The Quebec Life Cycle Inventory Database Project

Using the ecoinvent database to generate, review, integrate, and host regional LCI data

Published in: The International Journal of Life Cycle Assessment
First online: 30 May 2013
In issue: Due to be released

Authors: Lesage, Pascal
          Samson, Réjean

Contact ecoinvent: ecoinvent
                   Technoparkstrasse 1
                   8005 Zurich, Switzerland
                   support@ecoinvent.org

Main title: The Quebec Life Cycle Inventory Database Project

Sub-title: Using the ecoinvent Database to Generate, Review, Integrate and Host Regional LCI Data

Authors:
Pascal Lesage*, CIRAIG, École Polytechnique de Montréal, Département de génie chimique, C.P. 6079, succ. Centre-Ville, Montréal (Qc) Canada, H3C 3A7, pascal.lesage@polymtl.ca, Tel: +1.514.434.5620, Fax: +1.514.340.5913
Réjean Samson, CIRAIG, École Polytechnique de Montréal, Département de génie chimique, C.P. 6079, succ. Centre-Ville, Montréal (Qc) Canada, H3C 3A7

Abstract:

Purpose LCA in Quebec (Canada) is increasingly important. Yet, studies often still need to rely on foreign LCI data. The Quebec government invested in the creation of a Quebec LCI database. The approach is to work as an ecoinvent “National Database Initiative” (NDI), whereby the Quebec database initiative uses and contributes to the ecoinvent database. The paper clarifies the relationship between ecoinvent and the Quebec NDI and provides details on prioritization and data collection.

Methods The first steps were to select a partner database provider, and to work out the modalities of the partnership. The main criterion for partner selection was database transparency, i.e. availability of unit process data (gate-to-gate), necessary for database adaptation. This and other criteria, such as free access to external reviewers, conservation of dataset copyright, seamless embedding of datasets and overall database sophistication, pointed to ecoinvent. Once started, the NDI project proceeded as follows: (1) data collection was prioritized based on several criteria; (2) some datasets were “recontextualized”, i.e. existing datasets were duplicated and relocated in Quebec and linked to datasets representing regional suppliers, where relevant; (3) new datasets were created; and (4) Canadian environmentally extended supply-use tables were created for the ecoinvent IO repository.

Results and discussion Prioritization identified 500 candidate datasets for recontextualization, based on the relative importance of relative contribution of direct electricity consumption to cradle-to-gate impacts, and 12 key sectors from which about 450 data adaptation or collection projects were singled out. Data collection and private sector solicitation is underway. Private sector participation is highly variable. A number of communication tools have been elaborated
and a solicitation team formed to palliate this obstacle. The new ecoinvent database protocol (Weidema et al., 2011) increases the amount of information that is required to create a dataset, which can lengthen or, in extreme cases, impede dataset creation. However, this new information is required for the new database functionalities (e.g. providing multiple system models based on the same unit process data and regionalized LCA).

**Conclusion.** Being an NDI is advantageous for the Quebec LCI database project on multiple levels. By conserving dataset copyright, the NDI remains free to spawn or support other LCI databases. Embedding datasets in ecoinvent enables the generation of LCI results from “day 1”. The costs of IT infrastructure and data review are null. For these reasons, and because every NDI improves the global representativity of ecoinvent, we recommend other regional or national database projects work as NDIs.

**Keywords:** Quebec (Canada); LCI database; Regionalization; National database initiative; ecoinvent v3; Electricity mix
1. Introduction

1.1 Context

Quebec is the largest, by land mass, province of Canada (over 1.5 million km²) and the second most populous (over 8 million in 2012). Its economy is relatively large (were it a country, it would be 49th in terms of GDP and 27th in terms of GDP per capita [ISQ, 2013]) and diverse, with activities ranging from resource extraction and agriculture to aerospace and information technology. While a Canadian province, Quebec is distinct from the rest of Canada in many ways: culturally (it is predominantly French speaking and has been officially named “a nation within a united Canada” in a Canadian House of Commons motion), energetically (its electricity is over 96% from hydropower), etc. In the context of this journal, one distinction that is important is in the rapid importance, since the turn of the millennium, given to LCA. Quebec is home to one of the largest university research centers dedicated exclusively to LCA (Estrela, 2011), LCA underlies three of the sixteen principles of the Sustainable Development Act (MDDEP, 2006), LCA consultants and in-house LCA capacity is rapidly expanding, and the Quebec government is now piloting a large-scale carbon footprinting initiative (MDEIE, 2012). Despite this, Quebec LCI data is sparse and largely unavailable outside the organizations that generated it. This means that, compared to regions for which LCI databases exist, important resources need to be spent to create or adapt datasets and that additional uncertainty for the majority of datasets further upstream is largely inevitable. To remediate this, the Quebec government decided to fund a three-year Quebec LCI database development project.

The project is overseen by two ministries. The first is the Ministry of Sustainable Development, Environment and Parks, which is involved because of the environmental nature of the project. The second is the Ministry of Economic Development, Innovation and Exports, which sees in an LCI database a means for Quebec companies to better capture the advantage conferred by Quebec’s low carbon electricity grid mix, and hence to better distinguish themselves in markets where carbon footprints are considered. The project was conferred to the CIRAIG, an LCA research centre based in Montreal.

From the onset, it was determined that the project should not strive to build an LCI database from scratch, but should rather build upon and work with an existing LCI database. There were four main reasons for this. First, the costs and time associated with creating a new database were well beyond those available for the present project. Second, it made
possible the partial adaptation of datasets whereby only the most important flows were modified to account for the Quebec context without sacrificing completeness by ignoring other flows. Third, it evacuated the need to develop the database infrastructure (data formats, servers, database protocol, etc.). Finally, it allowed all new datasets to be directly integrated in a database, i.e. connected to upstream processes, meaning that full LCIs could be calculated from the get-go.

This paper explains why, both technically and contractually, ecoinvent was chosen as the partner database to develop a Quebec LCI database. It then outlines how the work to be carried out over a three year period was prioritized, and how work is progressing, stressing obstacles and the mechanisms that were put in place to overcome them.

2. Methods

2.1 Database selection

The key criterion for database selection was transparency, i.e. availability of datasets at a unit process (or gate-to-gate) level. The advantages of making data available at a unit process level in LCI databases have been discussed elsewhere (see e.g. UNEP-SETAC 2011 and Frischknetch, 2004). Of particular importance for the elaboration of a regional LCI database were:

* the availability of gate-to-gate data on exchanges. These unit process datasets can be used as initial templates for data collection. Sensitivity and contribution analyses can also help focus data collection on environmentally significant flows, potentially reducing the burden on both data collectors and solicited industry. Other flows can simply keep their initial (i.e. non-regionalized) value.

* the ability for changes in one unit process to be accounted for in the accumulated LCI of all downstream datasets, meaning that every time a dataset is regionalized, the LCI results of downstream datasets are also, to varying degrees, regionalized. This allows for the targeting of a few strategic modifications, such as adapting the electricity grid mix and waste management datasets, to increase the regional relevance of a great number of datasets.

Two databases were identified that met these criteria: US LCI and ecoinvent. Ultimately, the ecoinvent database was chosen for a number of reasons. Table 1 lists the technical advantages of ecoinvent. Non-technical aspects also were
important in making the decision. In its effort to globalise its database, the ecoinvent Center elaborated a contractual framework for collaborations with so-called National Database Initiatives (NDI). Basically, in exchange for a right to use NDI-generated datasets, the ecoinvent Center embeds the datasets in its global database and provides the NDI with technical support, dataset review and revenue sharing. The NDI also needs to provide environmentally-extended supply-use tables for its region (see Section 2.3.3). The use of ecoinvent software infrastructure to create, store and manipulate datasets ensured the start-up cost of the regional database would be low, and the free review process lowered running costs.

Since only a “right to use” to the developed gate-to-gate datasets are granted to ecoinvent, and not the associated intellectual property, it is possible for an NDI to freely publish its datasets independently of ecoinvent integrate the datasets in other existing databases and, eventually, in a stand-alone regional (or Canadian) database. By default, the cradle-to-gate inventory does not belong to the NDI since they are built on the background data of ecoinvent: however, a provision exists in the contract whereby even such LCI can be published by the NDI if a sufficient portion of the background impacts are associated with other NDI datasets.

Finally, the fact that ecoinvent is not dependent on fluctuating government funding and that it has a solid user base made it a seemingly stable foundation on which to build the regional database.
Table 1: Technical characteristics of the ecoinvent database that consolidated the decision to use it as partner for the creation of a Quebec LCI database

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality and coherence</td>
<td>A strong database protocol helps ensure coherence across datasets. Errors in datasets are minimized by internal checks, quality assurance and the use of unit process datasets by thousands of users for years.</td>
</tr>
<tr>
<td>Coverage</td>
<td>The database covers many of the core industrial sectors required to conduct LCA and consistently accounts for a substantial list of elementary flows.</td>
</tr>
<tr>
<td>Globalisation facilitation</td>
<td>ecoinvent v3 facilitates the inclusion of regional data via new features such as the ability to geolocalize datasets and to create “child” datasets that inherit all flows from the corresponding global average parent dataset unless otherwise specified.</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>Every flow in the database is qualified via common data quality indicators, which can then be translated to uncertainty. This is done using a pedigree matrix approach based on Weidema and Wesnaes (1996). This allows the adaptation work to identify low quality data that would profit from additional data collection as well as a mechanism to account for the reduction of uncertainty associated with regionalizing datasets.</td>
</tr>
<tr>
<td>Sophistication in v3</td>
<td>Some of the changes in version 3 of ecoinvent were directly in line with the CIRAIG’s research interests, notably the ability to use the same unit process datasets in multiple system models and the geolocalizing of inventory making the use of regionalized impact assessment methods possible.</td>
</tr>
</tbody>
</table>
2.2 Prioritization

Given the relatively short time frame of the project, data collection work was prioritized based on diverse criteria. No formal process was used to consider all the criteria: instead, priorities were set with government officials in light of the information generated by the analyses discussed below.

2.2.1 Prioritizing recontextualization

In some cases, simply specifying that a transforming activity is occurring in a given location can, in ecoinvent v3, significantly influence its associated accumulated LCI. This occurs when specifying the location changes how the unit process links to other datasets. We refer to the process of creating a regional (Quebec) version of an existing dataset in order to allow it to relink to other Quebec-specific datasets as “recontextualizing” (see Section 2.3.1).

In theory, recontextualization of any unit process dataset will help improve the regional relevance of its cumulated LCI as soon as at least one of its inputs or outputs is traded on a regional market. Recontextualization is however a partial and possibly insufficient approach since it does nothing to address the actual differences “within the gates” between the global average and the regional transforming activity. It misses key issues such as employed technology, energy efficiency, emission abatement and fuel mix. Recontextualization should therefore be seen as a “first pass” at regionalization, albeit a potentially very useful first pass.

One case where it is assumed to indeed be useful is for datasets where electricity consumption is environmentally significant. Because hydroelectricity represents over 96% of the electricity produced in Quebec, recontextualizing these unit processes can significantly improve the regional relevance of the associated accumulated LCI with minimal effort.

Initial candidates for recontextualization were identified by calculating the relative contribution of electricity consumption to cradle-to-gate impacts for any ecoinvent 2.2 unit process dataset that is not itself producing or transmitting electricity (3593 datasets). Only electricity directly consumed by the unit processes was considered. Datasets earmarked for recontextualization were those for which the contribution of electricity consumption was over 25% of total cradle-to-gate impacts for at least one of the four IMPACT 2002+ endpoint indicators.
2.2.2 Prioritizing sectors using environmental metrics

Three approaches were used to evaluate the environmental significance of industrial sectors.

The first approach used the Canadian National Pollutant Release Inventory (NPRI), Canada's legislated and publicly accessible inventory of emissions (Environment Canada, 2012). Only emissions for Quebec companies in 2010, and only impacts on human health as calculated using IMPACT 2002+ endpoint characterisation factors, were considered.

The other two approaches aimed to identify sectors that are key in the context of LCA rather than in absolute terms. For both, the 2002 Canadian Input-Output model produced by the University of Toronto (Bjorn et al., 2005) and hosted by Carnegie Mellon University (www.eiolca.net/) was used. One approach, referred to as “life cycle impacts of first tier inputs” (see Figure 1a for a visual description), considers the cradle-to-gate impacts of all goods and services directly referred to in a final demand vector. The second (Figure 1b), referred to as “sector contribution analysis”, calculates the contribution of a given sector to the life cycle impacts of the total goods and services referred to in the same final demand vector. Four scaling vectors were successively used for each approach: Quebec personal consumption, Quebec domestic demand, Quebec exports and re-exports, and Quebec gross production. The last is devoid of any physical sense as a final demand vector but was included as a weak proxy for the probability an LCA will be conducted of a given sector. For both approaches and the four final demand vectors, the sectors were compared on their contribution to climate change and human health impacts as calculated with IMPACT 2002+. Since the IO model represents the Canadian economy, which is actually quite different than the Quebec economy, results were fraught with uncertainty and the interpretation phase was delicate.
Figure 1: Difference between two IO-based prioritization approaches based on a simplified 4 sector model

In both figures 1a and 1b, letters (“A”, “B” and so on) represent different sectors, Figure 1a represents the “life cycle impacts of first tier inputs” approach. The impacts of sector A are those associated with the cradle-to-gate impacts of producing the amount of A specified in the final demand vector. Figure 1b represents the “sector contribution analysis” approach. The impacts of sector A are those associated with every instance of sector A, irrespective of the supply chain it is part of.
2.2.3 Other considerations for prioritizing sectors

Two other types of considerations were used in prioritizing work on the Quebec LCI database. The first was government policy: sectors that were directly targeted by government environmental policy for which life cycle information was judged to be relevant were added to the list of priority sectors. The second is research requirement: one sector (water treatment) was added to the priority list to compliment LCA research in water use impacts conducted at the CIRAIG (Boulay et al., 2011).

2.3 Data collection and adaptation

2.3.1 Recontextualization

Recontextualizing refers to the process of creating Quebec versions of existing datasets to relink them to other Quebec-specific datasets. Since the default market in ecoinvent v3 is global, recontextualization will only result in changes if (1) one or more of the unit processes’ intermediate exchanges are traded on non-global markets or (2) markets are effectively by-passed via a direct link between transforming activities, which is justified when the “regional” unit process is so closely linked to one of its suppliers that fluctuations in the former’s demand will directly be reflected in the supplier’s output (Weidema et al., 2011).

In a first pass, recontextualizing only focuses on datasets where direct electricity inputs are environmentally significant (see section 2.2.1). Future recontextualization may focus on waste management, harvested pulpwood and some construction materials, all of which are traded on non-global markets.

Once a unit process has been identified as a candidate for recontextualization, the following steps must be taken: (1) ensure the activity actually exists in the region of interest; (2) ensure there either exists a non-global market for one of the upstream inputs or a unit process which will be directly linked to; (3) create, using ecoEditor for version 3 (the free tool provided by ecoinvent Centre to encode and submit gate-to-gate datasets), a regional version of the Global dataset either by using a parent-child relationship (i.e. using an existing dataset with a global scope as a dynamic template) or by creating an entirely new dataset; (4) in ecoEditor, make changes to the activity’s annual production volume, the geographical correlation uncertainty score for exchanges, where relevant, and to the dataset description in appropriate fields; and (5) submit the dataset to ecoinvent via ecoEditor.
2.3.2 Data adaptation and new dataset creation

Although recontextualization can be useful, the core of the database project was the production of unit process datasets that represent the Quebec context. The general procedure for data generation is: onsite data collection (if relevant), horizontal aggregation (if relevant), internal data validation, data entry via the ecoEditor for version 3 software tool, and review by the ecoinvent editorial board.

A number of parallel approaches are used to generate these datasets. The first is to harvest datasets that have already been produced or are in the process of being produced by LCA practitioners in Quebec. In some cases, these have led to alliances (e.g. with the Quebec Wood Product Life Cycle Inventory initiative and the Carbon Footprint Quebec Pilot Project) or overarching agreements (e.g. with LCA consultants).

The second approach is to obtain new data directly from industry. While some industries, such as primary aluminium producers, understood the strategic advantage of being represented in an international LCI database and hence readily participated, it generally has been an uphill struggle to obtain private sector collaboration. Refusal can be attributed to lack of general awareness on the strategic value of life cycle information, fear that data will be misused, and lack of time. A number of tools to counter these obstacles have since been developed, such as documentation on the value of the project (website, brochures and presentations), boilerplate agreements concerning data confidentiality, presentations explaining horizontal and partial-vertical aggregation, simple and complete data collection templates, largely based on existing ecoinvent datasets, and a streamlined cradle-to-gate LCA report for the participating company based on the data that was collected. A team of solicitors was also put together to contact more companies than was possible when it was a task shared with actual data collection and other LCA work.

Finally, publically available data are used whenever relevant.

2.3.3 Supply-use tables

The NDI agreement signed with ecoinvent stipulated that the CIRAIG was to provide Canadian environmentally-extended supply-use tables to the ecoinvent Center. These were compiled using publically available data, including the supply-use tables and GHG emissions from Statistics Canada and pollutant emission data from the National
Pollutant Release Inventory. For now these data will be held in the ecoinvent “IO repository” and will serve as a basis for its hybrid model if and when the ecoinvent Centre decides to create one.

3. Results

3.1 Prioritization

The prioritization for recontextualization identified 587 unit processes for which direct electricity consumption contributed at least 25% of the life cycle impacts (see Figure 2). Figure 3 presents for climate change the percentage reduction in impacts resulting from a relinking of this first-tier electricity input to the Quebec grid mix.

![Figure 2: Identification of initial candidates for recontextualization based on the relative contribution of first-tier electricity consumption](image-url)
Figure 3: Share of life cycle climate change impacts attributable to first-tier electricity consumption in initial candidates for recontextualization and proportional impacts of same datasets when relocated in Quebec

The sector-level analyses resulted in a list of 12 priority sectors, presented in Table 2. Within each sector, individual unit processes and flows were identified, although data availability is ultimately the main guiding principle. A total of 450 new or adapted datasets were identified for potential inclusion in the database by the end of the project, with about 40% of these being of high priority.

Table 2: List of priority sectors and justification for their inclusion in the list
<table>
<thead>
<tr>
<th>Sector</th>
<th>Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture and food transformation</td>
<td>Agriculture: Both IO-based analyses. Food transformation: Data availability.</td>
</tr>
<tr>
<td>Basic chemicals</td>
<td>IO-based sector contribution analysis</td>
</tr>
<tr>
<td>Biofuels</td>
<td>Support for government policy</td>
</tr>
<tr>
<td>Energy vectors, heat and electricity</td>
<td>LCA experience, support for government policy, both IO-based analyses</td>
</tr>
<tr>
<td>Fertilizers and phytosanitary products</td>
<td>Both IO-based analyses</td>
</tr>
<tr>
<td>Mines and metals</td>
<td>Direct emissions, IO-based sector contribution analysis</td>
</tr>
<tr>
<td>Non-metallic mineral products</td>
<td>Direct emissions, IO-based sector contribution analysis</td>
</tr>
<tr>
<td>Pulp and paper</td>
<td>Direct emissions, economic importance</td>
</tr>
<tr>
<td>Road transport</td>
<td>Support for government policy, omnipresence in LCA</td>
</tr>
<tr>
<td>Waste management</td>
<td>Support for government policy, omnipresence in LCA</td>
</tr>
<tr>
<td>Water treatment</td>
<td>LCA research needs</td>
</tr>
<tr>
<td>Wood and wood products</td>
<td>Direct emissions, IO-based sector contribution analysis</td>
</tr>
</tbody>
</table>

3.2 Data collection

**Recontextualization:** Of the 587 datasets identified as initial candidates for recontextualization, only about 25 of the original candidate datasets (less than 5%) have for the moment been recontextualized. While some were dropped because they represented activities not occurring in Quebec, most are on hold because obtaining relevant Quebec production volumes was not straightforward. One difficulty is that some reference flows are difficult to express in terms of “production volume” occurring annually in a region (e.g. meters of soldering). For other reference flows, the available data was often of a different geographical scope (e.g. Canadian rather than Quebec) or aggregation (e.g. sector level rather than product level). The ecoinvent Centre has, in personal communications, recently downplayed the need for high quality absolute data on production volumes, indicating relative figures would be
acceptable, and defined various forms of acceptable proxy data for production volume, such as foreign production volumes scaled by GDP. This will allow the creation of a greater number of recontextualized datasets.

**New data collection:** While a handful of datasets have traversed all the steps from finding data sources to integration of reviewed data in the master database, many data collection projects are still in early data collection stages, or are stalled for lack of data supplier. The creation of a dedicated solicitation team has allowed us to make promising advancements.

4. Conclusions and recommendations

The CIRAIG is both supplier and user of the ecoinvent database. As users, the new functionalities of ecoinvent v3 are extremely promising and largely in line with the CIRAIG research program. As suppliers, some of the requirements can generate significant amounts of additional work. However, the ecoinvent Center has since the beginning proven to be quick to help with technical difficulties (e.g. database protocol interpretation, unit process modeling, ecoEditor bugs, etc.) and open to being challenged. The ecoEditor software, though sometimes a source of frustration due to speed issues and the inevitable bugs associated with new functions or software, has been an invaluable tool in the creation of ecoinvent v3 compliant ecoSpold 2 datasets. The CIRAIG, via the LCI database project, is co-funding the programming of some additional ecoEditor functionalities that will help data suppliers and database users.

More significant, however was the fact that the Quebec database is built on and contributed to an already populated database. The first advantage was the use, for a majority of our data collection and development projects, of existing ecoinvent gate-to-gate datasets. This allowed us, among other things, to efficiently produce data collection questionnaires, to identify flows that needed regionalization in priority, and to use “default” values for flows for which we were not able to obtain information (e.g. data on infrastructure). The second advantage is that the intermediate flows can normally be linked to existing datasets, allowing us to avoid introducing datagaps in the database (such as one can find in the US LCI with the use of “dummy processes” to account for intermediate flows with no corresponding transforming activities).

The benefits of submitting datasets to ecoinvent far outweigh the effort: existing gate-to-gate datasets allowed us to be efficient, and all services associated with review, integration and dissemination of datasets are provided for free.
What is more, given that the intellectual property of the gate-to-gate datasets remains with the NDI, the reviewed datasets can be used to constitute a stand-alone national database or be integrated in other third party databases.

The fact that the scope of the LCI database is focused on Quebec is associated with some limitations: while distinct, the Quebec economy is not

The government funding for the Quebec LCI database project will come to an end in December 2013. The CIRAIG is evaluating many scenarios to keep the database work alive beyond that date. At a minimum, all datasets will be made freely available. Efforts will be made to have long-term resources to maintain and update datasets that were submitted during the project as well as to continue supporting companies wishing to to submit data to the database. A more promising option would be to extend the scope of the project to a Canadian or North American level.

As a final recommendation to other groups wishing to develop a regional database: rapidly allocate dedicated resources to private sector solicitation, but at the same time start identifying alternative data sources. Consolidating private sector participation has proven to be both a very slow and uncertain process: companies are often long to convince and, once onboard, can pull out at any moment. Identifying secondary data sources as a failsafe approach for key sectors of flows is therefore important, and can also provide values to which the primary data can be compared as a first internal review step.

Acknowledgements

The authors would like to acknowledge the support of the two Québec ministries involved in the project (Ministère du développement durable, de l’environnement et des parcs and the Ministère du développement économique, de l’innovation et de l’exportation), all CIRAIG employees involved in the LCI database project and all our external collaborators who have provided data and insights.

References


