal article publications. Many journals allow the inclusion of supplementary materials to accompany journal articles. Journal articles that describe LCA studies could include a URL to a LCA dataset on footprinted.org (or any other online document that contains LCA data) so that readers can easily access details and examine the study in greater detail.

6.4 Potential Future Applications

It is hoped that others will implement their own LCA databases using the Open Sustainability Info code. The interested parties could be the facilitators of existing databases, which may benefit from easier management and new features. Another possible interested party could be the International Journal of LCA, which may benefit from having a journal-specific database to host datasets that have corresponding articles in the journal. For researchers that do not or cannot implement their own versions of OSI, they may make use of the footprinted.org database by submitting data on that platform.

It is also hoped that some of the future applications will be to simultaneously (if separately) use footprinted.org to build software tools that help communicate environmental issues to the public using reliable data and also help LCA practitioners easily build complex and robust models for their own benefit.

6.5 Conclusions

LCA is extremely valuable and has produced a vast collection of data. Yet the influence of this data is limited, the development of tools and visualizations modest, and the accessibility difficult. The exploration of open frameworks for LCA has revealed a small, burgeoning group of efforts and many opportunities.

The potential positive implications are exciting. However, engagement from the LCA community will have to be aggressively pursued. Uptake of existing open LCA projects is not significant. While the community has expressed some willingness to operate within open frameworks, new projects will have to address the gaps that the community is most concerned about without sacrificing the benefits of some of the best LCA databases. This means increased transparency while retaining quality.

However, continued apathy toward making LCA data more relevant outside the LCA community may actually be damaging. As software improves, the need for sustainability information increases the community risks losing authority status on the subject of sustainability metrics. Currently, few in the media or within NGOs and other interested parties refer to LCAs for their figures. They will seek out the most accessible environmental impact data. It is for everyone's benefit that that data be LCA data, because LCA data is the most detailed, accurate, reliable environmental impact data and because its worth resides in the environmental change it can influence.

REFERENCES

Asay, M. 2007. "Learning from a Failed Experiment in Open Source." *CNET*, October news.cnet.com/8301-13505_3-9797037-16.html.

Asay, M. 2010. "Symbian: A Lesson on the Wrong Way to Use Open Source." *GigaOM*, October http://gigaom.com/2010/10/28/symbian-a-lesson-on-the-wrong-way-to-use-open-source.

Anex, RP., and W Focht. 2002. "Public Participation in Life Cycle Assessment and Risk Assessment: A Shared Need." *Risk Analysis* 22:861-877.

Bagozzi, RP., and UM Dholakia. 2006. "Open Source Software User Communities: A Study of Participation in Linux User Groups." *Management Science* 52:1099-1115.

Bare, J., T Gloria, and G Norris. 2006. "Development of the method and U.S. normalization database for Life Cycle Impact Assessment and sustainability metrics." *Environmental Science & Technology* 40:5108-15.

Baumann, H. 1995. "LCA Use in Swedish Industry." *Forestry* 1:122-126.

Behlendorf, B. 2009. "How Open Source Can Still Save the World." *International Federation For Information Processing* 2009-2009.

Bieda, B. 2007. "The Use of Life Cycle Assessment (LCA) conception for Mittal Steel Poland SA Energy Generation - Krakow Plant Case Study." *1st International Conference on Energy and Sustainability*. New Forest, England. 1:165-174.

Bjorklund, AE. 2002. "Survey of approaches to improve reliability in LCA." *International Journal of Life Cycle Assessment* 7:64-72.

Bradley, F. 2009. "Discovering Linked Data." *Library Journal* 134:48-50.

Bretthauer, D. 2002. "Open source software: A history." *Information Technology and Libraries* 21:3-10.

Carlson, R., and J. Tivander. 2001. "Data definition and file syntax for ISO/TS 14048 data exchange." *CPM Chalmers University Of Technology*.

Carlson, R., and AC Pålsson. 2001. "First examples of practical application of ISO/TS 14048 Data Documentation Format." *CPM - Chalmers University Of Technology*.

Carlson, R., and AC Pålsson. 2006. "Implementation of integrated environmental information systems." *Centre for Environmental Assessment of Product and Material Systems – Chalmers University of Technology*.

Carlson, R., AC Pålsson, and J Tivander. 2003. "Data format mapping between SPINE and ISO / TS 14048". *Centre for Environmental Assessment of Product and Material Systems – Chalmers University of Technology*.

Ciroth, A. 2007. "OpenLCA – A new open source software for Life Cycle Assessment." *International Journal of Life Cycle Assessment* 12:209 - 210.

Ciroth, A, and H Becker. 2006. "Validation – The Missing Link in Life Cycle Assessment." *International Journal* 11:295 - 297.

Civic Apps for Greater Portland. 2010. “Civic Apps for Greater Portland.” civicapps.org (Accessed July 20, 2010).

Constant, D., S. Kiesler, and L. Sproull. 1994. “What’s Mine Is Ours, or Is It? A Study of Attitudes about Information Sharing.” *Information Systems Research* 5:400-421.

Cooper, J, and JA Fava. 2006. “Life-Cycle Assessment Practitioner Survey.” *Journal of Industrial Ecology* 10:12-14.

Crowston, K, H Annabi, and J Howison. 2003. “Defining open source software success.” *Proceedings of the 24th International Conference on Information Systems*. 3-17. Seattle, WA.

Crowston, K, J Howison, H Annabi. 2006. “Information Systems Success in Free and Open Source Software Development: Theory and Measures.” *Software Process: Improvement and Practice* 11:123-148.

Curran, MA, and P Notten. 2006. “Summary of Global Life Cycle Inventory Data Resources.” *SETAC/UNEP Life Cycle Initiative*.

Davis, C, I Nikolic, and GPJ Dijkema. 2010. “Industrial Ecology 2.0.” *Journal of Industrial Ecology* 14:707-726.

Delft University. n.d. “Enipedia”. http:// http://enipedia.tudelft.nl/ (Accessed January 11, 2011).

European Commission - Joint Research Centre - Institute for Environment and Sustainability. 2010. “LCA Database List.” http://lca.jrc.ec.europa.eu/lcainfohub/databaseList.vm (Accessed May 10, 2010).

Falkenstein, E., F Wellenreuther, and A Detzel. 2010. “LCA studies comparing beverage cartons and alternative packaging: can overall conclusions be drawn?” *The International Journal of Life Cycle Assessment* 15:938-945.

Fang, Y., and D Neufeld. 2009. “Understanding Sustained Participation in Open Source Software Projects.” *Journal of Management Information Systems* 25:9-50.

Finnveden, G., et al. 2009. “Recent developments in Life Cycle Assessment.” *Journal of Environmental Management* 91:1-21.

Fishbein, EA. 1991. “Ownership of Research Data.” *Academic Medicine* 66:129-133.

Fogel, K. 2009. “Producing Open Source Software - How to Run a Successful Free Software Project.” O’Reilly Media.

Stallman, RM. 2002. “Free Software, Free Society: Selected Essays of Richard M Stallman.” GNU Press. Boston, MA.

Frewer, LJ. 2000. “Public Participation Methods: A Framework for Evaluation.” *Science, Technology, and Human Values* 25:3-29.

Frischknecht, R., and G. Rebitzer. 2005. “The ecoinvent database system: a comprehensive web-based LCA database.” *Journal of Cleaner Production* 13:1337–1343.

Frischknecht, R., et al. 2005. “The ecoinvent Database: Overview and Methodological Framework” *The International Journal of Life Cycle Assessment* 10:3–9.

Frischknecht, R. 2005. “From heterogenous databases to unified and transparent LCI data.” *International Journal of Life Cycle Assessment* 10:1-2.

Frischknecht, R. 2006. “LCA Methodology Notions on the Design and Use of an Ideal Regional or Global LCA Database.” *International Journal of Life Cycle Assessment* 1:40-48.

Fuchs, C. 2006. “The implication of new information and communication technologies for sustainability.” *Environment, Development, and Sustainability* 10:291-309.

Gillham, B. 2008. “Small-Scale Social Survey Methods.” Continuum International Publishing Group.

Gul, S, T Shah, and T Baghwan. 2010. “Culture of open access in the University of Kashmir: a researcher’s viewpoint.” *Aslib Proceedings* 62:210-222.

Hall, W et al. 2006. “Creating a Science of the Web.” *Scientific American* 313:769-771.

Hischier, R., and B Weidema. 2009. “Improved data format meeting the challenges of future LCI databases.” *Proceedings from the 9th North American Conference on Life Cycle Assessment*.

Hogge, B. 2010. *Open Data Study: New Technologies*. London.

.Heijungs, R., and R Klejin. 2001. “Numerical Approaches to Life Cycle Interpretation.” *International Journal of Life Cycle Assessment* 6:141-148.

International Organization for Standardization. (2002). ISO 14040 : Environmental management -- Life cycle assessment -- Principles and framework.

International Organization for Standardization. (2002). ISO 14044 : Environmental management -- Life cycle assessment -- Requirements and guidelines.

International Organization for Standardization. (2002). ISO/TS 14048 : Environmental management -- Life cycle assessment -- Data documentation format.

International Organization for Standardization. (2006). ISO 14064 3: Greenhouse gases – Part 3: Specification with guidance for the validation and verification of greenhouse gas assertions.

Ishii, H, with C Csikszentmihalyi. 2010. “Small Business Applications of Sourcemap: A Web Tool for Sustainable Design and Supply Chain Transparency.” *CHI 2010: Sense and Sustainability*.

Jarvenpaa, SL, and DS Staples. 2000. “The use of collaborative electronic media for information sharing: an exploratory study of determinants.” *Information Systems* 9:129-154.

Kannan, R. 2004. “LCA/LCCA of oil fired steam turbine power plant in Singapore.” *Energy Conversion and Management* 45:3093-3107.

Kim, S., and BE Dale. 2008. “Life cycle assessment of fuel ethanol derived from corn grain via dry milling.” *Bioresource Technology* 99:5250-5260.

Klopffer, W. 1997. “Life Cycle Assessment - From the Beginning to the Current State.” *Environmental Science and Pollution Research* 4:223 - 228.

Klöpffer, W. 2006. “The Role of SETAC in the Development of LCA.” *International Journal of Life Cycle Assessment* 1:116-122.

Kolekofski, K. 2003. “Beliefs and attitudes affecting intentions to share information in an organizational setting.” *Information & Management* 40:521-532.

Kumar, H., A Azapagic, P Schepelmann, and M Ritthoff. 2010. “Options for broadening and deepening the LCA approaches.” *Journal of Cleaner Production* 18:120-127.

Lakhani, K., and RG Wolf. 2003. “Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects.” *SSRN Electronic Journal* 1-27.

Lee, J, P Ocallaghan, and D Allen. 1995. “Critical review of life cycle analysis and assessment techniques and their application to commercial activities.” *Resources, Conservation and Recycling* 13:37-56.

Lindman, J., J Juutilainen, and M Rossi. 2009. “Beyond the Business Model: Incentives for Organizations to Publish Software Source Code.” *International Federation for Information Processing* 47-56.

Masum, H., and M Tovey. n d. “Given Enough Minds - Bridging the Ingenuity Gap.”

Masum, H., and M Tovey. 2008. “Scaling up open problem solving.” *Collective Intelligence: Creating a Prosperous World at Peace*.

McCandless, D. 2010. “Information is Beautiful Hacks in India with David Cameron.” *The Guardian UK*, July 30 www.guardian.co.uk/news/datablog/2010/jul/30/information-beautiful-india-cameron.

Pattanakittipong, P. n d. “How to Develop LCI Database for Thai Industry: case study refinery products.”

Miller, DR, and D. John Doyle. 2007. “Understanding the open access data movement.” *Canadian Journal of Anesthesia* 54:949-950.

Miller, S., and TL Theis. 2008. “Comparison of Life-Cycle Inventory Databases: A Case Study Using Soybean Production.” *Journal of Industrial Ecology* 10:133-147.

Molina-murillo, SA., and TM Smith. 2009. “Exploring the use and impact of LCA-based information in corporate communications.” *International Journal of Life Cycle Assessment* 14:184-194.

Mongelli, I., S Suh, and G Huppes. 2005. “A Structure Comparison of two Approaches to LCA Inventory Data, Based on the MIET and ETH Databases.” *The International Journal of Life Cycle Assessment* 10:317–324.

Morris, P. W. G. and Hough, G. H. 1987. The Anatomy of Major Projects. John Wiley.

Murray-Rust, Peter. 2008. “Open Data in Science.” *Serials Review* 34:52-64.

Mutel, CL., and S Hellweg. 2009. “Regionalized life cycle assessment: computational methodology and application to inventory databases.” *Environmental Science & Technology* 43:5797-803.

Nash, JE. 1993. “Ownership and Outreach: A Model for Administration of Shared Data.” *Annals of the Missouri Botanical Garden* 80:304.

Open Tox. n d. “Open Toxicology.” http://opentox.org (Accessed September 17, 2010).

Ossés de Eicker, M., R Hischier, H Hurni, and R Zah. 2010. "Using non-local databases for the environmental assessment of industrial activities: The case of Latin America." *Environmental Impact Assessment Review* 30:145-157.

Ossés de Eicker, M et al. 2010. "The applicability of non-local LCI data for LCA." *Environmental Impact Assessment Review* 30:192-199.

Peters, S. 2009. "Open Source Is Changing the Way Work Gets Done." *International Federation For Information Processing* 2009-2009.

Pålsson, AC. 1999. "Review of LCI-data at SPINE @ CPM." *Centre for Environmental Assessment of Product and Material Systems – Chalmers University of Technology*.

Pålsson, AC. 2006. "Quality assessment of the ecoinvent and SPINE@CPM databases based on the ISO/TS 14048 data documentation format." *Centre for Environmental Assessment of Product and Material Systems - Chalmers University of Technology*.

PRÉ CONSULTANTS. 2006. "Introduction to LCA with SimaPro 7."

Redmill, F. J. 1990. "Considering quality in the management of software-based development projects." *Information and Software Technology* 32:18-22.

Robson, Colin. 2000. Small-scale Evaluation. Sage Publications, Inc.

Rook, P. 1986. "Controlling software development projects". *Software Engineering Journal* 1:7-16.

Rosener, J. 1978. "Citizen Participation - Can We Measure Its Effectiveness." *Public Administration* 38:457-463.

Seringhaus, MR, and MB Gerstein. 2007. "Publishing perishing? Towards tomorrow's information architecture." *Bioinformatics* 8:17.

Shadbolt, N, and T Berners-Lee. 2008. "Web Science Emerges." *Scientific American* 299:76-81.

Singh, BJ, A Krasowski, and SP Singh. 2011. "Life Cycle Inventory of HDPE Bottle-Based Liquid Milk Packaging Systems." *Packaging Technology and Science* 29:49-60.

Singleton, R. 2010. Approaches to Social Research. New York: Oxford University Press, Inc.

Skone, TJ. 2000. "What is life cycle interpretation?" *Environmental Progress* 19:92-100.

Spatari, S, M Betz, H Florin, M Baitz, and M Faltenbacher. 2001. "InLCA : Selected Papers Using GaBi 3 to Perform Life Cycle Assessment and Life Cycle Engineering." *International Journal of Life Cycle Assessment* 6:81-84.

Stam, W. 2009. "When does community participation enhance the performance of open source software companies ?." *Research Policy* 38:1288-1299.

Steen, B., and R Carlson. 1995. "SPINE A Relation Database Structure for Life Cycle Assessments." *Environmental Research*.

Suber, P. 2008. “Open Access in 2007.” *Journal of Electronic Publishing* 11.

Sue, VM, and LA Ritter. 2007. Conducting Online Surveys. Thousand Oaks: Sage Publications, Inc.

Tohidinia, Z., and M Mosakhani. 2010. “Knowledge sharing behaviour and its predictors.” *Industrial Management & Data Systems* 110:611-631.

Tomlinson, B., 2010. Greening through IT: Information Technology for Environmental Sustainability. Cambridge, MA: MIT Press.

Tovey, Mark. 2008. “Mass collaboration, open source, and social entrepreneurship.” *Collective Intelligence: Creating a Prosperous World at Peace*.

Turner, J. R. 1993. The Handbook of Project-based Management. McGraw-Hill.

Vagt, H et al. 2008. “D8 - LCA options for sustainable governance assessed” *Co-ordination Action for innovation in Life-Cycle Analysis for Sustainability*.

Vigon, BW., and AA Jensen. 1995. “Life cycle assessment: data quality and databases practitioner survey.” *Journal of Cleaner Production* 3:135-141

Wateridge, J. 1997. “How can IS/IT projects be measured for success?” *International Journal of Project Management* 16:59-63

Weidema, BP, and M.S. Wesnæs. 1996. “Data quality management for life cycle inventories–an example of using data quality indicators.” *Journal of Cleaner Production* 4:167–174.

Weidema, BP. 2000. “Increasing Credibility of LCA.” *International Journal of Life Cycle Assessment* 5:63 - 64.

Weidema, BP., et al. 2001. “Procedural guideline for collection, treatment, and quality documentation of LCA data.”

Wolf, MA., et al. n d. “ILCD format: Scope, Development, Compatibility” *European Commission – Joint Research Commission*.

Zapico, JL., N Brandt, and M Turpeinen. 2010. “Environmental Metrics.” *Journal of Industrial Ecology* 14:704-706.

APPENDIX A. LCA DATABASE REVIEW COLLECTED DATA

Table A-1 The Databases Reviewed

#	Name	URL
1	Australian Life Cycle Inventory Database Initiative	http://www.auslci.com.au/
2	Australian Life Cycle Inventory Data Project	http://www.cfd.rmit.edu.au/programs/life_cycle_assessment/life_cycle_inventory
3	BUWAL 250	http://www.umwelt-schweiz.ch/buwal/eng
4	Canadian Raw Materials Database	http://crmd.uwaterloo.ca
5	ecoinvent	http://www.ecoinvent.ch
6	EDIP	http://www.lcacenter.dk
7	Franklin US LCI	http://www.pre.nl
8	German Network on Life Cycle Inventory Data	http://www.lci-network.de
9	Japan National LCA Project	http://www.jemai.org.jp/lcaforum/index.cfm
10	Korean LCI	http://www.edp.or.kr/lcidb/english/lcidb/lcidb_intro.asp
11	LCA Food	http://www.lcafood.dk
12	Spine@CPM	http://www.globalspine.com
13	Swiss Agricultural Life Cycle Assessment Database (SALCA)	http://www.reckenholz.ch/doc/en/forsch/control/bilanz/bilanz.html
14	Thailand LCI Database Project	http://www.thailcidatabase.net/
15	US LCI Database Project	http://www.nrel.gov/lci
16	CPM LCA Database	http://cpmdatabase.cpm.chalmers.se/
17	DEAM™	http://www.ecobilan.com/uk_deam.php
18	DEAM™ Impact	http://www.ecobilan.com/uk_team05.php
19	EIME V11	http://www.codde.Fr
20	esu-services database v1	http://www.esu-services.ch/inventories.htm
21	GaBi databases 2006	http://www.gabi-software.com/
22	KCL EcoData	www.kcl.fi/eco
23	Option data pack	http://www.jemai.or.jp/CACHE/lca_details_lcaobj6.cfm
24	PlasticsEurope Eco-profiles	http://www.plasticseurope.org/content/default.asp?PageID=392
25	ProBas	www.probas.umweltbundesamt.de
26	SALCA	http://www.agroscope.admin.ch/
27	SimaPro database	http://www.pre.nl/simapro/inventory_databases.htm
28	LEGEP	http://www.legep.de
29	The Boustead Model	http://www.boustead-consulting.co.uk
30	Waste Technologies Data Centre	[Under Construction]
31	IDEMAT	http://www.idemat.nl/
32	Inventory of Carbon & Energy (ICE)	http://people.bath.ac.uk/cj219/
33	Foodprint	http://foodprint.awardspace.com/foodprintmethods.pdf
34	Okala	http://www.sustainableminds.com/product/methodology
35	USDA National Agricultural Library Digital Commons	http://riley.nal.usda.gov/nal_display/index.php?info_center=8&tax_level=1&tax_subject=757
36	European Aluminium Association	http://www.eaa.net/
37	Deutsches Kupferinstitut/ European Copper Institute	http://www.kupfer-institut.de/lifecycle/
38	European Federation of Corrugated Board Manufacturers (FEFCO)	http://www.fefco.org/
39	International Iron and Steel Institute (IISI)	http://www.worldsteel.org/
40	Nickel Institute	http://www.nickelinstitute.org/index.cfm/ci_id/11.htm

Table A-2 Access

#	Availability	Access	Language	Areas of Focus	Geographic Coverage	Number of Datasets
1	Active	Free	English	Electricity	Australia	1
2	Dead	Free	English		Australia	>100
3	Active	with Software	English, German, French	Packaging	Switzerland	16
4	Active	Free	English, France	Raw Materials	Canada	12
5	Active	Fee	English, Japanese, German		Global, Europe, Switzerland	1000
6	Dead	Fee	English		Denmark	>100
7	Active	with Software	English	Packaging	USA	>10
8	Active	Free	German, English		Germany	
9	Active	Fee	Japanese		Japan	>600
10	Active	Free	Korean		Korea	341
11	Active	Free	English	Food	Denmark	50
12	Active	Free	English	Industrial Materials	Global	100
13		Free	German	Agriculture	Switzerland	>100
14	Active	Free	Thai, English	Multi-industry	Thailand	160
15	Active	Free	English		USA	73
16	Active	Free	English, Swedish		Sweden	
17		with Software	English, French, German, Spanish		UK	300
18		with Software	English, French, German, Spanish		UK	
19	Active	with Software	Spanish French English	electrical, mechanical and electronic products	Canada, China, France, Japan	820
20	Active	Tiered	German English		Global	0

21		Fee	Afrikaans, Catalan, Chinese, Czech, Danish, Dutch, English, French, German, Hungarian, Italian, Japanese, Malay, Moldavian, Norwegian, Portuguese, Punjabi, Russian, Slovak, Spanish, Swedish, Thai, Turkish, Twi, Ukrainian, Vietnamese	Global		0	0
22	dead	Fee	English, Finnish, German, Swedish		Global		
23	Active	Fee	Japanese	chemical production, iron & steel and waste management process data	Japan		500
24	Active	Free	English	Plastics	Europe		84
25	Active	Free	German		Germany		8000
26	Active	Fee	English, French, German	Agriculture	Switzerland		700
27	Active	With Software	English				
28	Active	with Software	German		Germany/Italy		
29	Active	Tiered	English		Global		13000
30	Suspended	Free	English	Waste/Recycling Processes			
31	Active	Free	English, French, German, Portuguese, Spanish				>1000
32	Active	Free	English	Construction Materials	UK		350
33	Active	Free	English	Food	USA		90

34	Active	with Software	English	transportation, building materials, manufacturing and end-of-life processes		250
35	Active	Free	English	Agriculture	US	0 (as of May 2011)
36	Active	Free	English	Aluminium	Europe	
37	Active	Free	English	Copper	Europe	3
38	Active	Free	English	Carboard	Europe	1
39	Active	Free	English	Steel	Global	
40	Active	Free	English	Nickel		

Table A-3 Access

	Available Formats	Available file types	Last Update	LCI/ LCIA	Supplier	Supplier Type	Iterative/ Updated	Open Submission	For Software	ISO/TS 14048
1	ilcd, ecospold1	xml,xls	2011	LCI	Australian Life Cycle Assessment Society (ALCAS)	Association		Yes		Yes
2					Royal Melbourne Institute of Technology	Academic				
3			2004		Swiss Federal Office for the Environment	Government		No	SimaPro, GaBi	No
4		pdf	1998	LCI	University of Waterloo	Academic	No	No		No
5	ecospold1	xml, xls		Both	ecoinvent Centre	Academic	Yes	Yes	SimaPro,GaBi	Yes
6			2003	Both	Force Technology, IPU	Consulting	No	No	Gabi	
7				LCI	Franklin Associates, A Division of ERG	Consulting		No	SimaPro	Yes
8	ecospold			LCI	German Network on Life Cycle Data	Association				
9				Both		Government				
10		html		LCI	Korean Ministry of Commerce, Industry and Energy	Government				
11	SimaPro db	nx1	2007	Both	Ministry of Food, Agriculture and Fisheries, Danish Research Institute of Food Economics, Aalborg University	Academic/ Government	No	No	SimaPro	Yes
12	SPINE @ISO14048	html			Chalmers University of Technology	Academic				Yes
13					Agroscope - Swiss Government	Government				

14			2008		National Metal and Materials Technology Center (MTEC), Ministry of Industry Department of Industrial Works Thailand Research Fund (TRF), the Federation of Thai Industries (FTI) and the Thailand Environment Institute (TEI)	Governme nt	No			
15	ecospol d1, detailed spread sheet	xls, xml, spold	2007	LCI	NREL	Governme nt	Yes	Yes		yes
16	SPINE @ISO1 4048	html	Datas ets publis hed years rangin g from 1991-2010. Some data has collect ion years	Bot h	Center for Environmental Assessment of Product and Material Systems - CPM	Academic	No			yes
17				LCI	Ecobilan – Pricewaterhou seCoopers	Consulting				
18				LCI A	Ecobilan – Pricewaterhou seCoopers	Consulting				
19					Bureau Veritas CODDE	Consulting	Yes		EIME	
20	SimaPr o, ecospol d1	xml,nx1		LCI	ESU-services Ltd.	0				yes
21					PE International GmbH	Consulting				
22					Oy Keskuslaborat orio-Centrallaborat orium Ab, KCL	Consulting				
23				Bot h	(JEMAI)	Associatio n			Jemai-LCA-Pro	
24		pdf, xls	2005	LCI	PlasticsEurop e	Associatio n	No	No		
25		pdf,xls	2008	LCI	Federal Environmental Agency, Öko-Institut	Governme nt	Yes			

26			Both	Agroscope Reckenholz-Tänikon Research Station ART	Goverment				
27				PRé Consultants B.V.	Consulting		No	Simapro	
28				LEGEP Software GmbH	Consulting		No	LEGEP	
29				Boustead Consulting Limited	Consulting		No	The Bousted Model	No
30				UK Environment Agency	Government				No
31	html		Both	Delft University	Academic			IDEMAT	No
32	xls	2010	LCIA	University of Bath	Academic	No			No
33	xls	2008	LCIA	Alex Loijos	Academic	No	No		No
34		2010	LCIA	Sustainable Minds	Consulting	Yes	No	Sustainable Minds	No
35		n/a		USDA	Government		No		No
36	xls		LCI	European Aluminium Association	Industry Association		No		No
37	html,pdf		LCIA	Deutsches Kupferinstitut	Industry Association	No	No		No
38	pdf	2009	Both	European Federation of Corrugated Board Manufacturers (FEFCO)	Industry Association	Yes	No		No
39				International Iron and Steel Institute (IISI)	Industry Association		No		No
40	html,pdf	2003	LCI	Nickel Institute	Industry Association	No	No		No

APPENDIX B. FORMULATION OF THE SURVEY QUESTIONS

The questions were formed to reflect neutrality to all respondents, whether they represented the public or private sector. For instance, preferences for specific LCA file formats were not queried, as that would note an overall preference in the field for a particular proprietary format. The online survey was designed to be short and make efficient use of respondent time.

Questions consist of four areas: demographic questions, information sharing, data and databases (IT and ICT), and LCA practices. Survey questions were formulated to address the following areas of research, each of which contributes to an understanding of information sharing: demographic questions, information ownership, pro-social attitudes, organizational norms, perceived information value, personal cost to share, data and database IT, and LCA practices. Information ownership, pro-social attitudes, organizational norms, perceived information value, and personal cost to share all can help gauge the potential for information sharing within a group, and so are all subjects on which questions were be developed. The inclusion of questions on data and databases and LCA practices helps to establish existing norms and an understanding of the needs of the LCA community. This also can be used to help understand which issues could be addressed to provide “use value” to the community. Attitudes on information sharing became a subset topic, but the survey questions as a whole help understand both the needs and possibility for change within the community.

B.1.1 Demographic Survey

There are several demographic questions that can provide useful categorical divisors for other questions. Geographic region of residence can be telling as it can relate to cultural differences, etc. The following questions were asked:

Q1. Choose your region of residence:
Q2. Choose your country of residence (optional):

It is also useful to know about respondent experience with the LCA field and LCA data. Length of experience, role, and sector all inform the perspective, needs and attitudes toward LCA data. The following questions were asked:

Q3. How long have you been in the LCA field?
- 1-2 years
- 3-5 years
- 5-10 years
- 10+ years

Q4. How many LCA studies have you been involved with in your experience to date?
- 0-10
- 11 to 25
- 26 to 50
- 50+

Q5. What role(s) would best describe your interaction with LCA data to date?
- Decision-maker
- User of data
- User of software and databases
- Generator of software and databases
- Provider of data
- Other (please specify)

Q6. What sector best represents your work in LCA?
- Business - Corporate
- Business - Consultancy
- Government
- Non-profit/Civil Society
- Academic
- Other (please specify)

Q7. What level of economic activity does your organization operate within (optional)?
- primary (raw materials extraction)

secondary (manufacturing, processing, and construction)
tertiary (service and retail)
quaternary (office work, etc)
quinary (decision-making)

B.1.2 Information Sharing

B.1.2.1 Information Ownership

Kolekofski (2003, p. 2) defines information ownership attitudes as “*the tendency to treat information as a personally owned resource vice a corporate resource*.” Information ownership has a significant role to play in the intention to share knowledge in organizations.

Since our study community is made up of many distinct organizations and the information in question (LCA data) is sensitive to issues of ownership, an understanding of legal ownership must be inquired about. Most inquiries into information have been done within an organization, for which legality is not often an issue. Within a multi-organization research environment, information ownership becomes a legal as well as an attitudinal issue (Nash, 1993). The legal situation overshadows a researcher's attitude towards information sharing. A legal arrangement restricting information sharing renders personal attitudes towards information sharing practically moot. From a legal perspective there are many possible stakeholders that may be assigned legal ownership of research data. A researcher's parent institution, a private or public funding source, a private sector research partner, or the researchers themselves may have legal ownership. In order to inform us as to the legal state of research data we will ask the question below.

Q13: In your current and previous LCA research, who owns (owned) the resulting data? Do you feel that you have some ownership over your data?

B.1.2.2 Pro-social Attitudes

Existing pro-social attitudes were gauged using the following questions:

Q15: Do you conduct LCA studies in collaboration with others in the LCA field?

There are at least two anti-social attitudes that are specific to the academic community. Getting “scooped” is a particular concern of researchers, in which it is revealed that other researchers have already published identical research to another researcher's ongoing project, diminishing their ability to receive credit in the academic community. In discussions over resistance to open data and open access in academic communities, being “scooped” was listed as a major barrier to sharing data; researchers worried that revealing data for an initial paper may inspire other research groups to build on that research and scoop the original researcher's future research plans. It would be useful to know how many researchers in the LCA community share this concern. Mis-use of data, general risks of sharing data, and the possibility of incurring criticism are also listed as a concern in open data literature, and therefore were also included. The barrier of time-commitment was chosen because it is a common barrier in most voluntary actions. A lack of perception of benefit is also a common barrier in most voluntary actions, so it was also included in the form of the statement ‘I don't really see a benefit to others and to the field if I share my data’. A lack of knowledge on how to share data is a valid possibility for a barrier, so it also is included. The following question was presented as semi-close-ended because while these barriers are anticipated to be common indicators, respondents may have alternative barriers that should also be counted:

Q17: Select barriers to information sharing that applies to you (select all that apply):
I worry about the risks of sharing my data

I think the time-commitment for sharing data is probably too high
I worry that my data may be mis-used
I don't really see a benefit to me in sharing my data
I don't really see a benefit to others and to the field if I share my data
If I release my data, I may be scooped on future publications based on this data.
I don't want to open my data to criticism
I don't know how to share my data

The second anti-social attitude specific to open data is the concern of researchers that peers may examine their data critically, possibly leading to a demotion or invalidation of the value of the data and any research papers based on it. The following question was asked to gain insight into participants' concerns regarding their data being critiqued. It was posed as an open-ended question to invite respondents to explain the issues that struck them as important.

Q14: Do you receive critiques and comments on your LCA data/studies? How? Are you satisfied with existing avenues for soliciting feedback?

B.1.2.3 Organizational Norms

Constant et al. (1994: p. 5) defines a norm as "*shared beliefs about behaviors that people ordinarily do and behaviors that are right*." Constant et al. (1994) also found that organizational norms that support the belief that information is owned by the organization rather than the employee encourages information sharing. We can extend that easily to academic communities. Open access and open data movements have been pushing to normalize information sharing within academic communities, which has achieved limited success. For instance, the success of the open access movement can be gauged on 20 per cent of the papers published annually are open access (Hitchcock, 2004). Gul et al. (2010) found that open access is more prevalent in the sciences rather than the social sciences, so there may also be some established norm for sharing in the LCA field.

In order to gauge existing norms for information sharing in the LCA community, we asked the following questions:

> Q19: Do you share your data with others? Have others shared their data with you? Why or why not?

The following question was posed to understand current practices pertaining to publishing data: 1) how many members of the LCA community publish data at all (whether publishing detailed data is a norm) 2) if so, in what form is the data published. Understanding current practices will help us understand which are already utilized.

> Q10: Do you publish your data anywhere? If so, where? If not, why?

While this can give us an indication of what is being utilized, it is hard to gauge receptivity to potential changes or alternatives to publishing data. Users that are not publishing data are likely to have compelling reasons for not doing so, and may therefore may not be as receptive as other respondents to future changes that encourage data sharing.

As there is a special interest in new methods of collaboration like OSS communities, the following question was included to assess attitudes and norms relating to OSS concepts. The question was posed as closed statements because it is efficient for respondents and it is easy to extract quantified summary results.

> Q20: Check which apply:
> I use LCA databases that are free & open
> I use open source LCA software
> I am aware of open LCA databases
> I am aware of open source LCA software
> I am not aware of open LCA databases
> I am not aware of open source LCA software

Q22: Do you perceive benefits from the use of open source, open data, and linked data approaches to LCA?

B.1.2.4 Perceived Information Value

Jarvenpaa et al. (2000) specified that perceived information value is both an issue of quality and accessibility. Accessibility is a key factor in this project, as it is with all open projects. However, that could be considered a function of the design of a solution (in our case, our LCA information sharing media). We choose to then focus on gauging the perceived information value of the LCA data itself, which in this case could be both a quality and relevancy issue. Therefore, the participants were asked if they identified with the following statements (which were included in a large question about attitudes).

Q18: Select attitudes toward information sharing that apply to you (Select all that apply):
I think others will benefit from having access to my data
I think access to LCA data contributes to the quality and relevancy of academic data
I would like greater access to LCA data
I want to share my LCA data
I would like to learn about how to share my LCA data

B.1.2.5 Personal Cost to Share

Constant et al. (1994) notes that a possible barrier to information sharing may be the personal cost to share as it can take time and other resources. Therefore, the participants were asked if they identified with the following statement, which was included in a previous question that included other barriers to sharing:

Q: I think the time commitment to share my data is too high.

B.1.3 Data and Databases (IT and ICT)

Even if the intention to share is present, it is essential that tools and processes reduce friction and help turn intention into action. Therefore, we must gauge if there are any present barriers to the technology used and what those barriers are. It is anticipated that the identified technology barriers will be addressed in order to optimally facilitate the sharing of data.

The question was structured to allow for multiple answers and to allow for quick identification of some expected common answers, while also allowing respondents to provide alternative answers. While the information would have been valuable, file formats specific to LCA were not listed in order to retain a neutral stance within the LCA community between approaches. This question was posed as semi-closed because the options for storing LCA data are limited, but respondents may use uncommon or surprising options.

> Q12: How do you store your LCA data?
> Spreadsheet file
> LCA-specific file format
> a text or rich-text format
> Other (please specify)

Respondents were also asked open questions about their satisfaction and difficulties with existing LCA IT and ICT: data, databases, and data formats. The questions were posed as open because these are exploratory questions with no previous established criticism. Also, since this this is an opinion/experience based question, possible answers may be influenced by suggestions.

> Q8. Are you satisfied with existing LCA software and datasets? What do you feel are the gaps in existing software and datasets?
>
> Q9. Are you satisfied with the LCA file formats you use? What features drove your choice of data format? Do they contain sufficient indicators of data quality, detail, and metadata? Please explain.

B.1.4 LCA Practices

There are two prescribed parts of the LCA process in the ISO LCA standards that are of particular interest. The first part of the process of particular interest is the critical review stage. The critical review, the solicitation from other practitioners of feedback on data and methods, is ideal but not always done. It would be useful to know if this process, which requires data sharing and community interaction, is practiced regularly and effectively. It was posed as an open-ended question to be receptive to respondent insights.

> Q14. Do you receive critiques and comments on your LCA data/studies? How? Are you satisfied with existing avenues for soliciting feedback?

The second part of the LCA process is the iterative re-release of LCA studies. This practice acknowledges the imprecise, estimative nature of the LCA process. By editing and re-releasing, practitioners make better methodological and data decisions and reduce uncertainty, providing increasingly better. Also, as technology and other time-related factors change and are factored into new results, a re-released study remains useful. This occurs after a LCA study is published. It is desirable to know whether this part, which is essentially voluntary re-publishing, is practiced regularly. It was posed as an open-ended question to be receptive to respondent insights.

> Q16. Do you revise and re-release LCA studies (akin to iterative research prescribed by ISO 14040)? Why or why not?

B.1.5 General

Since this is an initial exploration, some general questions were suggested to leave opportunity for answers that did not fit with the information sharing models used to structure the questionnaire.

> Q23: Are there any issues on LCA data and data quality that you would like to discuss? Please elaborate.