NIST Technical Note
NIST TN 2353

Environmental Product Declaration (EPD) Database Creation and Public Life Cycle Inventory (LCI) Data Gap Assessment

Sindhu Ranganath
Gonzalo Rodriguez-Garcia
Joshua D. Kneifel
Purav Parekh

This publication is available free of charge from:
https://doi.org/10.6028/NIST.TN.2353

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

NIST Technical Note
NIST TN 2353

Environmental Product Declaration (EPD) Database Creation and Public Life Cycle Inventory (LCI) Data Gap Assessment

Joshua Kneifel
Applied Economics Office
Engineering Laboratory

Purav Parekh
Applied Economics Office
Engineering Laboratory
Department of Computer Science
George Washington University

Sindhu Ranganath
Gonzalo Rodriguez-Garcia
Applied Economics Office
Engineering Laboratory
Department of Engineering Management
and Systems Engineering
George Washington University

This publication is available free of charge from:
https://doi.org/10.6028/NIST.TN.2353

September 2025

U.S. Department of Commerce
Howard Lutnick, Secretary

National Institute of Standards and Technology
Craig Burkhardt, Acting Under Secretary of Commerce for Standards and Technology and Acting NIST Director

NIST TN 2353
September 2025

NIST Technical Series Policies

Publication History

Approved by the NIST Editorial Review Board on 2025-09-17

How to Cite this NIST Technical Series Publication

Ranganath, S., Rodriguez-Garcia, G., Kneifel, J., and Parekh, P. (2025) Environmental Product Declaration (EPD) Database Creation and Public Life Cycle Inventory (LCI) Data Gap Assessment. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Technical Note (TN) NIST TN 2353. https://doi.org/10.6028/NIST.TN.2353

Author ORCID iDs

Sindhu Ranganath: 0009-0005-8689-4336
Gonzalo Rodriguez-Garcia: 0000-0001-8084-8236
Joshua Kneifel: 0000-0002-3114-5531
Purav Parekh: 0009-0009-3250-5306

Contact Information

joshua.kneifel@nist.gov

Abstract

Given the increasing prevalence of using Environmental Product Declarations (EPDs) for product comparisons and selection, both domestically and globally, there is a need to ensure the results reported in EPDs are useful for such comparisons (i.e., transparent and standardized to ensure quality decision making). One challenge that was identified more than a decade ago was the lack of common data sources, which could undermine the comparability of EPDs and similar claims. Federal agencies have targeted this issue through the development of public secondary datasets and gap assessments of currently available public life cycle inventory (LCI) datasets.

This study complements public data gap assessments completed by the Environmental Protection Agency (EPA), Department of Energy (DOE), and NIST by focusing on three goals. First, develop and implement a repeatable process to create a machine-readable database of information reported in EPDs for a product category. Second, analyze the database to review reported data sources for various processes in all EPDs within the product category to determine which data sources are currently in use and assess whether public sources can replace commercial data sources while identifying any gaps in public LCI data. Third, use the results of the LCI data analysis to provide insights on developing representative inventories for common practice in existing EPDs for a product category. This study provides results for four product categories: cement, concrete, asphalt, and gypsum.

Keywords

Building materials; energy; environmental product declarations; life cycle assessment; life cycle inventory; representative inventory; resources; sustainability

Table of Contents

List of Tables

List of Figures

Preface

Given the increasing prevalence of using EPDs for product comparisons and selection, both domestically and globally, there is a need to ensure the results reported in EPDs are useful for such comparisons. One challenge for these comparisons is the lack of common public data sources for use in developing the LCA models used for estimating the results reported in EPDs. This study creates machine-readable databases of information reported in EPDs for four product categories (cement, concrete, asphalt, and gypsum) and uses those databases to complete a public LCA data gap assessment. Additionally, the LCA data analysis provides insights on developing representative inventories for common practice in existing EPDs for a product category.

Disclaimers

Certain commercial products are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose

Acknowledgments

The authors wish to thank all those who contributed ideas and suggestions for this study, including Dr. David Webb and Dr. Dave Butry of EL's Applied Economics Office, Dr. Pam Chu of NIST's Material Measurement Laboratory, and Dr. Li Piin Sung of EL's Infrastructure Materials Group. Thanks go to our colleagues at ERG and EPA for providing feedback on the study. A special thanks to Dr. Jonathan Broyles at the University of Colorado – Boulder for providing the concrete EPD database and feedback on the concrete-related analysis.

Author Contributions

Sindhu Ranganath: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Validation, Visualization, Writing- Original draft preparation.

Gonzalo Rodriguez-Garcia: Conceptualization, Methodology, Investigation, Data curation, Validation, Writing- Original draft preparation.

Joshua Kneifel: Funding acquisition, Resources, Conceptualization, Supervision, Project administration, Writing- Reviewing and Editing.

Purav Parekh: Data curation, Software

1. Introduction and Scope

There has been a steadily growing interest from consumers, industry, local and state jurisdictions in the U.S., and nation-states globally for improved reporting of the environmental and human health impacts associated with product purchases, particularly for construction-related materials [1–5]. A common approach to quantify these impacts is life cycle assessment (LCA), which provides a scientific methodology for calculating the impacts of a product or service over its entire life cycle in accordance with the International Organization for Standardization (ISO) 14040 and ISO 14044 standards [6, 7]. While LCA is the quantification tool, environmental statements are the communication mechanism for providing decision-makers with information about a product's environmental impacts. ISO has established fundamental principles and requirements for various types of environmental statements in ISO 14020 [8], including Environmental Product Declarations (EPDs), whose requirements are further developed in ISO 14025 [9].

An EPD is a standardized, third-party verified document that provides LCA-based information, as well as additional information on the environmental aspects of products [9]. EPD programs are often built for specific product categories, groups of products capable of fulfilling equivalent functions that might require EPDs to be consistent with a distinct Product Category Rule (PCR) in addition to the previously mentioned ISO standards. However, PCRs are not universal, and more than one PCR might exist for a given product category. Efforts to harmonize the development and use of PCRs include ISO 14027 on the development of PCRs and ISO 14029 on mutual recognition of EPDs and footprint communication programs [10, 11], as well as the PCR Open Standard from the American Center for Life Cycle Assessment (ACLCA) [12].

Using EPDs for documenting environmental and human health impacts of products has been common (and growing) since their introduction into building rating systems [2]. Additionally, EPDs are being used as the basis for product selection by building construction companies and building owners, as well as a range of U.S. jurisdictions and states [3]. Specifically, these programs require third-party verified EPDs compliant with ISO 14025 and ISO 21930 [13] standards for life-cycle stages A1-A3, known as “cradle-to-gate” because it includes impacts raw material extraction (A1), raw material transport (A2), and product manufacturing (A3), but excludes any impact after the product leaves the manufacturing site or “gate.” Along with domestic demand for products with EPDs, global demand is also growing as EPDs are increasingly being required in U.S. export markets [5].

Given the increasing prevalence of using EPDs for product comparisons and selection, both domestically and globally, there is a need to ensure the results reported in EPDs are useful for such comparisons. One challenge that was identified more than a decade ago was the lack of common data sources, which could undermine the comparability of EPDs and similar product claims [14]. The lack of identical assessment methodologies can also further hinder these efforts [15]. Some PCRs (e.g, concrete [16]) prescribe the use of secondary data sources to facilitate standardization and comparability of the EPDs. However, secondary data sources are still lacking, and those available are often commercial products and not necessarily representative of U.S. industrial practices.

Federal agencies in the United States have formalized collaboration to improve LCA secondary data through an interagency initiative, the Federal LCA Commons (FLCAC) [17]. Activities of the FLCAC include providing support to enhance standardization, measurement, reporting, and verification of LCA modeling. These activities will assist the industry in improving the transparency, trustworthiness, and comparability of results reported in EPDs, improving their competitiveness in the domestic and global marketplace. The Environmental Protection Agency (EPA), Department of Energy (DOE), National Renewable Energy Laboratory (NREL), and the National Institute of Standards and Technology (NIST) have already undertaken separate, but complementary activities to identify secondary LCA data needs for EPD development with a specific focus on embodied carbon of construction materials.

The EPA has provided several resources for improving the state of PCRs and EPDs developed by industry [18–21]], including Life Cycle Inventory Data Gap Assessment (DGA) that focused on identifying "life cycle inventory (LCI) data gaps in free-to-use and publicly accessible secondary LCI datasets" for large volume building materials. The report was developed by reviewing existing PCRs and communicating with PCR Committee members for a variety of high-priority construction materials to collect insights from industry and experts in the LCA field. The collected information was synthesized into a list of recommendations for prioritizing datasets to develop and fill public LCA data gaps in currently available public data sources (i.e., FLCAC mega-repository).

NREL has completed a structural path analysis (SPA) to identify hotspots for seven construction product categories, including concrete, asphalt, steel, and glass [22]. The analysis used the U.S. Environmentally Extended Input-Output (USEEIO) database [23] to identify data gaps in construction-related industries as a whole. It also included a comparison to identify consistencies and variations between the EPA data gap assessment and the underlying methodology.

NIST published a public data gap assessment complementary to those completed by the EPA and NREL by providing a more in-depth approach to identify and quantify the impacts of public data gaps that can be implemented across any product category to assist in prioritizing data gaps to address [24]. NIST developed a robust framework to identify and quantify public LCA data gaps, identified and quantified the impact of each data gap across each product category, and provided qualitative rankings of data gaps within and across the product categories considered for a variety of building materials.

NIST TN 2338 [24] is a complementary document to this study, with products selected to provide variability in the level of prescriptiveness for LCI data source selection and reporting within a PCR, and format variability in published EPDs for a given product category. This variability allows for both validating the framework in NIST TN 2338 while providing variability in published EPD information for this study. Additionally, the LCI for cement developed from the EPD database analysis in this study was used in NIST TN 2338.

The objectives of this study are threefold. First, develop and implement a repeatable (and, hopefully, automated) process to create a machine-readable database of information reported in EPDs for a product category. Second, analyze the machine-readable database to review reported data sources for various processes in all EPDs within the product category. Determine which data sources are currently in use and assess whether public sources can replace commercial ones while identifying any gaps in public LCI data. Third, use the results of the LCI data analysis to provide insights on developing representative inventories for common practice in existing EPDs for a product category.[1]

This study will provide results for four product categories: cement, concrete, asphalt, and gypsum. These categories were selected based on three factors:

(1) Complementary to previous efforts to evaluate public data gaps for construction materials, including those by the EPA, NREL, and NIST.

(2) Alignment with existing NIST programs and projects related to construction materials.

(3) Consider product categories with varying reporting formats and quantity of publicly available EPDs to assess the ability to automate EPD database creation across product categories.

Along with a summary of the findings, there will also be a discussion on potential next steps and future research building on this study and leveraging the associated published EPD databases.

[1] The representative inventories developed in this study have been used to assist NIST TN 2338 [24].

2. Study Methodology

The workflow for the methodology in this study is shown in Figure 1. The steps are as follows:

(1) Collect all publicly available EPDs for a product category.
(2) Convert the relevant/necessary EPD information into a machine-readable format.
(3) Analyze data sources for each process in the LCA model for each EPD.
(4) Aggregate the results for all EPDs to determine which data sources are used for each input into the product model.
(5) Use aggregated results to identify public data usage and prioritize public data gaps in the FLCAC. Where appropriate, data sources used were also compared to PCR guidance and requirements to determine whether PCRs are being followed consistently.

Each of these steps is discussed in more detail in the remainder of Section 2.

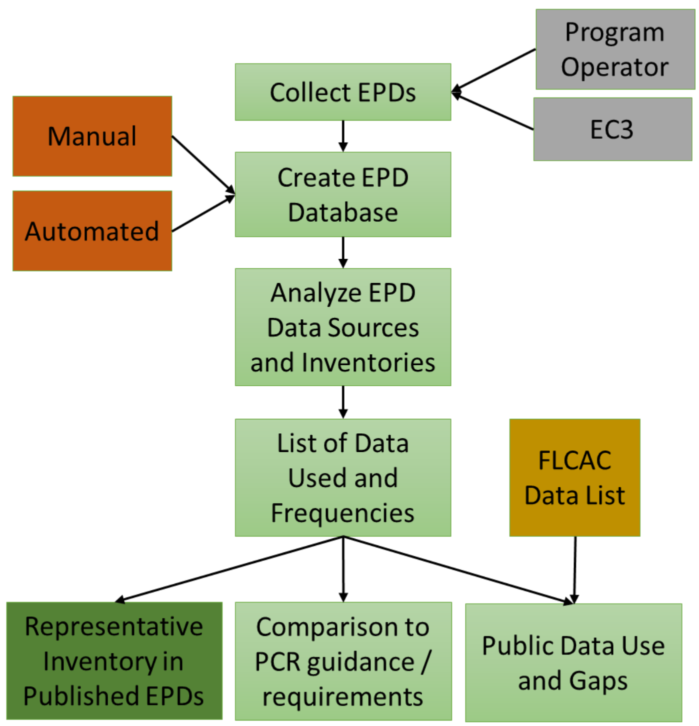

Figure 1 Study Methodology Workflow

Note: EC3 = Embodied Carbon in Construction Calculator; FLCAC = Federal LCA Commons

2.1. EPD Collection and Review

EPDs are available in varying quantities and formats depending on the product category, which makes the EPD collection process different for each product category. Each process will be discussed in detail in Section 2.2. The collected EPDs were primarily reviewed to understand the model inputs, including the database and software used for the LCA modeling. Additionally, modeling parameters (e.g., system boundary, allocation, cut-off), LCA data gaps (if specified in the EPDs), impact assessment results, and other administrative information were gathered and reviewed. PCRs were also reviewed for each product category to understand the prerequisite framework required to model the product under study. The collection process for each product category was dependent on the accessibility of the EPDs and the website on which the EPDs were hosted. For each of the product categories analyzed (i.e., concrete, cement, asphalt, and gypsum), a different methodology was needed, each of which is described in the remainder of this section.

2.1.1. Download Existing Digitized EPD Database

For concrete, a database (CSV format) of EPDs was created and published in Broyles. et al in November 2023, containing approximately 30 000 concrete EPDs of varying compressive strengths from the United States [25]. This database was unique to concrete and not available for any other product categories. The authors [25], also updated the database with an additional 16 750 EPDs by April 2025, and this updated database, with approximately 46 750 concrete EPDs, was used for the current analysis.

2.1.2. Collect EPDs from Program Operator

For all other product categories for which an EPD database was not already available, EPDs were downloaded directly from the Program Operator (PO) website(s) where feasible. This method was used to collect asphalt EPDs from the National Asphalt and Pavement Association (NAPA). The Emerald Ecolabel website, NAPA's EPD generator tool [26], organizes asphalt EPDs by state. This repository is regularly updated, and thus, all EPDs available as of January 2025 were gathered. A Google Chrome plugin called "DownThemAll" [27] was used to bulk download these EPDs, in the HTML format, the native format used by NAPA to create EPDs.

Although most POs host EPDs on their websites, their accessibility (especially bulk access capabilities) varies by PO and product category. For example, the POs for concrete products provide access to EPDs through a web tool that allows for downloading one EPD at a time. National Ready Mix Concrete Association (NRMCA) and ASTM International were contacted and confirmed that their websites currently lack bulk download capabilities.

2.1.3. Download from Embodied Construction Carbon Calculator (EC3)

For product categories without bulk access (i.e., cement and gypsum), Building Transparency's Embodied Construction Carbon Calculator (EC3) tool [28] was used to access currently valid EPDs. EC3 hosts various product categories for EPDs valid across different world regions, with EPDs available in their original format (e.g., PDF, HTML) and/or as openEPD JSON[2] files. Two approaches were used to access EPDs in the EC3 tool: (1) manual downloads for product categories with minimal EPDs, and (2) automated calls to the EC3 API for product categories for which the manual process would be too inefficient and time-consuming.

[2] openEPD is "an open data format for passing digital third-party verified EPDs among POs, EPD Databases, LCA tools, design tools, reporting, and procurement" [35].

The manual approach included applying filters within the EC3 tool to access region-specific and product-specific EPDs, ensuring that the selected EPDs were specific to the United States and valid as of the download date. Unless specified, no additional performance specification filters were applied. This methodology of accessing EPDs was applied to product categories that had fewer than 50 valid EPDs (i.e., cement). The end products retrieved were product-specific EPD files in their native format for the respective product categories.

Automated calls were made to the EC3 web API to collect the openEPD files for each EPD in a product category with enough EPDs to make manual downloading inefficient (i.e., gypsum). EC3's API allows for the filtered (e.g., region, product specificity, performance characteristics) download of the EC3-formatted information file (i.e., JSON file) for every product category. However, downloading the EPD source file (published PDF or HTML file) and/or the openEPD file must be completed for one individual EPD per API call. Therefore, the following automated process was implemented for bulk download of openEPD files:

1. Make an API call replicating a manual filtered search in the EC3 interface to identify valid EPDs for a given product category that are applicable in the U.S.
2. Extract unique identifiers (UUIDs) for each EPD in that filtered search
3. Create an API request to obtain the openEPD file by looping through each UUID

This approach has been implemented for obtaining the openEPD files for gypsum EPDs. Appendix A provides more details on the supplemental documents that provide the specific API call and the associated code that automates the process.

Bulk openEPD JSON file downloads were completed for product-specific EPDs for the U.S. that were valid as of the date of download. These JSON files were converted into CSV files for ease of use for further analysis. This approach was implemented for gypsum, but several issues were identified with this approach, and thus, the gypsum EPDs (PDF files) were also downloaded manually to ensure no EPDs were excluded and necessary information for the analysis in this study was available. openEPD is discussed further in Section 2.2 on database creation. There were various challenges faced while making the automated requests to the EC3 API. First, due to minimal documentation at the time of the requests, it was difficult to identify the correct API endpoints and the appropriate format to add the associated filters (query parameters). Once the endpoints and query parameters were identified, a script was developed to automate the API requests for each openEPD with the extracted UUIDs. However, multiple requests to the API using a loop were failing after a certain timeframe due to the "throttling" safeguard that was implemented in the EC3 API. Throttling is added to an API to prevent system overload by limiting the number of requests a user can make within a specific timeframe. To address this issue, a "delay" was added after several continuous API requests. This resolved the issue, and the extracted data from openEPDs was saved in two formats – a csv file and a JSON file, for further analysis.

2.2. Database creation

Current EPDs are typically available in PDF format (except for asphalt, which is in HTML). While PDFs offer several advantages (i.e., human-readable, common information format), the main challenge for this study is that they are not necessarily machine-readable, which is required to complete quantitative analysis of reported information in EPDs. Thus, an approach was developed to convert EPDs from PDFs to a machine-readable (CSV format) database for each product category.

The method applied for data extraction is dependent on product category-specific factors, such as the number of EPDs, consistency in reporting and formatting in EPDs (i.e., standardized versus non-standardized information), and availability of EPDs in their native format (i.e., PDF or HTML). Figure 2 provides a workflow diagram of the process to generate an EPD database dependent on those factors. Four different information format categorizations for EPDs are considered in this study: standardized HTML, standardized PDF, Non-Standardized PDF, and openEPD JSON.

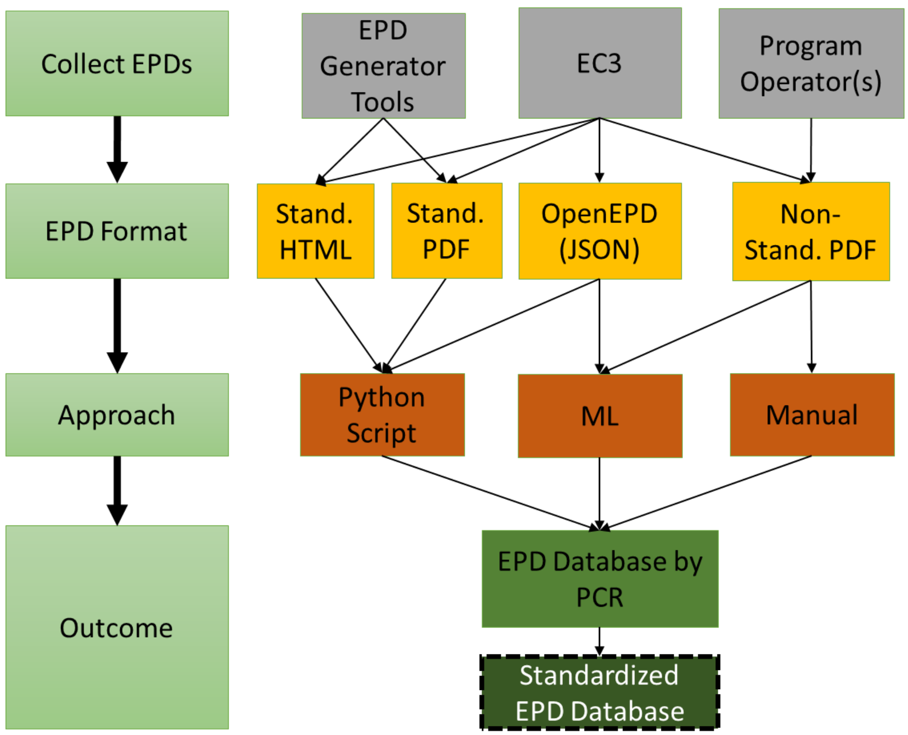

Figure 2 EPD Database Creation Methodology

Based on the information formats, there are three methods that could be used to extract the data from EPDs: manual, programming scripts, and machine learning (ML) / artificial intelligence (AI). Note that a combination of these methods may be applied as appropriate. Note that AI was not implemented in this study, but potential use cases will be mentioned in this document when referencing potential future research.

The result of the data extraction is a machine-readable database (CSV format) that provides data for selected fields reported in the source EPDs. Due to the varying information reported by the EPDs, these databases are not currently standardized across product categories. Building on this work, a future goal is to standardize across product categories to provide one format.

The manual data extraction process was specifically used in the initial stages of database creation. Product categories for which this technique was implemented had fewer than 50 EPDs (e.g., cement). Relevant sections of the EPDs were manually copied from the original EPD documents into a spreadsheet to create the CSV database. The chosen products had fewer EPDs, but the formatting and typesetting of these EPDs varied significantly by PO. The advantage of manually creating the database was the ability to carefully parse through the information in each EPD and select only relevant information, informing the development of an automated extraction process for other product categories. However, the downside was that this process was time-consuming, and it was difficult to validate the CSV database. Thus, an automated process is preferred for future database creation and updating.

Programming scripts (written in Python) were available from [25], which was originally developed to create a database for concrete EPDs. The database used for the current assessment is a product of [25] as-is except for a few manual edits. It is also assumed that the LCI data source for different processes found in the database is reported in the underlying EPD; a validation of the database against the underlying EPD was not conducted. The manual edits made were for the LCI data source information on processes that were determined to be coding errors. A few examples of the manual edits are below:

- Data source for a particular process was placed in the wrong column (e.g., the production process of slag cement was reported under a transport process, one column off)
- Data source represented two processes but was reported under only one of the processes (e.g., 50/50 fly ash slag mix only reported under fly ash)
- Minor text string differences (e.g., extra spaces)
- Uncategorized processes reported in “Any other item” that fall into a specific process or processes (e.g., 50/50 fly ash slag mix, rock dust).

These data sources were moved manually to the appropriate columns.

The “hard-coded” scripts for concrete were used as a proof of concept for another product category with standardized HTML files (i.e., asphalt). “Hard-coding” using programming scripts involves extracting specific information from predetermined lines within the PDF or HTML files. For example, the tenth line on the first page of an asphalt EPD contains specific information (e.g, declaration number). A Python script was written to extract this specific information from the tenth line of the EPD. This method was applied to extract information from each relevant line of an EPD to make a machine-readable database that has consistent formatting of reported information. The key advantage of hard-coding is that it automates the data extraction process while allowing precise selection of the required information for the database. However, the drawbacks of this method are that a considerable amount of time is spent on writing the scripts, albeit less than manually creating the database; it is also highly specific to a single product category and is only applicable to completely standardized EPD formats. Each product category typically has different PCR requirements that create variation in layout, design, and content reported. Even within a product category, there is often a lack of consistency in the reporting and formatting of EPDs. Additionally, some manual effort is needed to identify the essential information for which the hard-coding should be tailored to collect the data.

For categories that do not have consistency in reporting information and formatting of the EPD (i.e., non-standardized format; gypsum), this study used the openEPD JSON available for the respective EPDs as a standardized reference document to extract the essential information. The openEPD JSON was converted to CSV and transferred into a file for all EPDs in the product category (i.e., gypsum board). The most significant drawback of this approach is that not all information reported in an EPD is provided in the openEPD files. For the purposes of this study, some of the missing information was necessary to complete the analysis, specifically, the material composition of the inputs to the product and any LCA-specific data source information reported in the EPD. Due to the non-standardized format, it was infeasible to write hard-coded scripts that accurately and efficiently extracted this missing information. Instead, a hybrid approach of ML with scripts was leveraged to read and extract the necessary information from the non-standardized EPDs (i.e., gypsum EPD PDFs). This information was appended to the CSV-formatted EPD database built using the openEPD JSON.

For this study, Microsoft Azure AI's Document Intelligence[3] was used to parse through the PDF, which returned a JSON of the extracted information. The required "missing information" on material content was then further extracted using a JavaScript script and appended to the CSV file. The associated JavaScript code is discussed in Appendix A. This "off-the-shelf" hybrid approach was effective for most gypsum EPDs to automatically extract the information, which could then be saved to a machine-readable format (CSV). However, it was not effective for EPDs representing more than one product because the table structure and required data manipulation varied too significantly. Thus, the multi-product EPD data extraction was completed manually.

The approach used for this study was successful in extracting the necessary data into a machine-readable format. Some post-processing through scripting was necessary to combine with the data extracted through the hybrid "off-the-shelf" approach and then append to the CSV database generated from the openEPD file.

Due to time and labor constraints, a manual approach was implemented to collect the data source information in addition to the previous methods of database creation mentioned. Gypsum EPDs also provided additional information, such as LCA system boundary, database disclosed, primary data disclosure, representativeness, software disclosed, and so on, that were not available on the openEPD JSON files. Thus, the relevant information was manually copied from EPD PDFs to the csv files.

Future research could consider leveraging AI (e.g., large language models - LLMs) by uploading the individual EPD PDF files and using prompt engineering techniques to extract the information needed. Appendix A provides some prompts that were developed for, but not implemented in, this study that may be useful for future work in applying prompt engineering techniques to replicate the manual process. Although outside the scope of this study, this initial application of LLMs for gypsum has provided significant lessons learned and a template for developing a fully automated process leveraging LLMs. The manual processes could be automated through more

[3] Disclaimer: Certain commercial products are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

complex techniques such as fine-tuning and retrieval augmented generation (RAG) using the generated databases and the source EPD files, or a combination of these methods. Additional details on the proposed approach will be discussed further in Section 4.4 (Future Research). A fully automated ML/AI approach would be appealing because the PDF structure of EPDs varies widely across and within product categories.

The resulting CSV databases for the different products include common data fields such as declaration number, PCR used, impact assessment results, and other foundational information. These fields are agnostic to the product category and are essential for creating a third-party verified Type III environmental declaration.

Fields unique to each product category, along with any additional columns created to provide insights specifically about the product category, are listed under the respective product categories in the following sections. The level of disclosure requirements varies by product category, which may lead to variations in disclosure both across product categories (i.e., EPDs in a category may consistently provide more information than another category) and within product categories (i.e., some EPDs may provide more information than others within the same category). One such area of disclosure variation is documentation of LCI data sources.

2.3. Database analysis

Once the CSV database for the EPDs of a product category has been created, it can be used to analyze a wide range of research questions. This study will focus on the process inventories and LCI data sources for a given product category to:

- Quantify the frequency of processes appearing in published EPDs.
- Determine whether PCR requirements (if available), were followed in data source usage and modeling.
- Complete a public LCI data gap assessment: summarizing and analyzing the reported LCI data sources in the EPD and providing insights into public LCI data gaps
- Develop an inventory that is representative of the process flows reported in published EPDs.

Additional analysis that was completed outside of this scope may also be briefly summarized in Section 4.4 if deemed relevant for discussion to highlight other potential use cases for the EPD databases.

2.4. Semantics

The focus of this study is to analyze the data sources used in the LCA modeling reported in valid EPDs. In the study, databases used to model different processes, as well as processes for which public data is used, are determined to inform FLCAC about the data to be developed or updated to address these gaps. A variety of LCA databases were identified across the EPD databases, with varying levels of use for each across the four product categories: concrete, cement, asphalt, and gypsum. The selection of specific databases for a specific process can be driven by a range of factors:

- ecoinvent (EI) – a proprietary (i.e., commercial) database [30]
- United States Life Cycle Inventory (USLCI) [31]– a publicly available, free LCA database maintained by the FLCAC mega repository [32]
- Datasmart LCI, under the name US-EI – a proprietary database focused on North America, combining original data together with data from USLCI v.1.60 and a modified version v.2.2 to be representative of U.S. operations [33].
- GaBi Thinkstep – a proprietary (i.e., commercial) database [34]

A few other databases appear minimally, including GREET [35] and USEEIO [23]

In addition to LCA databases, EPDs also use other data sources listed below:

- Facility-specific data
- Manufacture-specific LCA report or a third-party verified type III EPD
- Data provided directly by industry, which is not available in currently published LCA databases (i.e., industry average data).
- LCA reports published in peer-reviewed journal articles.

Relevant flows include services, energy, waste to be disposed of, and material outputs. Of these, energy, material, and service flows are collectively referred to as products or product flows on openLCA. Products/product flows and wastes are either produced or treated by processes (also referred to as providers on openLCA). Data source is an umbrella term used to refer to any data used in LCA modeling, which could be either processes from an LCA database or non-LCA data like EPDs. A dataset is how a database provides a particular process. It contains resources from nature and emissions to nature, required to "produce" that product, "provide" that service, or "treat" that waste. This study focuses on identifying gaps in the availability of production, combustion, treatment, and service process data that provide services, energy, waste disposal, and material output.

In this study, data gaps refer to "processes" or "background datasets" that are currently unavailable in the public database. In some instances, the environmentally relevant flows (inputs in the form of raw materials, fuel, transportation, etc.) are available in a public database, but the database lacks a process that treats and/or generates that flow. In USLCI, such flows are identified as "cutoff flows" (a classification specific to the USLCI database). For example, the input "explosives, at the plant" is listed in the database, but the corresponding process that describes the resources used and emissions generated during its production or usage is missing. Other times, both the process and the environmentally relevant product/waste flow are absent from the database. For example, EI includes the raw material "clinker" that is provided by multiple processes (e.g., "market for clinker" and "clinker production"). USCLI does not currently have "clinker," the flow, or a production process for clinker. These two examples highlight the two reasons for a process data gap:

(1) Absence of processes representing the production, service, or treatment of a cutoff flow present in USLCI.

(2) Absence of both the relevant flow and the process producing/treating the flow.

The data gap assessment in this study is specifically aligned with the processes prescribed or recommended in the PCR. For instance, the concrete PCR prescribes using U.S. averages from the USLCI database for both combination and single-unit truck transport, covering long and short hauls with gas or diesel power. Based on these prescriptions, there are no identified data gaps for service processes that provide for truck transportation within the scope of the concrete PCR. However, this alignment with the PCR also means that any processes not prescribed in the PCR are not considered for the data gap assessment. For example, regional differences in transport may be significant enough that the absence of geographically specific processes could constitute a data gap but would not be identified by this study.

The remainder of this study will discuss the findings from the data source gaps analysis of the EPD databases for each of the four product categories considered: concrete, cement, asphalt, and gypsum. The data gap analysis only notes whether a process is available as a public dataset and does not assess the quality of the processes available.

2.5. Data gap definitions

The authors acknowledge that there are various approaches to categorizing public LCI data gaps. For this study, a “public dataset” is a dataset available on the FLCAC mega-repository. Data gaps have been organized to the best ability of the authors into three categories: primary, secondary, and tertiary.[4] There is a “no data gap” categorization, which occurs when a process is available in a public dataset. This categorization holds regardless of whether the public data is always, sometimes, or never reported as being used in EPDs.

A data gap is labeled a “primary” data gap if either (1) a process that could impact the LCA results is not available as a public dataset or (2) there is not enough information to identify the specific unavailable process itself. The latter are considered “unclassifiable” because the “process” reported is named too generically, such that the name could represent multiple processes and cannot be identified.

A data gap is labeled a “secondary” data gap if (1) a process that could impact the LCA results is not available, but a similar production process (i.e., proxy) is available as a public dataset or (2) there exists a public data source that has either been prescribed in a PCR or used in an EPD to account for the upstream impacts, but has not yet been incorporated into a public dataset (e.g., supplier specific data/reports/EPDs).

A data gap is labeled a “tertiary” data gap if the absence of the process does not affect the results of the EPD. In this study, there are product categories where a PCR identifies and/or EPDs use recovered materials as input materials. According to ISO 21930 [13], the use of recovered materials must be declared, but the only impacts to be considered are those from their use and any processing needed to make the recovered material usable. For example, the concrete PCR

[4] Note that the subjective nature of this process led to cases where the authors were not in complete agreement of categorization for specific individual data gaps.

evaluated here indicates that the treatment and transport of recovered materials should be included in the EPDs.

A recovered material identified by the PCR or used by an EPD could be linked to multiple process models that may not be available in a public database. One example is a process that generates a recovered material, likely as waste. If such a production process is not available on a public database, this process is a tertiary data gap because its absence does not affect the result of EPDs under analysis. In addition to a production process, there should also be a process for the final disposal of that waste/recovered material. This is the destination the waste/recovered material would have reached if the waste under assessment had not been used as recovered material (landfill, incineration, etc.). This end-of-life (EOL) disposal process is also considered a tertiary data gap because it acts as an alternative scenario with no effect on the results in the EPD. If the waste is produced within the product system boundary and is not 100 % internally recycled, the producer bears the responsibility for the production and disposal of the waste, and these processes impact the LCA result.

In between the waste production and the waste disposal processes, there may be one or more waste treatment processes that modify the waste and make it suitable for disposal or use as a recovered material in the process under assessment. If treatment is needed for the waste to become usable, and that process is not available as a public dataset, then it is recommended to consider it a primary data gap because it has an impact on the LCA results for the EPD. As mentioned above, the impacts of the treatment of the recovered material must be included in the EPD but might not be reported separately. Thus, waste treatment processes may be reported in this study as a primary data gap because the inputs needed for said treatment process may be included as part of the inventory, but not explicitly reported in an EPD.

3. Findings

This section of the report presents LCI data source findings from the EPD databases analyzed for the four product categories for which the methodology was applied: concrete, cement, asphalt, and gypsum. The challenges encountered were specific to each product category. For each product, the public data availability status is defined per the categorization in Table 1, using a combination of lettering and numbering conventions as needed. Color coding complements the categorization and highlights data gaps throughout the text when identified. A subset of this categorization is implemented for products as appropriate. Table 2 also shows an additional breakdown of data gaps for asphalt, to differentiate between gaps identified by the PCR and additional data gaps identified by the approach implemented in this report.

Table 1 Public Data Status Legend

Legend	Priority / Category
A public production process is available and is used by EPDs/prescribed by the PCR	X
A public production process is available, but not used by EPDs/prescribed by the PCR	Y
Publicly available datasets are used along with proprietary datasets	V
Recovered material identified by the PCR	Z
Input mentioned, but the background dataset used for modeling the process has not been listed by the EPDs	W
Primary data gaps: Processes not available as a public data set	1
Secondary data gap: Prescribed IA or MSP EPD as a data source, but the report is not converted into a public dataset	2
Secondary data gap*: Data source identified as FLCAC equivalent is considered a proxy	2*
Tertiary data gap: Processes whose absence does not affect the results of the EPD	3

Table 2 Asphalt Public Data Status Legend

Legend		Priority
	A public production process is available and is used by EPDs/prescribed by the PCR	
	Proxy process prescribed by the PCR	
	Data gaps – identified by the PCR	
Primary Data Gap	identified as a data gap by the PCR and the report	1
	identified as a data gap per this report's classification, but a data source prescribed in the PCR	
Secondary Data gap	Identified as a data gap by the PCR, but a 'proxy' process has been identified per this report	2*
	Data source prescribed by the PCR, but is classified as a secondary data gap per report, as the prescribed PCR data source is considered a proxy (marked as "**Secondary***")	
	Data source prescribed by the PCR, but classified as a secondary data gap per the report because no model is available in a public LCA database (Marked as "**Secondary**")	2
Tertiary data gap	Recovered material identified by the PCR	3
	Additional recovered material identified by the report	

3.1. Concrete

The POs involved in preparing concrete EPDs are NRMCA and ASTM International, which have implemented a standardized template for EPDs in PDF format. As of April 2025, EC3 provides approximately 74 800 valid product-specific concrete EPDs in the U.S. conforming to the NSF Concrete PCR v2.3 [28]. The filters applied on EC3 to validate the number of US-specific concrete EPDs are:

- 'Jurisdiction – USA'
- 'Validity – As of the date of search'
- 'PCR – NSF international PCR for concrete'

EC3 was used for validation purposes, but as mentioned in Section 2.2, the EPD database for concrete was sourced from Broyles et al. [25], eliminating the need for an active process to download EPDs and create the database. All the EPDs in the concrete product category disclose the data sources used in EPD generation, which can be attributed to the prescriptive nature of the PCR that specifies data sources to be used in LCA modeling.

The data gap assessment for concrete EPDs will be conducted in two steps: (1) assess the public data gaps within the PCR-prescribed background datasets, and (2) assess public data gaps within any other processes reported in EPDs that were not prescribed by the PCR.

3.1.1. PCR Public Data Gap Assessment

The PCR specifies that manufacturer-specific product (MSP) EPDs for raw materials used in concrete manufacturing are the preferred data source for LCA modeling (Section 7.1.9 of NSF Concrete PCR v2.3). In cases where such data is unavailable, Appendix A of the PCR prescribes default data sources that are consistent with ISO 21930:2017, a normative reference document that serves as the foundation for the development of concrete PCR.

3.1.1.1. PCR Prescribed Data

Appendix A of the concrete PCR prescribes 39 data sources and identifies two recovered materials for which no dataset is needed. Table 3 provides an overview of the data sources prescribed in the PCR and identifies whether the prescribed processes are available within public databases. Thus, the table highlights any data gaps for specific processes within the FLCAC.

Table 3 Public Data Status for concrete – Based on the concrete PCR prescribed data sources

	Flow	PCR-prescribed data sources	FLCAC equivalent	Gap Priority	Category
Transport - Land	*Road: Combination Truck*	Long haul, diesel-powered - RNA			X
		Short haul, diesel-powered - RNA			X
		Short haul, gasoline-powered - RNA			X
	Road: Single Truck	Long haul, diesel-powered - RNA			X
		Long haul, gasoline-powered - RNA			X
		Short haul, diesel-powered - RNA			X

	Flow	PCR-prescribed data sources	FLCAC equivalent	Gap Priority	Category
		Short haul, gasoline-powered - RNA			X
	Rail	Train, diesel-powered - US			X
Transport - Water	*Ocean Freighter*	Ocean Freighter, avg fuel mix [a]			X
	Barge	Barge, average fuel mix - US			X
		Barge, diesel-powered"			X
Energy – Elect.	*Electricity generation*	EI - Electricity process by NERC Region	Electricity; at user; consumption mix		Y
Energy - Site energy	*Natural Gas*	USLCI - Natural gas, combusted in industrial equipment			X
	Diesel	USLCI - Diesel, combusted in industrial equipment			X
	Gasoline	USLCI - Gasoline, combusted in equipment			X
	Residual fuel oil	USLCI - Residual fuel oil, combusted in industrial boiler			X
Raw Material - Binder	*PC*	1. MSP EPD [b]		2	
		2. IA EPD	PC, at plant		Y
	BHC	1. MSP EPD [b]		2	
		2. IA EPD [c]		2	
	PBHC	1. MSP EPD [b]		2	
		2. IA EPD [c]		2	
	PLC	1. MSP EPD [b]		2	
		2. IA EPD [c]		2	
	Slag cement	1. MSP EPD [b]		2	
		2. IA EPD [c]		2	
	Fly ash	Recovered material	Prod.: USLCI: Bituminous coal, combusted in industrial boiler USLCI: Electricity, bituminous coal, at power plant		X, Z
			Treatment	1	Z
			Disposal	3	Z
	Silica fume	Recovered material	Production	3	Z
			Treatment	1	Z
			Disposal	3	Z
Raw Materials Aggregates	*Fine agg – natural* / *Coarse aggr. - natural*	EI - Gravel, round| gravel and sand quarry operation [d]	ANL/C&C: Fine & coarse aggregates		Y
	Coarse aggregate – crushed / *Fine aggregate – crushed*	EI - Gravel, crushed production [d]		1	
	Lightweight aggr.	EI - Expanded clay production [d]		1	
	Water	EI – Tap water, at user [d]		1	
	Chemical	*1. IA EPD (EFCA EPD)* [c]		2	

	Flow	PCR-prescribed data sources	FLCAC equivalent	Gap Priority	Category
Raw Materials	*admixtures*	*2. EPD for water-reducing admx [c] or 2. MSP EPD [b]*		2	
Admixtures	*Chemical admixture – plasticizer*	*1. IA EPD (EFCA EPD) [c]*		2	
		2. EPD for water-reducing admx [c] or 2. MSP EPD [b]		2	
	Chemical admixture – coloring	*1. IA EPD (EFCA EPD) [c]*		2	
		2. EPD for water-reducing admx [c] or 2. MSP EPD [b]		2	

Legend	Priority / Category
A public production process is available and is prescribed by the PCR	X
A public production process is available but not prescribed by the PCR	Y
Recovered material identified by the PCR	Z
Primary data gaps: Processes not available as a public data set	1
Secondary data gap: Prescribed IA or MSP EPD as a data source, but the report is not converted into a public dataset	2
Tertiary data gap: Processes whose absence does not affect the results of the EPD	3

[a] Potentially a typing/printing error on the current PCR – The process prescribed in the PCR for 'Ocean freighter' is representative of train transport
[b] Not publicly available. Likely Proprietary data source
[c] Potential inclusion in a public database if not proprietary
[d] Electricity should be substituted by U.S. electricity

Prescribing multiple data source (background LCA dataset) options for an input material, energy source, or transportation service acknowledges the different service processes that may be provided by the datasets and allows for the selection of the most representative option for each EPD. For example, for the transportation service "Road: Combination Truck," the PCR prescribes multiple processes that vary based on the distance traveled (short-haul and long-haul) and the type of engine powering the trucks (gasoline or diesel). Additionally, data source options provide flexibility in cases where MSP EPDs are not currently available.

3.1.1.2. PCR Public Data Gaps

The 39 data sources prescribed by the PCR cover the 28 processes. Six additional processes have been linked to the two recovered materials identified by the PCR, all of which have been listed in Table 4.

Table 4 Processes for which data sources have been prescribed by the concrete PCR

	Processes		Processes
1	*Combination Truck, long haul diesel*	18	Blended hydraulic cement production
2	*Combination Truck, short-haul diesel*	19	Performance-based hydraulic cement production
3	*Combination Truck, Short-haul gasoline*	20	PLC production

	Processes
4	*Single Truck, Long haul diesel*
5	*Single Truck, Long haul gasoline*
6	*Single Truck, Short-haul diesel*
7	*Single Truck, Short-haul gasoline*
8	*Train, diesel-powered*
9	*Ocean Freighter, average fuel-powered*
10	*Barge, diesel-powered*
11	*Barge, average fuel-powered*
12	**Electricity generation**
13	*Natural Gas Combustion in the Industry*
14	*Diesel combustion in the industry*
15	*Gasoline combustion in the industry*
16	*Residual fuel oil combustion industry*
17	**PC production**
21	Slag cement production
22	**Natural (fine and Coarse) aggregate production**
23	Crushed (fine and Coarse) aggregate production
24	Lightweight aggregate production
25	Tap water production
26	Chemical Admixture production [b]
27	Plasticizing admixture production
28	Coloring admixture production
29[a]	*Fly ash production*
30[a]	Fly ash treatment
31[a]	Fly ash disposal
32[a]	Silica fumes production
33[a]	Silica fumes treatment
34[a]	Silica fume disposal

Note: The PCR does not prescribe publicly available datasets for processes in **bold**, while for *italicized* processes PCR prescribes publicly available datasets.

[a] *Additional processes identified per this study's methodology for the recovered material identified by the PCR*
[b] *Although PCR does not prescribe background data sources for every type of admixture used in concrete production, the 'Chemical' admixture will be treated as generic, whose hierarchy of data prescription will be considered when conducting data source usage and data gap analysis for different types of admixtures reported in the concrete PCR.*

Of the 28 processes, the fifteen pertaining to transportation and site energy are publicly available (*italicized* on Table 4). There are three processes available in public databases that are not prescribed by the PCR (highlighted in '**Bold text'** on Table 4):

- **Site electricity usage**
- **Production of natural fine (sand) and natural coarse (gravel) aggregate**
- **Production of Portland cement**

USLCI currently includes production processes for **electricity, sand/gravel,** and **Portland cement**. The PCR prescribes using EI's "Electricity" for the given NERC region, where manufacturing occurs, despite USLCI offering the same level of regionalized data. One potential reason is that USLCI's electricity may be considered outdated because it was based on 2010 data, while EI 3.4's electricity data is from 2015. Recent updates in the USLCI database include the incorporation of the FLCAC Electricity Baseline (eLCI) library based on 2018 data, which is an option for the next PCR (expected in 2025) to prescribe USLCI for electricity production. The production of sand and gravel has also been recently released in a public database on the FLCAC based on the ANL production process for fine & coarse aggregate, which could be considered for inclusion in the prescribed data sources in the next PCR. Portland cement represents a unique case because there exists a production process in USLCI, but the PCR requires the use of MSP EPDs whenever possible and the use of IA EPDs otherwise. The LCI used as the basis of the IA EPD is publicly available and could be used to update the USLCI model, which could then be prescribed by the PCR for usage when an MSP EPD is not available. An example model based on the IA EPD using public data was developed and used for analysis in NIST TN 2338, published as part of a NIST dataset of LCA models for construction materials [24], and will be available from NIST's repository on the FLCAC.

Thus, the eighteen processes associated with concrete are not classified as data gaps (**Bold** and *italicized* text in Table 4).

The PCR also prescribes either MSP EPDs or IA EPDs for production of seven other binders (e.g., alternative cements) and admixtures:

- BHC
- PBHC
- PLC
- Slag cement
- Chemical admixtures
- Chemical admixture – plasticizer
- Chemical admixture – coloring

There are publicly available IA EPDs, along with limited MSP EPDs, for each of these binders and admixtures, while no data is available in a public database. Thus, the seven production processes for the other binders and admixtures are secondary data gaps (Figure 3).

There are no publicly available production processes for production of the following raw materials: ’crushed fine and crushed coarse aggregates,’ ‘lightweight aggregates,’ and ‘tap water’. Thus, these three production processes are classified as primary data gaps.

The PCR identified two wastes (fly ash and silica fumes) used as inputs, which are categorized as recovered materials. Six additional processes have been linked to these two recovered materials per this study’s methodology (Table 4). There are processes in USLCI that account for the production of fly ash (“bituminous coal combustion” and “electricity from bituminous coal,” which labels the output as “Disposal, ash, and flue gas desulfurization sludge, to unspecified reuse”. However, there are no production processes that would generate silica fumes as an output (silicon metal or ferrosilicon alloys). Thus, the ‘production process for silica fume’ is considered a tertiary data gap as it does not impact the results of the LCA. Besides the production of silica fumes, there are no disposal processes for either fly ash or silica fumes; these two disposal processes for the recovered materials—‘disposal of fly ash’ and ‘disposal of silica fumes’ are also considered tertiary data gaps. The potential ‘treatment process turning waste fly ash to usable input material’ or the ‘treatment process of turning silica fumes into usable input material’ for concrete is also missing from the database, and these treatment processes are considered primary data gaps. Treatment processes for fly ash and silica fumes are missing from public databases but should still be included in the EPDs of Ready-Made Concrete per the PCR.

Figure 3 summarizes the data gaps for the 34 processes identified (Table 4). Nineteen processes, including production of fly ash, processes pertaining to energy and transport, and three production processes publicly available but not prescribed by the PCR, are not classified as data gaps. The remaining 15 processes are data gaps, five of which are primary data gaps, including the three production processes of the raw materials that are not publicly available, and the two treatment processes for the recovered materials. Seven secondary process data gaps were identified for the production of binders and admixtures. Finally, there are three tertiary process data gaps— production of silica fumes and disposal of both fly ash and silica fumes.

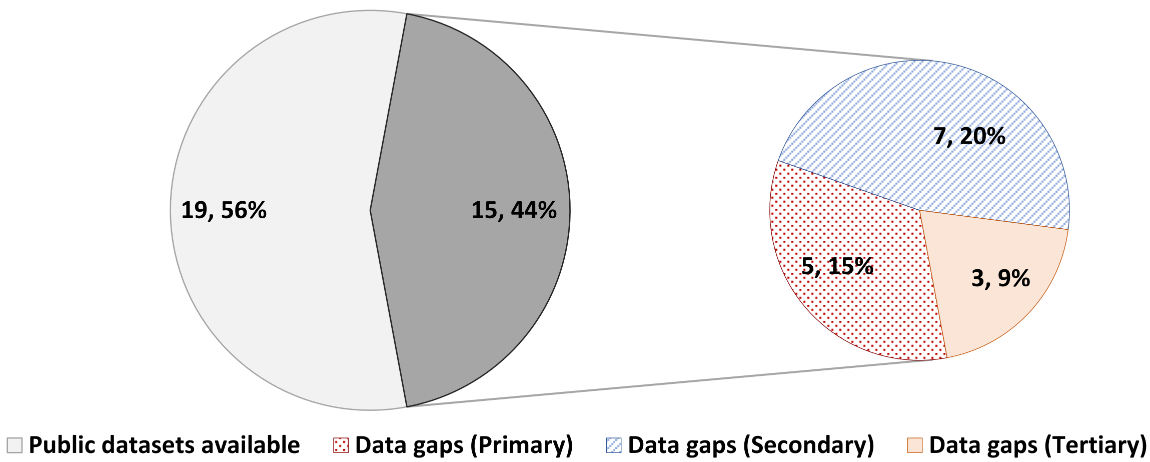

Figure 3 Public Data Gap Analysis for background data prescribed in the concrete PCR

In parallel with addressing public data gaps, strategies for incorporating access to manufacturer-specific data into public databases could also be explored (e.g., attaching a system process to each EPD). These strategies could include mechanisms to maintain intellectual property confidentiality while transparently providing necessary data to EPD users involved in product labeling or procurement decisions.

3.1.2. EPD Data Usage Compared to PCR Data Prescriptions

All the concrete EPDs report the background data sources used for modeling the respective processes, and these processes are listed in Table 5. In all, 47 processes and the usage of three recovered materials have been identified on the concrete EPDs, of which 26 processes have prescribed background datasets in the concrete PCR. These 26 processes are highlighted in **Bold** in Table 5. The three recovered materials usage identified on the EPDs are 'Fly ash,' 'silica fumes,' and 'steel fibers.' Two of these, 'Fly ash' and 'Slag,' are also identified by the PCR (also highlighted in **Bold**), and the six processes linked to these, per this study's approach, have been described in Section 3.1.1. Since steel fibers are not identified by the PCR, additional processes linked to it will be identified and described in Section 3.1.3.

Table 5 Processes reported in concrete EPDs

	Processes		***Processes***
1	**Accelerating Admixture production [a]**	11[e]	Expanded clay production for metakaolin
2	**Air-entraining Admixture production [a]**	26	Microfibers production
3	**Hardening-accelerator admx production [a]**	**27**	**Municipal water production**
4	**Other Admixtures production [a]**	**28**	**Natural Gas combustion in industry [c]**
5	**Plasticizing Admixture production [a]**	29	Non-Hazardous Solid Waste disposal
6	**Retarding Admixture production [a]**	30	Oils/Lubricants/Greases prod.
7	**Set accelerator Admixture production [a]**	31	Carbon black pigment production
8	**Superplasticizer Admixture production [a]**	32	Pigments production
9	**Waterproofing Admixture production [a]**	**33**	**PC production**
10	**Crushed aggregates Production**	**34**	**PLC production**
11	**Lightweight aggregates Production**	**35**	**Train, diesel-powered**

12	**Natural aggregates Production**
13	Other Aggregates production
14	Recycled concrete Aggregate prod.
10 [d]	Crushed aggr. production for a blend of aggr.
12 [d]	Natural aggr. production for a blend of aggr.
15	**Barge, average fuel-powered**
16	Carbon Cure cement production
17	Citric acid production for cleaning chemicals
18	Phosphoric acid prod. for cleaning chemicals
19	CNG HD Short-haul truck Transport
20	CO_2 production
21	**Diesel combustion in industry** [b]
22	Dolomite production
23	**Electricity generation (NERC regions)**
24	Ground Glass Pozzolan production
25	Limestone production
36	Rock dust production
37	HD Combination Truck – Short haul, RD
38	**Ocean Freighter, avg fuel powered**
39	**Slag Cement production**
40	**Combination Truck – long-haul diesel**
41	**Combination Truck – short-haul diesel**
42	**Single Unit Truck – short-haul diesel**
43	**Single Unit Truck – long-haul diesel**
44	On-site solar generation
45	Wash well (water)
46	Propane Production
47	Propane combustion
Rec Mat 1	**Fly Ash**
Rec Mat 2	**Steel Fibers**
Rec Mat 3	Silica Fumes

Note: Processes in **bold** have data sources prescribed by the PCR

[a] *Although PCR does not prescribe background data sources for every type of admixture used in concrete production, the ‘Chemical’ admixture will be treated as generic, whose hierarchy of data prescription will be considered when conducting data source usage and data gap analysis for different types of admixtures reported in the concrete PCR.*
[b] *Of the 45 147 EPDs that have reported the “combustion process of diesel in industry,” 120 EPDs (≈ 0.3 % of the EPDs) use a background dataset from the Datasmart LCI database, and the remaining use a dataset from USLCI to model the combustion process. Although the datasets used to model the combustion processes are from two different databases, they are considered equivalent.*
[c] *Like diesel, about 14 850 EPDs report the combustion of Natural gas, of which about 0.4 % use a background dataset from the Datasmart LCI database, and the remaining EPDs use a background dataset from USLCI. Although these background datasets are from different databases, they are considered equivalent.*
[d] *EPDs reporting ‘Blend of aggregates’ as an input material disclose the usage of two background datasets, one representing the ‘crushed aggregate production’ and the other, ‘Natural aggregate production,’ both from US-EI. These background datasets have also been used to model the production processes of two other raw materials (crushed and natural aggregates). Since the production processes reported are the same across different raw materials, each background dataset appears twice but is counted once.*
[e] *EPDs reporting metakaolin disclose the usage of ‘expanded clay production’ from US-EI as a background dataset. The same background dataset has also been used to model production of lightweight aggregates. Since the production processes are the same across different raw materials, the background dataset appears twice but is counted once.*

The PCR also prescribes the data sources for processes that have not been currently used by any EPDs analyzed for modeling. However, the availability of these data prescriptions may aid any future EPD development that requires the usage of one of these processes. The processes for which background datasets have been prescribed but have not been used in the EPDs analyzed are:

- Combination Truck, Short Haul gasoline
- Single Truck, Long haul gasoline
- Single Truck, Short-haul gasoline
- Production of BHC
- Production of PBHC
- Barge, diesel-powered

- Gasoline combustion in the industry
- Residual fuel oil combustion industry
- Production of coloring admixtures

Table 6 illustrates a comparison between PCR-prescribed background data sources with the data sources used in the EPDs. Processes whose data sources align with PCR prescriptions have been shaded light gray, and the text highlighted in **bold** represents processes using public datasets**.**

To model the production of four processes[5] and the usage of one recovered material, different data sources are used across different EPDs (*italicized* in Table 6). These processes are '*production of other admixtures*,' '*production of PC*,' '*production of PLC*,' and '*production of slag cement*.' The recovered material is '*fly ash*.' The transportation and combustion of energy and fossil fuel processes are predominantly modeled using USLCI datasets. For production of raw materials, datasets from proprietary databases and/or MSP or IA EPDs are mostly used. Three instances have been identified where the data sources used to model the processes have some deviation from the PCR-prescribed data sources. These are the production processes of 'other admixtures,' 'PC,' and 'slag cement'. For admixtures in general, the PCR prescribes using EFCA IA EPD for impacts of producing the respective admixture and, if unavailable for a particular category of admixtures, the PCR prescribes using the EFCA IA EPD for water-reducing admixture or an MSP EPD. In most cases where concrete EPDs disclose "other admixture" as an input material, the EPDs report the impacts of the respective admixture production using the "highest impact in each impact category of the six EFCA admixture EPD," which is not always the additive with the highest impact to all impact categories.[6]

For PC production, the PCR prescribed background data source is MSP EPD whenever possible; if not, the usage of IA EPD has been prescribed. But 0.1 % of the EPDs use both IA EPD and supplier-specific data to model the production of the PC[7]. To model the production of slag cement, all EPDs reporting the process either use IA EPD or supplier-specific data per PCR prescriptions. However, about 30 % of the EPDs that use IA EPD to report the impacts of slag cement production do so without including the A2 impacts.

---

[5] For all four processes that have been modeled using data sources across different EPDs, the PCR prescribes a hierarchy of background data sources

[6] While water resisting admixture has the highest impact in most of the impact categories, there are some impact categories like "Acidification potential, accumulated exceedance ", "Water (user) deprivation potential, deprivation-weighted" among a few others, where other admixtures like retarders, hardening accelerators, etc., have higher A1-A3 impact than water resisting admixtures.

[7] Please note that this statement has been made based on the data source reported in the database; a spot check has not been conducted to see if the respective EPDs use multiple data sources to model the production of Portland cement.

Table 6 Comparing the data sources for the EPD reported process with the data sources prescribed in the concrete EPDs

SL No	***Process***	***Data source used by EPD***	***PCR Prescribed data source***	***# of EPDs using a data source*** [l]	***# of EPDs reporting process*** [l]
1	Accelerating Admixture production [a]	EFCA EPD	1. IA EPD (EFCA EPD)	64	64
			2. EPD for water-reducing admx "or"		
			2. MSP EPD		
2	Air-entraining Admixture production [a]	EFCA EPD	1. IA EPD (EFCA EPD)	17 113	17 113
			2. EPD for water-reducing admx "or"		
			2. MSP EPD		
3	Hardening-accelerator admixture production [a]	EFCA EPD	1. IA EPD (EFCA EPD)	259	259
			2. EPD for water-reducing admx "or"		
			2. MSP EPD		
4	*Other Admixtures production* [a]		1. IA EPD (EFCA EPD)		11 284
			2. EPD for water-reducing admx "or"		
		Supplier-specific data	*2. MSP EPD*	697	
		Highest impact in each impact category of the six EFCA admx EPD		10 587	
5	Plasticizing Admixture production [a]	EFCA EPD	1. IA EPD (EFCA EPD)	29 729	29 729
			2. EPD for water-reducing admx "or"		
			2. MSP EPD		
6	Retarding Admixture production [a]	EFCA EPD	1. IA EPD (EFCA EPD)	7 678	7 678
			2. EPD for water-reducing admx "or"		
			2. MSP EPD		
7	Set accelerator Admixture production [a]	EFCA EPD	1. IA EPD (EFCA EPD)	549	549
			2. EPD for water-reducing admx "or"		
			2. MSP EPD		
8	Superplasticizing Admixture production [a]	EFCA EPD	1. IA EPD (EFCA EPD)	11 013	11 013
			2. EPD for water-reducing admx "or"		
			2. MSP EPD		
9	Waterproofing Admixture production [a]	EFCA EPD	1. IA EPD (EFCA EPD)	198	286
			2. EPD for water-reducing admx "or"	88	
			2. MSP EPD		
10	Crushed aggr. prod	US-EI: Gravel, crushed, at mine [b / m]	EI - Gravel, crushed production [b, m]	23 666 / 1464 [(m)]	23 666
11	Lightweight aggr. Prod.	US-EI: Expanded clay, at plant [c / n]	EI - Expanded clay production [c, n]	724 / 2 [(n)]	724
12	Natural aggregate prod	US-EI: Gravel, round, at mine [d / m]	EI- Gravel, round [d, m]	37 267 / 1464 [(m)]	37 267

SL No	*Process*	*Data source used by EPD*	*PCR Prescribed data source*	*# of EPDs using a data source [l]*	*# of EPDs reporting process [l]*
13	Barge, average fuel	**USLCI: Barge, avg fuel mix**	Barge, average fuel mix	3 957	3 957
14	Diesel combustion in industry	**USLCI/US-EI: Diesel, combusted in industrial equipment [f]**	USLCI: Diesel, combusted in industrial equipment	45 147	45 147
15	Electricity gen. (NERC)	EI: Market for medium voltage	EI: Market for medium voltage	43 818	43 818
16	Water production	US-EI: Tap water, at user [e]	EI – Tap water, at user [e]	42 201	42 201
17	Natural Gas combustion in industry	**USLCI/US-EI: Natural gas, combusted in industrial boiler [g]**	USLCI: Natural gas, combusted in industrial boiler	14 851	14 851
18	*PC Production*	*Supplier-specific data*	*1. MSP EPD*	*13 900*	*400 44*
		PCA, IA EPD	*2. IA EPD*	*26 103*	
		IA EPD + supplier-specific data [i]		*41*	
19	*PLC Production*	*Supplier-specific data*	*1. MSP EPD*	1581	4029
		PCA/CAC, IA EPD	*2. IA EPD*	2448	
20	Diesel train	**USLCI: train, diesel-powered**	Train, diesel-powered	20 895	20 895
21	Ocean Freighter	**USLCI: Ocean Freighter, avg fuel**	Ocean Freighter, avg fuel mix	25 172	25 172
22	*Slag Cement prod.*	*Supplier-specific data*	*1. MSP EPD*	*72*	*11 083*
		MS (CRH) 50% Slag (transportation to the mixing facility included)		*201*	
		SCA IA EPD [j]	*2. IA EPD*	*10 810*	
23	CT – short-haul diesel	**USLCI: CT, short-haul, diesel**	CT, short-haul, diesel	44 854 [k]	44 854
24	CT – long-haul diesel	**USLCI: CT, long-haul, diesel**	CT, long-haul, diesel		
25	SUT – short-haul diesel	**USLCI: SUT – short-haul diesel**	SUT – short-haul diesel		
26	SUT – long-haul diesel	**USLCI: SUT – long-haul diesel**	SUT – long-haul diesel		
Rec. Mat. 1	*Fly Ash*	*byproduct of coal combustion; no upstream manufacturing impacts*	*Recovered materials [h]*	*22 869*	*23 070*
		MS (CRH) 50% Fly Ash (transport to the mixing facility included)		*201*	
Rec. Mat. 2	Silica Fume	byproduct of silicon & ferrosilicon alloy prod.; no upstream manufacturing impacts	Recovered materials [h]	348	348

Note: Processes in **bold** are modeled using public datasets across EPDs, while *italicized* processes are modeled using different datasets across EPDs.

[a] *Although PCR does not prescribe background data sources for every type of admixture used in concrete production, the 'Chemical' admixture will be treated as generic, whose hierarchy of data prescription will be considered when conducting data source usage and data gap analysis for different types of admixtures reported in the concrete PCR.*
[b, c, d, e] *The concrete PCR prescribes using the EI processes substituted with U.S. electricity, and the Datasmart (US-EI) LCI data replaces links in the EI v.2.2 unit processes with U.S.-specific data (largely energy and fuel) to better represent production in the U.S. Thus, the four US-EI processes (distinguished by b, c, d, and e) are considered equivalent to fulfill PCR's requirement* [33].
[f, g] *The two processes, "combustion of diesel in industrial equipment" (f) and "combustion of Natural gas in industrial boiler" (g) from USLCI and Datasmart (US-EI) are considered equivalent.*
[h] *The environmental impacts allocated to these materials are limited to the treatment and transportation required to use as a concrete material input.*
[i] *About 0.1 % of the EPDs use both IA EPD and supplier-specific data to model the production of slag cement.*
[j] *About 30 % of the EPDs using the SCA IA EPD to report the impact of slag cement production do so without the A2 impacts.*
[k] *Most EPDs report multiple transport service processes, for example, some EPDs report the usage of both short and long-haul combination trucks, some report the usage of SUT and combination trucks.*
[l] *Please note that the number of EPDs using a specific data source has been made based on what has been reported in the database; a spot check has not been conducted to verify the database with the actual EPDs.*
[m] *background data is also used to model the production of 'blend of aggregates'*
[n] *background data is also used to model the production of 'metakaolin'*

Since the data sources used by the concrete EPDs to model the 26 processes and the two recovered materials highlighted in Table 6 adhere to the PCR prescriptions, the data gap analysis conducted for the PCR-prescribed background datasets in Section 3.1.1 holds for these processes. The upcoming section will detail the data gap analysis for the remaining 21 processes and the remaining recovered material without background datasets prescribed in the concrete PCR.

3.1.3. EPD Public Data Gap Assessment

The Table 7 lists the data sources and conducts a data gap analysis for processes reported in the concrete EPDs without a PCR-prescribed data source.

Table 7 Data gap assessment for processes without a background data prescription in the concrete PCR

Flow	EPD data sources	FLCAC equivalent	Gap Priority	Category
Other Aggregates	Supplier-specific data		2	
Recycled Concrete Aggregates	1. Custom process, based on US-EI "Gravel, crushed, at mine		1	
	2. Supplier-specific data [f]		2	
Blend of aggregates	US-EI: Gravel, crushed, at mine [a]		1 [(a)]	
	US-EI: Gravel, round, at mine [a]	ANL/C&C: Fine & coarse aggr.		Y
Carbon Cure cement	Supplier-specific data		2	
Ground Glass Pozzolan	Supplier-specific data		2	
Dolomite	US-EI: Dolomite at plant		1	
Limestone	1. US-EI: Limestone, crushed, for mill [g]	USLCI: Limestone, at mine		Y

Flow	EPD data sources	FLCAC equivalent	Gap Priority	Category
	2. Supplier-specific data		2	
Metakaolin	US-EI: Expanded clay, at plant [b]		1 [(b)]	
Microfibers	ELCD: "PP fibres, crude oil based, production mix, at plant, PP granulate, without additives [e]		1	
Steel Fibers	byproduct of steel production; no upstream manufacturing impacts	Prod: Steel; cold rolled coil		X, Z
		Treatment: Primary	1	Z
		Disposal: Tertiary	3	Z
Carbon black pigment	EI: Market for Carbon black		1	
CO_2	Supplier-specific data		2	
Pigments	USEEIO: Synthetic dye and pigment			X
Rock dust	Supplier-specific data		2	
Wash water (well)	Not mentioned [c]		1	W
Oils/Lubricants/Greases	EI: Market for lubricating oil	USLCI: Lubricant feedstock,		
Cleaning chemicals	EI: Market for citric acid		1	
	EI: Market for industrial-grade Phosphoric acid		1	
CNG HD Short-haul truck	GREET: CNG HD truck short Haul, GREET WTW, from waste			X
RD CT HD Short-haul Truck	GREET: HD Truck, combination, short-haul, renewable diesel			X
On-site solar	Not mentioned [c]	US Electricity Baseline: Electricity – SOLAR – (Balance authority)	2*	W
Propane combustion	Production: USLCI/US-EI [d]: Liquefied petroleum gas, at refinery			X
	Combustion: USEPA Emission factor '08	1. Operation of LPG equipment; IA; > 19 kW & < 56 kW Operation of LPG equipment; IA; > 56 kW and < 560 kW 2. USLCI: Liquefied petroleum gas, combusted in industrial boiler		Y
Non-Hazardous SW disposal	US-EI: Disposal of Municipal Solid Waste		1	

Legend	**Priority / Category**
A public production process is available and is used by EPDs	X

A public production process is available, but not used by EPDs	Y
Recovered material identified by the PCR	Z
Input mentioned, but the background dataset used for modeling the process has not been listed by the EPDs	W
Primary data gaps: Processes not available as a public data set	1
Secondary data gap: Prescribed IA or MSP EPD as data source, but the report is not converted into a public dataset	2
Secondary data gap*: Data source identified as FLCAC equivalent is considered a proxy	2*
Tertiary data gap: Processes whose absence does not affect the results of the EPD	3

[a] *The EPD reporting Blend of aggregates discloses the usage of two background datasets: 'crushed aggregate production' and 'Natural aggregate production' from US-EI. The same background datasets have also been used to model the production processes of two other raw materials (crushed and natural aggregates). Since the production processes reported are the same across different raw materials, each background dataset appears twice but is counted once in the EPD data gap analysis (Table 3).*

[b] *The EPDs reporting metakaolin disclose the usage of 'expanded clay production' from US-EI as a background dataset. The same background dataset has also been used to model the production of lightweight aggregates. Since the production processes are the same across different raw materials, the background dataset appears twice but is counted once in the EPD data gap analysis (Table 3).*

[c] *On-site solar and wash water wells have been mentioned as input items used for the production of concrete mixture, but the background data sources used to model their respective processes have not been identified within the EPD database. Regionalized processes for solar electricity generation at the balance authority level exist within the FLCAC. These processes are for utility-scale electricity production and are therefore only a proxy for the generation of electricity using onsite panels. If it is necessary to model on-site electricity generation using photovoltaics, it is recommended to use those from the balance authority closest to the concrete plant, as it accounts for local solar incidence, which may be similar to the one where the concrete plant is located.*

[d] *Production processes from USLCI and US-EI are considered equivalent*

[e] *Although the ELCD has been providing free-of-charge, well-documented Life Cycle Inventory (LCI) data on resource consumption and emissions for many processes. This process is considered a data gap because it is not present in the US-focused public LCA database.*

[f] *Only 2 of the 107 EPDs reporting "Recycled Concrete Aggregates" as an input material use supplier-specific data as a data source; the remaining 105 EPDs use a custom process based on the US-EI dataset. (Table 8)*

[g] *About 57 % of the EPDs reporting Limestone as an input material use proprietary datasets from the US-EI database; the remaining EPDs use supplier-specific data. (Table 8)*

Of the 21 processes identified in Table 7, four processes have been modeled using public datasets by all the EPDs reporting their use:

- Propane production
- Renewable Diesel Combustion Truck, Heavy Duty, Short-haul Truck
- Compressed Natural Gas Heavy Duty Short-haul Truck
- Pigment production

Additionally, three processes—'production of limestone,' 'production of lubricating oil,' and 'Combustion of Propane in industrial equipment'—are also available as public datasets, although EPDs do not use them to model their respective processes. In all, these seven processes have public datasets and are not classified as data gaps.

Seven reported production processes are currently unavailable on a public database as datasets and are considered primary data gaps. The EPDs that report these processes have used background datasets either from a proprietary database:

- Production of recycled concrete aggregates
- Production of Carbon black pigments
- Production of Dolomite
- Production of microfibers
- Production of citric acid for cleaning chemicals
- Production of Phosphoric acid for cleaning chemicals
- Disposal of municipal solid waste

Additionally, two input items— 'wash water well' and 'on-site solar [8]' —have been mentioned without listing the background dataset used for the modeling. However, on the public database, datasets describing the resources used and emissions generated during their production or usage are missing, and hence, 'producing wash water from wells' is considered a primary data gap. However, 'generating electricity from solar' is considered a secondary data gap* because of the availability of the proxy data set on the public database.

Supplier-specific data (MSP EPDs) were used to model the impacts of the five raw material production processes; however, the production processes of these raw materials are not available as public datasets. Thus, these five processes are considered secondary data gaps

- Production of other aggregates
- Production of Carbon-cure cement
- Production of ground glass pozzolan
- Production of CO_2
- Production of Rock Dust

Some EPDs also listed 'steel fibers,' a recovered material used as an input in concrete mixture production. Per this study's methodology, three additional processes have been linked to the recovered material (Table 7). A process that produces steel coil is already available on the public database. Due to its availability on the public database, the production process of steel is not considered a data gap. It should be noted that despite its availability, the production of rolled steel does not impact the LCA results of concrete. Besides the production of steel fibers, there are no disposal processes for these steel fibers; the disposal of the steel fibers does not impact the results of the concrete LCA. Thus, the absence of a 'disposal process for the steel fiber' is considered a tertiary data gap. However, the potential 'treatment process turning the recovered steel cords into usable input material' for concrete is also missing from the database, and this treatment process impacts the results of the LCA and thus is considered a primary data gap. Treatment processes for recovered material are missing from public databases but should still be included in the EPDs of Ready-Made Concrete per the PCR.

In all, 24 processes have been identified in Table 7 and shown in Figure 4, of which the eight processes available on the public database are not considered a public data gap, including four production processes for which public datasets have been used, three production processes for which public datasets are available but not used, and one production process for the recovered

[8] *A spot check has not been conducted to check if the EPDs listing the wash water well or on-site solar report any background data sources used to model the process.*

material for which a public dataset is available. Nine processes, including the seven production processes for which the datasets are currently not available on a public database, one production process for which background data has not been mentioned in EPDs, but the public database does not have representative datasets that can model their production, and one treatment process of recovered material, are considered primary data gaps. Six production processes, of which five use supplier-specific data to model the production processes, and one production process for which a proxy data set is available, are considered secondary data gaps. One disposal process of recovered material is considered a tertiary data gap.

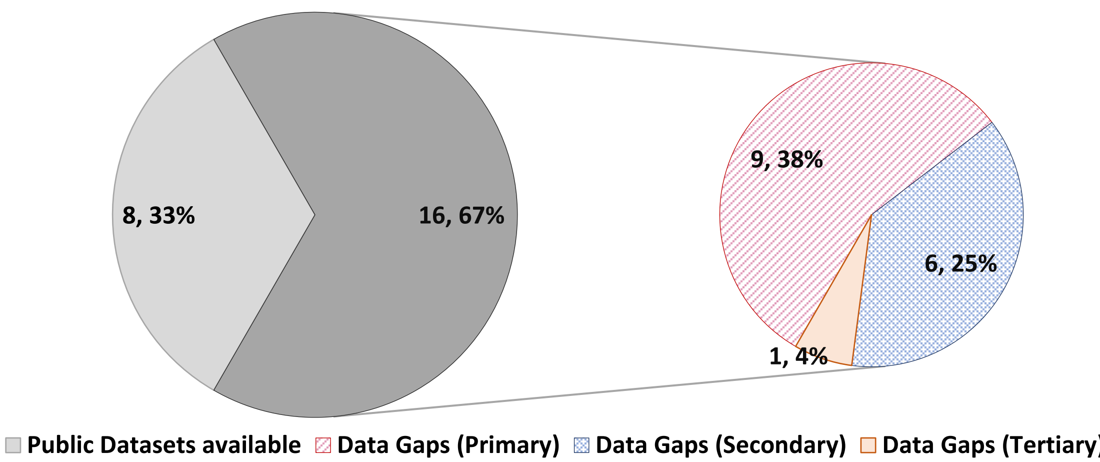

Figure 4 Data gap analysis for EPD reported processes without PCR prescribed data sources

Table 8 highlights the data source usage for the processes identified in the EPDs without PCR data prescriptions. The processes that use different data sources across EPDs are '*RCA production*' and '*Limestone production*' (*italicized* in Table 8). The datasets in **bold** represent public datasets (Table 8).

Table 8 Background data source reported by concrete EPDs that do not have PCR data prescriptions

SL No	Process	Data source used	# EPDs using the data source [d]	# EPDs reporting the process [d]
1	Other aggregate production	Supplier-specific data	8 987	8 987
2	*RCA production*	*1. Custom process, based on US-EI "Gravel, crushed, at mine*	*105*	*107*
		2. Supplier-specific data	*2*	
[a]	Crushed aggr. production for blend of aggregates	US-EI: Gravel, crushed, at mine and US-EI Gravel, round, at mine	1 464	1 464
[b]	Natural aggr. production for blend of aggregates			
3	Carbon Cure cement prod.	Supplier-specific data	611	611
4	Ground Glass Pozzolan prod	Supplier-specific data	13	13
5	Dolomite production	US-EI: Dolomite at plant	82	82
6	*Limestone production*	*1. US-EI: Limestone, crushed, for mill*	*3 067*	*5 416*
		2. Supplier-specific data	*2 349*	
[c]	Metakaolin production	US-EI: Expanded clay, at plant	2	2

7	Microfibers production	ELCD: PP fibers, crude oil based, production mix, at plant, PP granulate, without additives [e]	268	268
8	Carbon black pigment prod.	EI: Market for Carbon black	2	2
9	CO_2 production	Supplier-specific data	45	45
10	Pigments production	**USIO: Synthetic dye & pigment**	**21**	**21**
11	Rock dust production	Supplier-specific data	63	63
12	Wash water (well) prod.	Not mentioned [f]	15	15
13	Oils/Lubricants/Greases production	EI: Market for lubricating oil	37 694	37 694
14	Citric acid production for cleaning chemicals [j]	EI: Market for citric acid and EI: Market for industrial-grade Phosphoric acid	28 311	28 311
15	Phosphoric acid production for cleaning chemicals [j]			
16	CNG HD Short-haul truck	**GREET: CNG HD truck short Haul, GREET WTW, from waste**	**618**	**618**
17	RD Combination HD Short-haul Truck	**GREET: HD Truck, combination, short-haul, RD**	**3 646**	**3 646**
18	On-site solar	Not mentioned [f]	199	199
19	Propane Production [g]	**Production: USLCI/US-EI [h] Liquefied petroleum gas, at refinery** and USEPA Emission factor '08	8 306	8 306
20	Propane Combustion [g]			
21	Non-Hazardous SW disposal	US-EI: Disposal of Municipal Solid Waste	39 962	39 962
Rec.	Steel Fibers Production	byproduct of steel production; no upstream manufacturing impacts	4	4
Mat.	Steel Fibers Treatment			
3 [i]	Steel Fibers disposal			

Note: Processes in **bold** have publicly available datasets, while *italicized* processes are modeled using different datasets across EPDs.

[a, b, c] *The background datasets used for modeling these three production processes (distinguished by a, b, and c) are also used to model other production processes as prescribed by the PCR (Table 3), and thus they are not counted again for data gap analysis.*

[d] *Please note that the number of EPDs using a specific data source has been made based on what has been reported in the database; a spot check has not been conducted to see if the EPDs use a data source to model the respective process.*

[e] *Although the ELCD has been providing free-of-charge, well-documented Life Cycle Inventory (LCI) data on resource consumption and emissions for many processes. This process is considered a data gap since it is not present in the U.S. federal LCA database.*

[f] *On-site solar and wash water wells have been mentioned as input items used for the production of concrete mixture, but the background data sources used to model their respective processes have not been identified within the EPD database. Regionalized processes for solar electricity generation at the balance authority level exist within the FLCAC. These processes are for utility-scale electricity production and are therefore only a proxy for the generation of electricity using onsite panels. If it is necessary to model on-site electricity generation using photovoltaics, it is recommended to use those from the balance authority closest to the concrete plant, as it accounts for local solar incidence, which may be similar to the one where the concrete plant is located.*

[g] *Although Propane is reported as one fuel source, the data sources used, model two processes—'production of propane' and 'combustion of propane.' Hence, this study separates them into two separate processes: the production and the combustion process.*
[h] *The process, "Liquefied petroleum gas, at refinery" from USLCI and Datasmart (US-EI) are considered equivalent.*
[i] *Steel fiber, a reported recovered material, is linked to three processes per the study's methodology.*
[j] *Like Propane, although cleaning chemicals are reported as a single input source, the data sources used, model two processes —'production of citric acid' and 'production of industrial phosphoric acid.' Hence, this study separates them into two separate processes.*

3.1.4. Public data gap inferences

As shown in Table 8, 21 processes (less than 45 %) require data source prescriptions in the upcoming PCR revisions. A general trend observed across concrete EPDs is that for most processes, EPDs consistently select the same data source. Thus, the potential next step toward background data harmonization could focus on the PCR, clarifying the precedence of data sources and selecting a default source for yet unprescribed processes. This could involve ensuring MSP EPDs are always data sources used to model the production of a particular flow whenever available. If they are not, industry-wide EPD should be used as a second-best option. These requirements are consistent with ISO 21930's Section 7.19 [13]. However, these requirements are inconsistently applied in the current PCR. For processes where these data sources are not available (e.g., transport or energy), a general requirement could reduce the number of data sources used. One such requirement could be: "When no EPDs are available, processes from a public database shall be used whenever available. Only when a process has no publicly available dataset should commercial databases be used."

Amidst ongoing efforts to harmonize data usage, three factors have been identified in this study that may be contributing to the observed variations in database selection: (1) inconsistent or unclear data requirements and prioritization within this product category, (2) a lack of interoperability across public databases, and (3) consistency in data selection using unprescribed databases. An example of (1) identified by this study is the use of the same prioritization ranking for two different data sources for admixtures (MSP EPDs and water-reducing admixtures). An example of (2) is the lack of interoperability between GREET and USLCI, two of the most used public LCA databases. For (3), certain processes without specific prescribed data sources consistently rely on a single data source (e.g., ground glass pozzolans rely on an MSP EPD). Prescriptions for specific input materials could be specified in a future version of the PCR. For other input materials (e.g., cleaning chemicals), EI or US-EI are selected because these production processes are not available in public databases. Addressing each of these three factors would increase the harmonization of data selection within this product category. Figure 5 summarizes the data source selection for various processes in concrete EPDs. Overall, the highest reliance is on IA or MSP EPDs, which are supply-chain-specific and enhance the accuracy and representativeness of the models. USLCI is the third most utilized database.

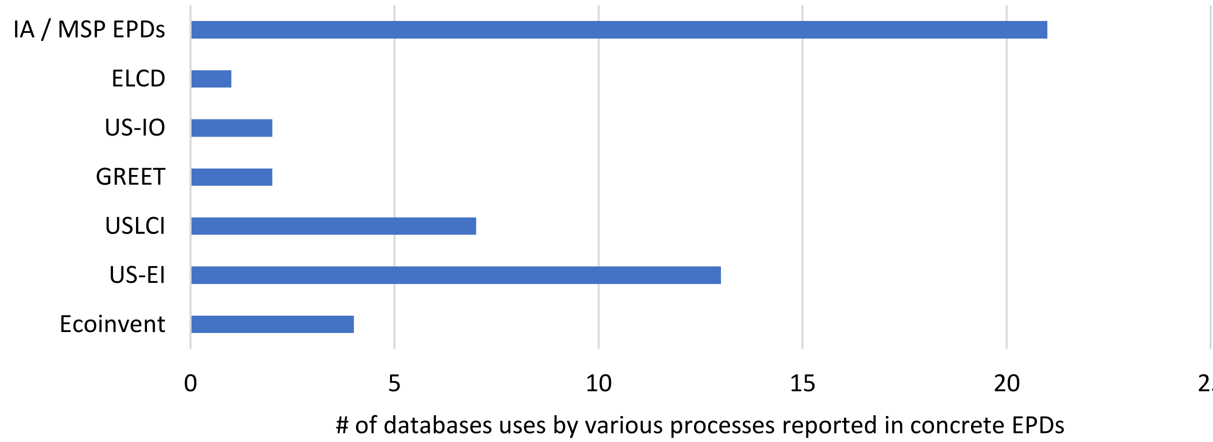

Figure 5 Database usage for various processes in concrete EPDs

A structured approach for filling data gaps could begin with creating publicly available datasets for production or treatment processes currently unavailable in a public database. Next, efforts could focus on converting reports into models within a public database (e.g., USLCI). Consolidating all the prescribed processes within the public database(s) would provide a unified and transparent repository. Finally, tertiary data gaps could be filled. In parallel with addressing public data gaps, strategies for incorporating access to manufacturer-specific data into public databases could also be explored (e.g., attaching a system process to each EPD). These strategies could include mechanisms to maintain intellectual property confidentiality while transparently providing necessary data to EPD users involved in product labeling or procurement decisions.

3.2. Cement

The cement EPDs are required to conform to the NSF PCR for Portland, Blended, Masonry, Mortar, and Plastic (Stucco) Cements v2 or later [36]. As of the date of analysis, 33 cement EPDs had been manually downloaded from the EC3 website, and the associated database was created by manually extracting the relevant data into a spreadsheet. Cement EPDs aggregate information for multiple products manufactured at a single facility by the product manufacturer. As a result, a single EPD may report information on multiple products.

3.2.1. PCR Public Data Gap Assessment

The cement PCR does not prescribe or recommend datasets for LCA modeling other than requiring that data on electricity production be representative of the location at the NERC region (or similar), preferably including transmission and distribution losses.

3.2.2. EPD LCI data reporting

Only 39 % (13) of cement EPDs have disclosed information regarding the LCI datasets used in EPD preparation[9], as illustrated in Figure 6. This lack of transparency is counter to that of concrete EPDs, even though both PCRs have the same POs (i.e., NRMCA and ASTM International). Additionally, the disclosure of used datasets is not uniform across EPDs created by a given PO. Figure 6 shows that 38 % and 44 % of EPDs developed by ASTM International and NRMCA, respectively, do not disclose LCI data sources.

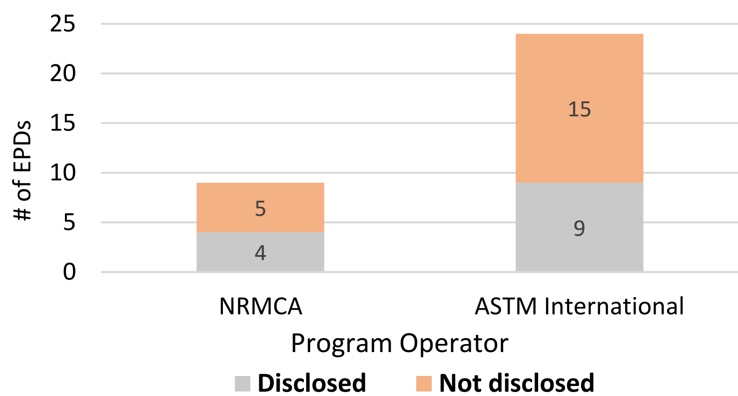

Figure 6 Background data disclosure in cement EPDs based on program operators.

Table 9 demonstrates the 24 processes identified in the cement EPDs; of those, eleven processes are reported by only one EPD (rows shaded gray in Table 9). Given the limited disclosure of LCI data sources in the cement EPDs, the cement EPD database analysis can only be completed on the subset of the EPDs that reported processes to assess whether a USLCI dataset could have been picked to model the process in all instances it was reported.

Table 9 Number of EPDs reporting different processes from the respective database

SL No	Process	Total EPDs including the process	ULSCI	US-EI	EI	Others
1	***Short-haul Truck Transport***	***13***	***13***			
2	***Long Haul Truck Transport***	***13***	***13***			
3	Train Transport	4	3		1	
4	Ocean Freight Transport	1			1	
5	Barge Transport	3	2		1	
6	Medium Voltage Electricity Generation	1			1	
7	High-Voltage Electricity Generation	12		11	1	
8	Natural gas production	10	1		9	
9	***Petroleum Coke production***	***11***	***1*** [b]	***10*** [b]		
10	***Diesel combustion in industrial boiler***	***4***	***1*** [a]	***3*** [a]		
11	Hard Coal (Anthracite) production [e]	7			7	

[9] The section 9.3 of the cement PCR says, "The EPD **shall** include the following: (A)..., (B)..., **(**C). A table outlining the primary sources of data used to complete the upstream material LCI background data including the date or version number.

SL No	Process	Total EPDs including the process	ULSCI	US-EI	EI	Others
12	Bituminous coal production	1	1			
13	Gasoline combustion in industrial equipment	1	1			
14	Ammonia production	1			1	
15	Diethylene glycol production	1			1	
16	Glycine production	1			1	
17	*Clay production [c]*	*1*			*1*	
18	Clinker production	1				1
19	*Silica Sand production [d]*	*2*			*2*	
19[d]	*Silica sand production for foundry sand [d]*	*1*			*1*	
17[c]	*Clay production for foundry sand [c]*	*1*			*1*	
11[e]	Anthracite production for foundry sand [e]	1	1			
20	Gypsum production	3		1	1	1
21	Hydrated Lime production	1	1			
22	Iron ore production	1	1			
23	Limestone production	11	2			9
24	Water production	1			1	

Note: *Italicized* processes are modeled using the same datasets across all EPDs reporting the respective process. Processes in **bold** are modeled using same public datasets across different EPDs reporting the respective process.

[a, b] The Diesel combusted in industrial boiler (a) processes from USLCI and US-EI are considered equivalent although they are datasets from different databases, similar to petroleum coke (b).
[c, d] The EPD reporting foundry sand discloses the usage of three background datasets: 'silica sand production' (d) and 'clay pit operation'(c) from EI and 'anthracite coal production' from USLCI. The two production processes from EI are also used as background data for producing other raw materials (silica sand and clay). Since the production processes are the same across different raw materials, each background dataset appears twice but is counted once.
[e] Hard coal or Anthracite are provided by datasets from different databases, yet both of these datasets represent the production of a single raw material—hard coal (anthracite)

The transportation processes are predominantly modeled using USLCI datasets, with some instances of reliance on EI. For the production and combustion of energy and fossil fuels, datasets from US-EI and EI are mostly used, while USLCI datasets are utilized sparingly. Notably, there are no processes within the energy category whose provision relies exclusively on USLCI. The production of raw and ancillary materials is not modeled using public datasets.

For production processes reported by only one EPD, it is infeasible to determine whether the background dataset used in modeling the process would have been consistent across all the EPDs had they all reported their background data. Of the remaining thirteen processes reported by more than one EPD, six processes are modeled using the same dataset across all EPDs (text *italicized* in Table 9): *Short Haul truck, long haul truck, Petroleum coke* [10]*, Diesel*[11]*, Clay, and Silica sand*. Of which, four are modeled using public datasets (Highlighted in **bold** in Table 9).

3.2.3. EPD public data gap assessment

Table 10 identifies the 32 data sources covering 24 processes, as shown in Table 9 and summarizes public data gaps. There are public datasets available for six of the 24 processes, and they were used in all reported instances. Thus, these six processes are not considered data gaps: 'Short-haul truck transport,' 'Long-haul truck transport,' 'Petroleum coke production [12],' 'Diesel combustion[13],' 'Bituminous coal production,' and 'Gasoline combustion'. Four processes are available as public datasets and used in some EPDs. Due to their availability in a public database, these four production processes are also not considered data gaps: 'Train transport,' 'Barge Transport,' 'Natural Gas production,' and 'Hard coal/Anthracite production'. The commercial database (EI) is generally used for natural gas, while the public database (USLCI) is more commonly used for transport activities, as illustrated in Table 9. Five processes are available on a public database but have not been reported as used in any EPDs. Although not used, these processes are available in a public database. Thus, they have not been classified as data gaps: 'Ocean transport,' 'Medium-voltage electricity,' 'Ammonia production,' 'Gypsum production,' and 'Limestone production'.

Table 10 Public Data Status for cement EPDs.

	Flows	Data sources used by EPDs	FLCAC equivalent	Gap Priority	Category
Transport: Land	*Train*	1. USLCI – Transport, train, diesel-powered			V
		2. EI – Transport, freight train, market (US)			V
	Truck	USLCI – Transport, combination truck, long-haul, diesel-powered			X
		USLCI – Transport, combination truck, short-haul, diesel-powered			X
Transport: Ocean	*Ocean transport*	EI – Transport, freight, sea, transoceanic ship	USLCI: Transport, ocean freighter, residual fuel oil powered		Y
	Barge transportation	1. USLCI - Transport, barge, diesel-powered			V
		2. EI – Transport, freight, inland waterways, barge			V

[10] The Petroleum coke, at refinery processes from USLCI and US-EI are considered equivalent although they are datasets from different databases
[11] The Diesel combusted in industrial boiler processes from USLCI and US-EI are considered equivalent although they are datasets from different databases
[12] The Petroleum coke, at refinery processes from USLCI and US-EI are considered equivalent although they are datasets from different databases
[13] The Diesel combusted in industrial boiler processes from USLCI and US-EI are considered equivalent although they are datasets from different databases

	Flows	Data sources used by EPDs	FLCAC equivalent	Gap Priority	Category
Energy:	*High Voltage Electricity*	1. US-EI – High voltage electricity at grid		1	
		2. EI - Electricity, high voltage			
Electricity	*Medium Voltage Electricity*	EI - Electricity, medium voltage	USLCI: Electricity; at grid; consumption mix		Y
	Natural gas	1. EI - Natural gas, high pressure			V
		2. USLCI - Natural gas, processed, at plant			V
Energy:	*Petroleum coke*	US-EI / USLCI [f] – Petroleum coke, at refinery			X
Fossil Fuel	*Diesel*	US-EI / USLCI[f] - Diesel, combusted in industrial equipment			X
	[Hard] Coal	EI – Hard coal, hard coal operations, and hard coal preparation [g]	USLCI: Anthracite coal, at mine		V
	Bituminous coal	USLCI - Bituminous coal, at mine			X
	Gasoline	USLCI – Gasoline, combusted in industrial equipment			X
Raw Materials:	*Ammonia*	EI - Ammonia, liquid, ammonia production, steam reforming, liquid	USLCI: Ammonia, steam reforming, liq. at plant		Y
Chemicals	*Diethylene glycol (GA 1)*	EI – Diethylene glycol production	USLCI: Ethylene glycol, material production, organic compound	2*	
	Glycine (GA 2)	EI – Glycine production		1	
	Clay	EI – Clay, clay pit operation [b]		1 [(d)]	
Raw Materials:	*Clinker*	Supplier-specific data		2	
	Silica sand	EI – Silica sand production [c]		1 [(e)]	
	Foundry Sand	EI – Silica sand production [c]		3 [(e)]	
		EI – Clay, clay pit operation [b]		3 [(d)]	
Others		USLCI – Anthracite coal, at mine [g]			V
			Production	3	
			Treatment	1	
			Disposal	3	
	Gypsum	1. US-EI, Gypsum mineral, at mine	ANL C&C: Gypsum; at quarry		Y
		2. EI– Gypsum, mineral			Y
		3. AthenaLCI – Gypsum natural, at quarry/mine			Y
	Hydrated lime	EI – Lime. Hydrated, lose weight	USLCI: Quicklime, at plant [h]	2*	
	Iron ore	EI - Iron ore, beneficiated, 65 % Fe		1	
	Limestone	1. EI – Limestone, crushed for mill, production	USLCI: Limestone, at mine		Y
		2. Manufacture-specific Primary Data			Y

Flows	Data sources used by EPDs	FLCAC equivalent	Gap Priority	Category
Water	EI – Tap water production, underground water without treatment		1	

Legend	Priority / Category
A public production process is available and is used by EPDs	X
Publicly available datasets used along with proprietary datasets	V
A public production process is available, but not used by EPDs	Y
Primary Data gaps: Processes not available as a public data set	1
Secondary Data gaps: Prescribed IA or MSP EPD as a data source, but the report is not converted into a public dataset	2
Secondary Data gaps*: Data source identified as FLCAC equivalent is considered a proxy	2*
Tertiary Data gaps: Processes whose absence does not affect the results of the EPD	3

[a] *The US-EI Diesel combusted in industrial equipment can also be provided by an equivalent process from the USLCI database; thus, the process is not considered a data gap.*

[b,c] *The EPD reporting foundry sand discloses the usage of three background datasets, the three background datasets are 'silica sand production' (c) and 'clay pit operation'(b) from EI, and 'anthracite coal production' from USLCI. The two production processes from EI are also used as background data for producing other raw materials (silica sand and clay). Since the production processes are the same across raw materials, the two background datasets, although appearing four times in total, are counted twice.*

[(d), (e)] *Although foundry sand has three production processes reported in one of the EPDs, it is typically a waste product that cement manufacturers purchase for a nominal price from foundries* [37]. *According to this study's methodology, the production of foundry sand is considered burden-free. Therefore, the inputs required to manufacture foundry sand do not impact cement production. Consequently, the gaps related to manufacturing foundry sand are considered tertiary gaps, even though these production processes (distinguished by (d) and (e)) might be considered primary gaps where their usage directly impacts cement production.*

[f] *Production processes from USLCI and US-EI are considered equivalent*

[g] *Hard coal or Anthracite is provided by datasets from different databases, yet both datasets represent the production of a single raw material—hard coal (anthracite)*

[h] *Authors acknowledge that quicklime and hydrated lime have different chemical compositions, but the asphalt PCR prescribes the production process of quicklime as a background dataset for hydrated lime production, and per this document's approach, this production process for quicklime has been identified as a proxy, thus the same production process is also identified as a proxy background dataset for hydrated lime in the table above*

Six reported production processes are currently unavailable on a public database, and the EPDs that report these processes have used background datasets from a proprietary database:

- Generation of high-voltage electricity
- Production of Glycine
- Production of clay
- Production of silica sand
- Production of Iron ore
- Production of tap water

'Production of clinker' has been modeled using manufacturer-specific primary data and has been categorized as a secondary data gap. The last two production processes, 'production of diethylene glycol (one of the grinding aids)' and 'production of hydrated lime,' are also not available on the public dataset. Although the actual production processes of these raw materials are not available, a similar production process (i.e., proxy) is available on the public database. These two production processes are classified as secondary data gap*.

One EPD reports using an EI process for production of 'diethylene glycol' (DEG), one of the grinding aids. DEG and ethylene glycol (EG) are both produced through partial hydrolysis of ethylene oxide, and USLCI has a process for production of EG. However, this USLCI process is only considered a proxy because EI reports slightly different inputs and outputs for production of EG and DEG. For this study, their production processes are considered similar, but not equivalent. Therefore, the production of DEG is a secondary data gap.

Beyond the 24 processes identified in the cement EPDs, three additional processes have been linked to foundry sand per this study's approach, increasing the total considered processes for this analysis to 27. Although the only EPD reporting the usage of foundry sand in the production of cement has disclosed using three background datasets (production of silica sand, clay, and anthracite) to model the production of foundry sand, the literature shows that foundry sand used in the production of cement is typically a waste product recovered from the foundries by the cement producers for a nominal price [37]. Hence, the three additional processes have been linked to the production, potential treatment, and disposal of foundry sand.

There are no processes in a public database that account for the production or disposal of foundry sand; hence, the 'production process of foundry sand' and 'disposal process of foundry sand' are considered tertiary data gaps. The potential 'treatment process turning the waste foundry sand into usable input material' for cement is also missing from the database, and these treatment processes are considered primary data gaps. Prior to use, spent foundry sand requires crushing or screening to reduce or separate oversized materials that may be present [38]. Although these treatment processes are missing from public databases, they should be included in the EPDs of cement per the PCR.

Figure 7 summarizes the data gaps for the 27 processes identified (Table 10). Fifteen processes, including six processes for which only a public dataset has been used across all EPDs, four processes for which public and proprietary datasets have been used, and five processes for which proprietary datasets are used despite the availability of equivalent public datasets, are not considered data gaps. Seven are primary gaps, including the production of five raw materials, the generation of high-voltage electricity, and the treatment of foundry sand. Three production processes for which either a proxy or manufacturer-specific primary data has been used are classified as secondary data gaps. Finally, two tertiary data gaps—the production and disposal of foundry sand.

A structured approach for filling data gaps could begin with creating publicly available production processes for the eight currently unavailable processes. Next, efforts could focus on creating actual production processes for the secondary data gaps to replace similar production processes (i.e., proxies). Once these priority gaps are addressed, tertiary data gaps could be filled.

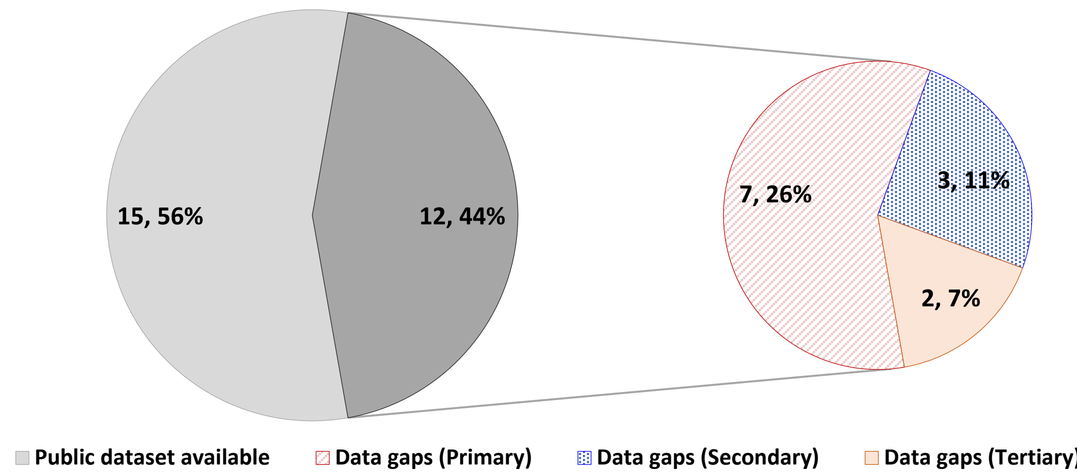

Figure 7 Public data gap analysis of the processes highlighted in Table 10

3.2.4. Public Data Gap Inferences

In summary, of the 24 processes identified in Table 9, seven processes (about 29 %) have been modeled using different datasets across different EPDs (Table 9 and Table 10). These are:

- Train transport
- Barge transport
- High-voltage electricity
- Natural gas production
- Gypsum production
- Limestone production
- Hard coal/Anthracite

Of the remaining seventeen processes, eleven (46 %) are reported in only one EPD (highlighted in gray on Table 9). Thus, it is infeasible to evaluate dataset consistency. The remaining six (25 %), reported in more than one EPD, are consistently modeled using the same background dataset across different EPDs reporting the process. Figure 8 illustrates the usage of different databases for all the processes reported in the cement EPDs.

Not all EPDs disclose datasets, which limits the ability to generalize the assessment conducted in this study. To further the background data harmonization in cement EPDs, a critical first step would be to ensure that the background data being used is reported. Further, to make the EPDs machine-readable, separating multiple products currently grouped together under a single EPD by a single producer's plant location into individual EPDs could be considered.

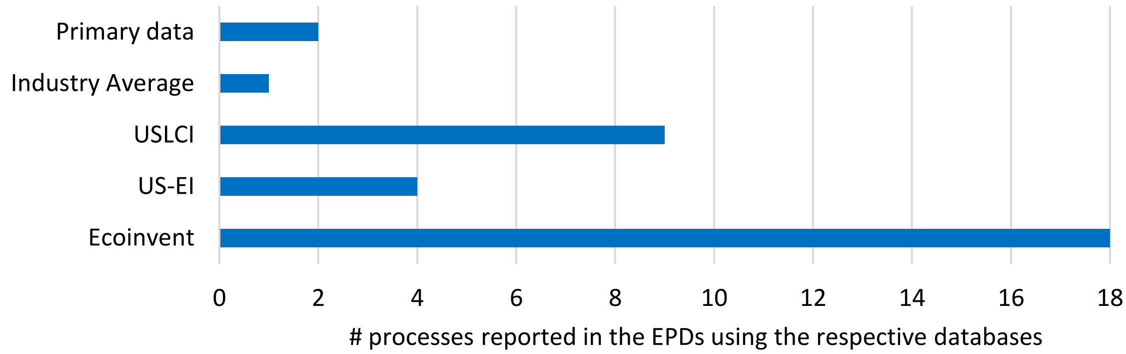

Figure 8 Database usage for different processes reported by EPDs

3.3. Asphalt

The PO for asphalt EPDs is the National Asphalt Pavement Association (NAPA), which has implemented a standardized EPD template in HTML format. As of January 2025, there are at least 4048 valid product-specific asphalt EPDs across the U.S. These EPDs conform to NAPA's PCR for Asphalt Mixtures, version 2.0 [39]. The asphalt EPD database was created by bulk downloading the EPDs and extracting the required information through hard-coding the HTML data into a spreadsheet, as mentioned in Section 2.1.2.

3.3.1. PCR public data assessment

The asphalt PCR prescribes upstream data sources required for asphalt EPDs as listed in Annex 1 [40]. Annex 1 also identifies data gaps in said data sources, which are further elaborated in this section.

3.3.1.1. PCR public data gap prescription

The asphalt PCR actively promotes transparency in choosing data sources by prescribing publicly available, free data sources for LCA modeling in Annex 1 [40]. Section 7.1.9 of the PCR – "Selection of Data and Data Quality Requirements" – further defines the criteria and requirements for selecting both background and foreground data. Annex A of the asphalt PCR identifies 99 background datasets. Table 11 summarizes the PCR-prescribed data sources and highlights the respective data gaps as identified by the PCR.

For foreground data, plant-specific annual data from 2017 or later are prescribed for asphalt production (A3), and mix-specific data associated with the extraction, upstream production, and transportation of raw materials are prescribed for the A1 and A2 stages of the LCA modeling.

For background data, the PCR prioritizes the use of publicly available data sources over proprietary data sources for LCA modeling. The asphalt PCR prescribes the use of facility- and product-specific (MSP) EPDs modeled according to TRACI v2.1 as background datasets. If MSP EPDs are not available, then IA EPDs modeled according to TRACI v2.1 shall be used. In cases where neither MSP nor IA EPDs are available, the PCR prescribes the use of a freely available public database included in Annex 1. These datasets include critically reviewed LCA studies that comply with ISO 14040/14044.

Section 7.19.2 of the PCR also encourages developing an MSP or IA EPD, for which potential upstream data for asphalt mixture modeling to use the datasets prescribed in Annex 1 for common input materials to ensure improved consistency and comparability across materials.

Table 11 Asphalt PCR data gap analysis – based on data gaps identified by the PCR

	Flow	**PCR-prescribed data sources**
Raw Material: Binder	*Asphalt binder, refined from crude oil*	North American avg: Asphalt binder, no additives, consumption mix, at terminal, from crude oil [41]
		North American avg: Asphalt binder, 0.5 % PPA (by weight of asphalt binder), consumption mix, at terminal, from crude oil [41]
		North American avg: Asphalt binder, 3.5 % SBS (by weight of asphalt binder), consumption mix, at terminal, from crude oil [41] *<when SBR or SBS is the modifier>* [a]
		North American avg: Asphalt binder, 8 % GRT (by weight of asphalt binder), consumption mix, at terminal, from crude oil [41]
	Bio-based asphalt binder	**Data gap**
	Emulsified asphalt	**Data gap**
	Cutback asphalt	**Data gap**
	Natural asphalt binder (Gilsonite & Trinidad Lake asphalt)	**Data gap**
Raw Material: Aggregate	*Crushed stone, sand, gravel*	[42] for Coarse aggregate from crushed stone
	Recycled aggregates	[42] for Coarse aggregate from crushed stone
	Slag aggregate	**Data gap**
Raw Mat.: RAP/RAS	*RAP*	[42] for RAP and RAS
	RAS	[42] for RAP and RAS
Raw Material:	*AA- hydrated lime*	USLCI - Quicklime, at plant
	AA – (LA) Amidoamines	**Data gap**
	AA - (LA) Imidazolines	**Data gap**
	AA - (LA) Organo-silanes	**Data gap**
	AA - (LA) Polyamines	**Data gap**
	BME: Biopolymers	**Data gap**
	BME: GRT	Asphalt binder, 8% GRT (by weight of asphalt binder), consumption mix, at terminal, from crude oil [41]
	BME: Natural rubber	**Data gap**
	BME: Polychloroprene latex	**Data gap**
	BME: RET	**Data gap**
	BME: SBR	Asphalt binder, 3.5 % SBS (by weight of asphalt binder), consumption mix, at terminal, from crude oil [41]
	BME: SBS	Asphalt binder, 3.5 % SBS (by weight of asphalt binder), consumption mix, at terminal, from crude oil [41]
	BMEx: Bio-based oils	**Data gap**
	BMEx: Lignin	**Data gap**
	BMEx: Petroleum oils	**Data gap**
	BMEx: REOB / VTAE	**Data gap**
	BMEx: Sulfur	**Data gap**

	Flow	PCR-prescribed data sources
	BMNA): Gilsonite	**Data gap**
	BMNA: Trinidad Lake asphalt	**Data gap**
	BMO: Biochar	**Data gap**
	BMO: PPA	North American avg: Asphalt binder, 0.5 % PPA (by weight of asphalt binder), consumption mix, at terminal, from crude oil [41].
Binder Additives	*BMP: EA copolymer*	**Data gap**
	BMP: EPM	**Data gap**
	BMP: EPDM	**Data gap**
	BMP: EVA	**Data gap**
(Generic Examples)	*BMP: Polyethylene*	**Data gap**
	BMP: Polyolefins	**Data gap**
	BMP: Polypropylene	**Data gap**
	BMP: Recycled plastics	**Data gap**
	BMRA: Aromatic extracts	**Data gap**
	BMRA: Paraffinic oils	**Data gap**
	BMRA: Tall oil-based products	**Data gap**
	BMRA: Vegetable oil-based products	**Data gap**
	NF: cellulose	**Data gap**
	NF: Minerals	**Data gap**
	NF: Rock wool	**Data gap**
	SF: Aramid	**Data gap**
	SF: Fiberglass	**Data gap**
	SF: Polyester	**Data gap**
	SF: Polypropylene	**Data gap**
	MF: Baghouse fines	<Recovered material, Cut-off rule>
	MF: Crusher fines	Coarse aggregates from crushed stone [43]
	MF: Fly ash	<Recovered material, Cut-off rule>
	MF: Lime	USLCI: Quicklime, at plant
	MF: Portland cement	PCA: Portland cement, at plant [44]
	MF: slag cement	SCA: Slag cement [45]
	Pigments: Iron oxide	**Data gap**
	Pigments: Titanium dioxide	**Data gap**
	Recycled plastic, dry method	**Data gap**
	WMA, chemical: Amine surfactants, emulsifiers, and other chemical additives	**Data gap**
	WMA, organic: waxes and fatty acid amides	**Data gap**
	WMA, others: Zeolites (aluminosilicates)	**Data gap**
	WMA, others: Hybrid technologies (blends of chemical and water-based technologies)	**Data gap**
	Boat	NREL'21: Transportation by barge, diesel

	Flow	PCR-prescribed data sources
Transport		NREL'21: Transportation by ocean freighter, diesel
	Pipeline	NREL'21: Transportation by pipeline, natural gas
	Train	NREL'21: Transportation by train, diesel
	Truck	NREL'21: Transportation by combination truck, diesel
		NREL'21: Transportation by combination truck, gasoline
		NREL'21: Transportation by refuse truck, diesel
		NREL'21: Transportation by refuse truck, gasoline
Energy	*Electricity*	NREL'21: (Consumption-based data) Balancing Authority
Energy: Fuel Combustion (Boilers, Burners, And Other Heating Equipment)	*Biodiesel*	NREL'21 and [46]: Soy biodiesel, production, at plant with combustion emissions data by Miller'08
	Brown grease (aka grease trap oil, FOG)	USLCI and [46]: Soy biodiesel, production, at plant with combustion emissions data by Miller'08
	coal, anthracite	NREL'21: Anthracite coal, combusted in industrial boiler
	coal, bituminous	NREL'21: Bituminous coal, combusted in industrial boiler
	coal, lignite	NREL'21: Lignite coal, combusted in industrial boiler
	Diesel	NREL'21: Diesel, combusted in industrial boiler
	Landfill gas	NREL'21: Natural gas, combusted in industrial boiler (proxy)
	LNG	NREL'21: Natural gas, combusted in industrial boiler
	LPG (propane)	NREL'21: Liquified petroleum gas, combusted in industrial boiler
	Natural gas	NREL'21: Natural gas, combusted in industrial boiler
	Recycled fuel oil (used oil)	Recycled fuel oil, combusted in industrial boiler [42]
	Residual fuel oil	NREL'21: Residual fuel oil, combusted in industrial boiler
	Renewable diesel	NREL'21: Diesel, combusted in industrial boiler (proxy)
	RNG	NREL'21: Natural gas, combusted in industrial boiler
	Yellow grease (aka recycled vegetable oil)	NREL'21 and [46]: Soy biodiesel, production, at plant with combustion emissions data by Miller'08
Energy: Fuel combustion (Internal combustion Engines)	*Biodiesel*	NREL'21 and [46]: Soy biodiesel, production, at plant with combustion emissions data by Miller'08
	Compressed Natural Gas (CNG)	NREL'21: Compressed natural gas equipment operation, industry average, >56 kW and <560 kW
		NREL'21: Compressed natural gas equipment operation, industry average, >19 kW and <56 kW
	Diesel fuel	NREL'21: Diesel, combusted in industrial equipment
	Gasoline	NREL'21: Gasoline, combusted in industrial equipment
	LNG	NREL'21: Natural gas, combusted in industrial equipment
	LPG/propane	NREL'21: Propane, combusted in equipment [b]
	Renewable diesel	NREL'21: Diesel, combusted in industrial equipment
EOL	*Waste and by-products*	US EPA WARM inventory for "MSW, landfilling of asphalt concrete" with plant-specific transport distances by [42]

Legend
Public dataset available prescribed by the PCR
Proxy process prescribed by the PCR
Data gaps – identified by the PCR
Recovered material identified by the PCR

[a] *Asphalt binder can be modified using polymers such as SBS, Solution SBR, PPA, and GRT* [42]. *But to model the modified asphalt binder with SBS as an additive, SBR is used as a proxy* [41]
[b] *List of Potential Typing Errors: Could not locate the process on a public repository.*

3.3.1.2. PCR-based data gap summary

Figure 9 summarizes the data gaps in the prescribed data sources for the asphalt PCR. According to the PCR, 47 of the 99 identified background datasets (production or combustion processes) are publicly available. Further, 47 background datasets (production processes) are not publicly available and are identified as data gaps**.** All the data gaps identified are missing production processes for raw materials, most of which belong to the inventory category ‘additives.’ Annex 1 of the PCR also prescribes proxy background datasets of the following three inputs: ‘Recycled aggregates,’ ‘Landfill gas,’ and ‘Renewable diesel.’ The proxies prescribed for combustion processes for ‘Landfill gas’ and ‘Renewable diesel’ are available as models in an open-source LCA database. The proxy data source prescribed for ‘Recycled aggregates’ is a publicly available report, which could be used to create a production process on a public LCA database. The remaining two processes, ‘Baghouse fines’ and ‘fly ash,’ are marked as recovered materials by the PCR.

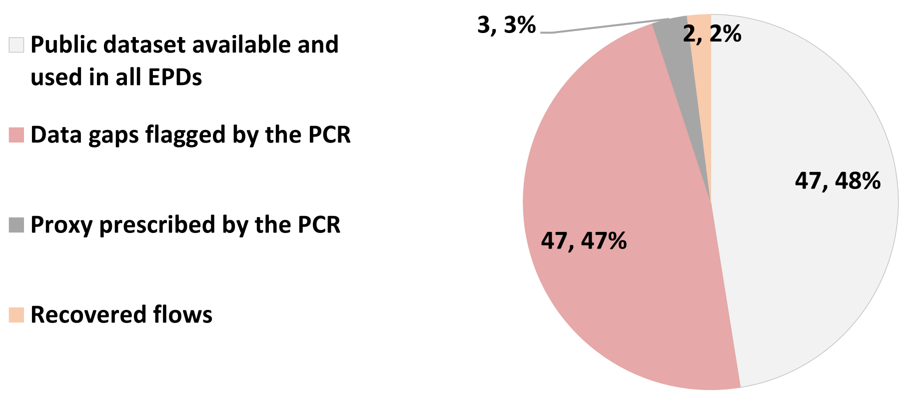

Figure 9 Public data gap analysis for data sources prescribed by the asphalt PCR

3.3.1.3. Data gap summary - per this study

Although the PCR included its own list of data gaps and proxies, it does not include a definition of what constitutes a data gap or what distinguishes a proxy from other models. Therefore, the background datasets listed in Table 11 are reevaluated according to the data gap definition presented in Section 3. Any background dataset (production/combustion/treatment process) present on the FLCAC database that could address a data gap identified by the PCR will be considered a “proxy.”

Based on this study’s approach, 108 background datasets (production/combustion/treatment processes) were identified in the asphalt PCR as reported in Table 12. There were three pairs of raw materials with different names, but were provided by the same production process, and are treated as only three production process data gaps, because the prescribed production process/background datasets were the same. More information about the raw material pairs is provided in Table 12 notes and the following paragraphs.

Table 12 Asphalt PCR data gap analysis – based on this report's definition of data gap

	Flow	Classification (PCR)	Classification (this report)	FLCAC process	Gap Priority
Raw Material: Asphalt Binder	*Asphalt binder, No additives*	Available	Available	USLCI: Asphalt binder, no additives, consumption mix, at terminal, from crude oil	
	Asphalt binder, 0.5 % PPA	Available	Available	USLCI: Asphalt binder, 0.5 % PPA, consumption mix, at terminal, from crude oil	
	Asphalt binder, 3.5 % SBS/SBR	Available	with SBR: Available	USLCI: Asphalt binder, 3.5 % SBS, consumption mix, at terminal, from crude oil [a]	
			with SBS: Secondary*	USLCI: Asphalt binder, 3.5 % SBS, consumption mix, at terminal, from crude oil [a]	2*
	Asphalt binder, 8 % GRT	Available	Available	USLCI: Asphalt binder, 8 % GRT, consumption mix, at terminal, from crude oil	
	Bio-based asphalt binder	Data gap	Primary	-	1
	Emulsified asphalt	Data gap	Primary	-	1
	Cutback asphalt	Data gap	Primary	-	1
	Natural asphalt binder	Data gap	Primary	-	1
Raw Material: Aggregate	*Crushed stone, sand, and gravel*	Available	Secondary	Crushed stone [42]	2
			Available	Sand: ANL/C&C: Fine & coarse aggregates	
			Available	Gravel: ANL/C&C: Fine & coarse aggregates	
	Recycled aggregates	Proxy	Secondary	-	2
	Slag aggregates	Data gap	Primary	-	
RAS & RAP	*RAP*	Available	Available	US EPA - CDDM: Processing of reclaimed asphalt pavement	
	RAS	Available	Available	US EPA - CDDM: Processing of asphalt shingles; ground	
Raw Material: Additives	*AA - hydrated lime [b]*	Available	Secondary*	USLCI: Quicklime, at plant	2*
	AA (LA): Amidoamines	Data gap	Primary	-	1
	AA (LA) Imidazolines	Data gap	Primary	-	1
	AA (LA) Organsilanes	Data gap	Primary	-	1
	AA (LA) Polyamines	Data gap	Primary	-	1
	BME: Biopolymers	Data gap	Primary	-	1
	BME: GRT	Available	Secondary*		2*
	BME: Natural rubber	Data gap	Primary	-	1
	BME: Polychloroprene	Data gap	Primary	-	1
	BME: Reactive ET	Data gap	Primary	-	1
	BME: SBR	Available	Secondary*		2*
	BME: SBS	Available	Secondary*		2*
	BMEx: Bio-based oils	Data gap	Primary	-	1
	BMEx: Lignin	Data gap	Primary	-	1

	Flow	Classification (PCR)	Classification (this report)	FLCAC process	Gap Priority
(Generic Examples)	*BMEx: Petroleum oils*	Data gap	Secondary*	USLCI: Crude oil production	2*
	BMEx: REOB / VTAE	Data gap	Primary	-	1
	BMEx: Sulfur	Data gap	Secondary*	USLCI: Sulfur, at plant	2*
	BMNA: Gilsonite	Data gap	Primary	-	1
	BMNA: Trinidad Lake asphalt	Data gap	Primary	-	1
	BMO: Biochar	Data gap	Primary	-	1
	BMO: PPA	Available	Secondary*		2*
	BMP: Ethylene acrylate	Data gap	Primary	-	1
	BMP: EPM	Data gap	Primary	-	1
	BMP: EPDM	Data gap	Primary	-	1
	BMP: EVA	Data gap	Primary	-	1
	BMP: Polyethylene	Data gap	Secondary*	USLCI: HDPE/LDPE, virgin resin	2*
	BMP: Polyolefins	Data gap	Primary	-	1
	BMP: Polypropylene [d]	Data gap	Secondary*	USLCI: PP, virgin resin	2*
	BMP: Recycled plastic [c]	Data gap	Secondary*	USLCI: Recycled Post-consumer, PET & HDPE	2*
	BMRA: Aromatic extracts	Data gap	Primary	-	1
	BMRA: Paraffinic oils	Data gap	Primary	-	1
	BMRA: Tall oil products	Data gap	Primary	-	1
	BMRA: Vegetable oil products	Data gap	Secondary*	USLCI: Soy oil, refined, at plant	2*
	NF: cellulose	Data gap	Primary	-	1
	NF: Minerals	Data gap	Primary	-	1
	NF: Rock wool	Data gap	Primary	-	1
	SF: Aramid	Data gap	Primary	-	1
	SF: Fiberglass	Data gap	Secondary*	USLCI: Glass fiber; at plant	2*
	SF: Polyester	Data gap	Secondary*	USLCI: Unsaturated polyester, resin, at plant	2*
	SF: Polypropylene [d]	Data gap	Secondary*	USLCI: PP, virgin resin	2*
	MF: Baghouse fines	Cut-off	Production: Available	Same process as the production of asphalt mix modeled for the EPD	
			Disposal primary	-	1
	MF: Crusher fines	Available	Secondary	-	2
	MF: Fly ash	Cut-off	Production: Available	USLCI: Bituminous coal, combusted in industrial boiler USLCI: Electricity, bituminous coal, at power plant	

	Flow	Classification (PCR)	Classification (this report)	FLCAC process	Gap Priority
			Treatment: Primary		1
			Disposal: Tertiary	-	3
	MF: Lime [b]	Available	Secondary*	USLCI: Quicklime, at plant	2*
	MF: PC	Available	Available	USLCI: Portland cement [e]	
	MF: slag cement	Available	Secondary		2
	Pigments: Iron oxide	Data gap	Primary		1
	Pigments: TiO_2	Data gap	Primary		1
	Recycled plastic, dry method [c]	Data gap	Secondary*	USLCI: Recycled Post-consumer, PET & HDPE	2*
	WMA, chemical: Amine	Data gap	Primary		1
	WMA, organic: waxes	Data gap	Primary		1
	WMA, others: Zeolites	Data gap	Primary		1
	WMA, others: Hybrid	Data gap	Primary		1
Transport	*Barge*	Available	Available	USLCI: Transportation by barge, diesel	
	Ocean freighter	Available	Available	USLCI: Transportation by ocean freighter, diesel	
	Pipeline	Available	Available	USLCI: Transportation by pipeline, natural gas	
	Train	Available	Available	USLCI: Transportation by train, diesel	
	Combination truck, diesel	Available	Available	USLCI: Transportation by combination truck, diesel	
	Combination truck, gasoline	Available	Available	USLCI: Transportation by combination truck, gasoline	
	Refuse truck, diesel	Available	Available	USLCI: Transportation by refuse truck, diesel	
	Refuse truck, gasoline	Available	Available	USLCI: Transportation by refuse truck, gasoline	
Electricity	*Electricity*	Available	Available	USLCI: (Consumption-based data) Balancing Authority	
Energy: Fuel Combustion	*Biodiesel*	Available	Production: Available	USLCI: Soy biodiesel, at plant	
			Combustion: Secondary		2
	Brown grease (aka grease trap oil, FOG)	Available	Production: Tertiary		3
			Treatment: Primary		1
			Combustion: Secondary		2
	coal, anthracite	Available	Available	ULCI: Anthracite coal, BC	
	coal, bituminous	Available	Available	USLCI: Bituminous coal, BC	
	coal, lignite	Available	Available	USLCI: Lignite coal, BC	

	Flow	Classification (PCR)	Classification (this report)	FLCAC process	Gap Priority
(Boilers, Burners, And Other Heating Equipment)	*Diesel*	Available	Available	USLCI: Diesel, BC	
	Landfill gas	Proxy	Secondary*	USLCI: NG, BC	2*
	LNG	Available	Secondary*	USLCI: NG, BC	2*
	LPG (aka propane)	Available	Available	USLCI: LPG, BC	
	Natural gas	Available	Available	USLCI: NG, BC	
	Recycled/used fuel oil	Available	Secondary		2
	Residual fuel oil	Available	Available	USLCI: RFO, BC	
	Renewable diesel	Proxy	Secondary*	USLCI: Diesel, BC	2*
	RNG	Available	Secondary*	USLCI: NG, BC	2*
	Yellow grease (aka recycled vegetable oil)	Available	Production: Tertiary		3
			Treatment: Primary		1
			Combustion: Secondary		2
Energy: Fuel combustion	*Biodiesel*	Available	Production: Available	USLCI: Soy biodiesel, at plant	
			Combustion: Secondary		2
	CNG >56 kW and <560 kW	Available	Available	HEO & Pavement: CNG equipment operation, IA, >56 kW and <560 kW	
	CNG >19 kW and <56 kW	Available	Available	HEO & Pavement: CNG equipment operation, IA, >19 kW & <56 kW	
(ICE/ Internal Combustion Engine)	*Diesel fuel*	Available	Available	USLCI: Diesel, combusted in industrial equipment	
	Gasoline	Available	Available	USLCI: Gasoline, combusted in industrial equipment	
	LNG	Available	Secondary*	USLCI: NG, IE	2*
	LPG/propane	Available	Secondary*	USLCI: LPG, BC	2*
	Renewable diesel [f]	Available [f]	Secondary* [f]	USLCI: Diesel, IE [f]	2*
EOL	*Waste and by-products*	Available	Secondary		2

Legend		**Priority**
	A public production process is available and is used by EPDs	
Primary Data Gap	identified as a data gap by the PCR and the report	1
	Identified as a data gap per this report's classification, but a data source prescribed in the PCR	
Secondary Data gap	Identified as a data gap by the PCR, but a 'proxy' process has been identified per this report	2*
	Data source prescribed by the PCR, but is classified as a secondary data gap per report, as the prescribed PCR data source is considered a proxy (marked as "Secondary* ")	
	Data source prescribed by the PCR, but classified as a secondary data gap per the report because no model is available in a public LCA database (Marked as "Secondary")	2
Tertiary data gap	Recovered material identified by the PCR	3
	Additional recovered material identified by the report	

Note: the cell color denotes the PCR data gap classification of a process prescribed, and the text color indicates the data gap classification of a process per the report's methodology

[a] *Asphalt binder can be modified using polymers such as SBS, Solution SBR, PPA, and GRT* [42]. *But to model the modified asphalt binder with SBS as an additive, SBR is used as a proxy* [41]

[b, c, d] *The production of these pairs of raw materials constitutes only one data gap per pair of raw material, as the production process is assumed to be the same amongst each pair ('Antistrip agents- hydrated lime' and 'Mineral fillers: Lime,'(b) 'Binder Modifier Plastic: Recycled plastics' and 'recycled plastic, dry method' (c) AND 'Synthetic Filler: Polypropylene' and 'Binder Modifier Plastic: Polypropylene' (d))*

[e] *for PC, although the PCR prescribes the usage of IA EPD, the authors would also like to highlight the production process of PC available in the USLCI database called "Portland cement, at plant."*

[f] *ICE-Renewable diesel is considered a PCR-identified proxy because Industrial boiler combustion Renewable diesel is a PCR proxy*

Of the 47 data gaps identified by the Asphalt PCR (Table 11 and Figure 9), ten have equivalent (i.e., proxy) processes available on the USLCI database. Therefore, they are labeled as secondary data gaps* on Table 12 and summarized in Table 13. The remaining 37 data gaps identified by the PCR remain data gaps per this study's definition because no public dataset was identified for these processes.

Table 13 Similar datasets that could be used for Asphalt PCR identified data gaps

PCR Data gap	Identified proxy dataset
Production of 'Binder modifier Extender (BMEx) – petroleum oil'	USLCI: Crude oil production
Production of 'BMEx – Sulfur'	USLCI: Sulfur, at plant
Production of 'Binder modifier plastic (BMP) – Polyethene (PE)	USLCI: HDPE/LDPE, virgin resin
Production of 'BMP – Polypropylene (PP)' Production of 'Synthetic Fiber (SF): PP' *	USLCI: PP, virgin resin
Production of 'BMP – 'Recycled plastic' Production of 'Recycled plastic, dry method' *	USLCI: Rcyl Post-con, PET & HDPE
Production of 'Binder modifier recycling agent (BMRA) – Vegetable oil'	USLCI: Soy oil, refined, at plant
Production of 'SF: Fiberglass'	USLCI: Glass fiber; at plant
Production of 'SF: Polyester'	USLCI: Unsaturated polyester, resin, at plant

**(Note that although BMP-polypropylene and SF-polypropylene are listed as two separate raw materials in the PCR, they are counted once for calculation purposes, as previously mentioned. Similarly, BMP-recycled plastics and recycled plastic-dry method are also counted only once).*

For several raw materials, classifying their production process as a primary data gap is a conservative assumption, as little or no information is known about the process itself. For example, amidoamines or zeolites are a group of chemicals whose general composition might be known but not their production process (e.g., zeolite X, zeolite Y). While any information on material inputs is a starting point for a data gap analysis, such as the one conducted here, without detailed information on what the material is and where it is from, the results are merely estimations.

As previously stated, the PCR identifies the three proxy processes, 'BC of RD,' 'BC of landfill gas,' and 'production of recycled aggregates.' The PCR prescribes a publicly available report for the production of recycled aggregates, but the production process model is not available as a dataset in an LCA database. Thus, the production of recycled aggregates is considered a secondary data gap per this study's methodology. Similarly, the two BC processes were also classified as secondary data gaps because no model for the combustion of these fuels has been identified in the public database, although the processes prescribed by the PCR are considered reasonable alternatives (thus secondary data gaps*). In addition to these three processes, the IC of 'renewable diesel' is also considered a secondary data gap*. The PCR does not label this process as a proxy, despite prescribing the use of fossil fuel-based Diesel combustion instead of RD[14].

There are ten other processes (labeled "Secondary*" on Table 12) that are classified as secondary data gaps because no suitable model has been identified in the public databases. Four of the secondary data gaps are production additives: 'Styrene butadiene styrene (SBS),' 'Styrene butadiene rubber (SBR),' 'ground rubber tire (GRT),' and 'Polyphosphoric acid (PPA)'. As an alternative, the PCR prescribes the production of asphalt binder containing those additives. Although the PCR does not label these as proxies, it is understood that the additive itself does not make the binder and is not equivalent. The reasoning for labeling the remaining six data gaps Secondary* is documented in Table 14.

Table 14 Secondary data gaps whose model was not considered a proxy by the PCR

Data gap	Alternative prescribed by the PCR is considered a proxy, and the process is classified as a secondary data gap because
Production of Asphalt binder with SBS additive	SBR and SBS are different synthetic rubbers, and therefore, asphalt binders modified with SBS (the model prescribed by the PCR) are also expected to have different environmental profiles [41, 42]
Production of Antistrip agent – Hydrated Lime / Mineral Filler – Lime[a].	Quicklime, CaO, is chemically different from hydrated lime, $Ca(OH)_3$.
IC of LNG	The prescribed 'Natural gas, combusted in industrial equipment' does not include the liquefaction process
BC of LNG	As for IC of LNG. In addition, differences are expected between the emissions of IC and BC
BC of RNG	The prescribed process refers to fossil natural gas, but RNG has a different origin
IC of LPG	Differences are expected between the emissions of IC and BC

[a] Note that although Antistrip agent-hydrated Lime and Mineral filler-lime are listed as two separate raw materials in the PCR, they are counted once for calculation purposes, as previously mentioned.

Nine other production or combustion processes that are not labeled as data gaps by the PCR (due to the availability of public reports on the process) are classified as secondary data gaps in this study (marked as "Secondary" on Table 12). This is due to the lack of a dataset in a public LCA database. These nine flows are:

- Production of Crushed aggregates
- Production of Mineral filler: Crusher fines
- Production of Mineral fillers: Slag
- BC of Biodiesel
- BC of brown grease

[14] However, the PCR prescribes the use of BD of diesel, to model the combustion of boiler combusted RD and classifies the usage of BC Diesel as a proxy.

- BC of recycled fuel oil
- BC of yellow grease
- IC of Biodiesel
- Waste and by-products

The PCR prescribes a single public report for 'Crushed stone, sand, and gravel' despite these being three different raw materials with potentially distinct production processes. For that reason, these three materials are separated for the analysis conducted here. The production processes of sand and gravel are not considered data gaps due to a recently released public database on the FLCAC [47, 48]. However, neither this database nor others in the FLCAC include a model for production of crushed aggregates. Thus, it is considered a secondary data gap, as the public report is available to develop the model.

Although a production process for biodiesel exists in USLCI, there is no process for its combustion. Similarly, the combustion process models for yellow and brown greases are currently unavailable on the public LCA database, needing to rely on the reports prescribed by the PCR for their modeling. In all, the production or combustion of 31 materials and fuels are considered secondary data gap (Figure 10) (eight processes mentioned in Table 13, two proxy proxies identified by the PCR, the additional proxy that should have been identified by the PCR, ten processes identified as "Secondary*****," and nine processes, along with one proxy PCR identified proxy process identified as "Secondary" in Table 12). The PCR identified two wastes used as inputs—fly ash and baghouse fines – and identified them as recovered materials. Five additional processes have been linked to these two recovered materials per this study's methodology (Table 12). There are processes in USLCI that account for the production of fly ash ("bituminous coal combustion" and "electricity from bituminous coal" with outputs "Disposal, ash, and flue gas desulfurization sludge, to unspecified reuse").

The potential 'treatment process turning waste fly ash into usable input material' for asphalt is also missing from the database. This treatment process is considered a primary data gap since it impacts the LCA results.

The baghouse fines that are captured in the baghouse may be routed directly to the asphalt production facility or stored in a silo for subsequent use as a mineral filler additive in the mix [49], thus may not need any specialized treatment. The baghouse fines that are not 100 % closed-loop recycled in the plant (i.e., fines designated for beneficial re-use or considered waste) may impact the LCA results [42]. However, the production of baghouse fines is not a data gap, since it is the same process as the production of asphalt mix modeled for the EPD. There are also no disposal processes for either fly ash or baghouse fines; the 'disposal process for fly ash' is considered a tertiary data gap because its absence does not impact the results of the EPD, whereas the 'disposal of baghouse fines' is considered a primary data gap. Typically, the disposal processes for waste generated outside the product system are treated as tertiary data gaps in the current analysis. For example, disposal of fly ash is a byproduct of coal combustion from power plants and is considered the responsibility of the external producer. In the case of baghouse fines, it is estimated that 80 % to 90 % of asphalt producers whose plants are equipped with baghouses attempt to recycle as much of the dust back into their own paving mixes as possible [49].

However, some fraction is landfilled, which is within the system being analyzed [42]. Thus, the disposal of baghouse fines potentially impacts the LCA results, and the absence of a disposal process makes it a primary data gap. When baghouse fines are internally recycled, they are considered burden-free inputs and have no disposal processes.

The 'production of yellow grease' and the 'production of brown grease' are considered tertiary data gaps despite the PCR prescribing a public dataset for both ('Soy biodiesel, production, at plant'). Although these greases may be used for the production of biodiesel, they are not equivalent to soy biodiesel because they are originally waste products [50]. Therefore, grease could be considered recovered materials and burden-free. This is qualitatively different from secondary data gaps for which proxy processes are available. In all, three tertiary data gaps have been identified and can be seen in Figure 10.

However, the 'treatment process turning yellow into combustible fuel' and 'treatment process turning brown greases into combustible fuel' are also missing from the public database, and these treatment processes are considered primary data gaps. Although these treatment processes (of yellow and brown grease) are missing from public databases, they should be included in the EPDs of asphalt per the PCR.

Thus, including the 37 primary data gaps previously identified and the three treatment processes of recovered waste (fly ash, yellow grease, and brown grease), and the disposal process of baghouse fines, there are 42 primary data gaps (Figure 10). In addition to the primary, secondary, and tertiary data gaps, 32 background datasets (production or combustion processes) are readily available on a public LCA database.

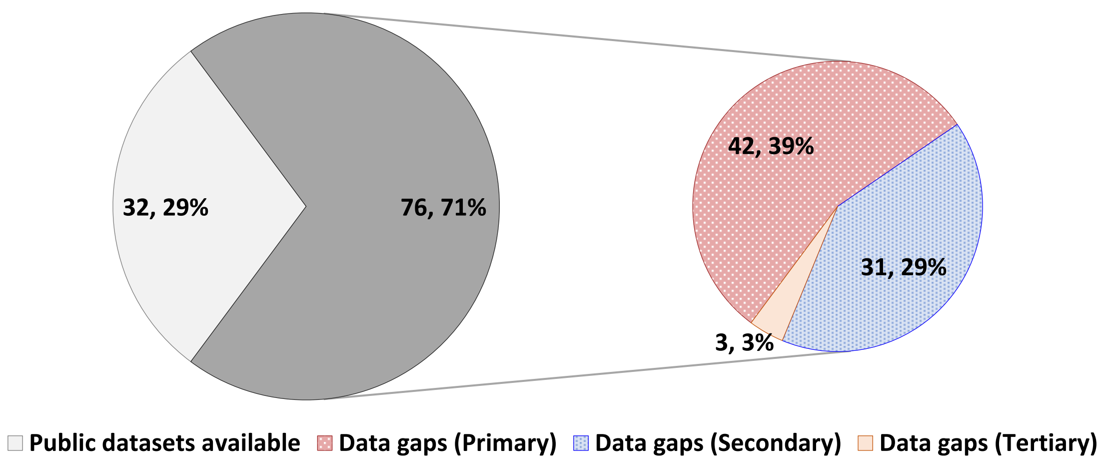

Figure 10 Asphalt PCR data gap assessment - per NIST methodology

3.3.2. EPD data usage compared to the prescribed data sources

The 4048 asphalt EPDs gathered for this study are produced using NAPA's verified EPD generating tool - Emerald Ecolabel. This tool was designed specifically for the asphalt mix production industry to create the most up-to-date supply chain-specific EPDs. The tool was developed in conformance with ISO 14025, ISO 21930, and the PCR for Asphalt Mixtures. Thus, the underlying

LCA modeling method is used for all asphalt mixture EPDs produced with no deviations from the prescribed data sources.

3.3.3. EPD public data gaps

This section summarizes the data gaps based on the processes already identified in the EPDs. Figure 11 shows the frequency of the 27 unique data gaps[15] (across 2953 total instances) that were identified in the Asphalt EPD database.

Of these 27 data gaps, twenty are classified as primary and seven as secondary. The primary data gaps are the different 'AA: Amidoamines,' 'AA: Organo-silanes,' 'AA: Polyamines,' 'WMA: Zeolites,' 'WMA chemicals,' 'WMA-hybrid tech,' 'WMA-wax/fatty acids,' 'Fibers & Fillers: Cellulose,' 'Fibers & Fillers: Baghouse fines,' 'Fibers & Fillers: Aramids,' 'Other Additives: Natural rubber,' 'Other Additives: polychloroprene latex,' 'Other Additives: VTAE,' 'Other Additives: Blends of plastic and rubber polymers,' 'Other Additives: Bio-based oils,' 'Fibers & Fillers: Mineral,' 'Other Additives: Paraffinic Oils,' and 'Binder + TA: three binders added with terminal additives (Table 12)' (labeled with "1" and highlighted in red in Figure 11). 'Polypropylene,' 'recycled plastics,' 'sulfur,' 'polyester,' 'petroleum oil,' 'fiberglass,' and 'vegetable oil-based products' are the seven secondary data gaps (Table 12) (labeled with "2" and highlighted in blue in Figure 11).

Like Amidoamines, which were highlighted in the Section 3.3.1.3, the three binders with terminal additives classified as primary data gaps are a conservative assumption, with little or no information known about the terminal additive itself. The fact that they are additives suggests that they are used in small quantities, with some hint provided about the stage at which they are added to the asphalt mixture. But very little can be understood about their production process; the additives added at the terminal can also be different across different producers. Such larger groups of raw materials limit potential comparisons because their production processes can neither be identified nor quantified. EPDs that include such primary data gaps cannot be compared because the number and kind of additives are unknown.

[15] Note that SF-polypropylene and plastics-polypropylene have been counted only once, as discussed in Section 3.3.1.3, their production processes are assumed to be similar.

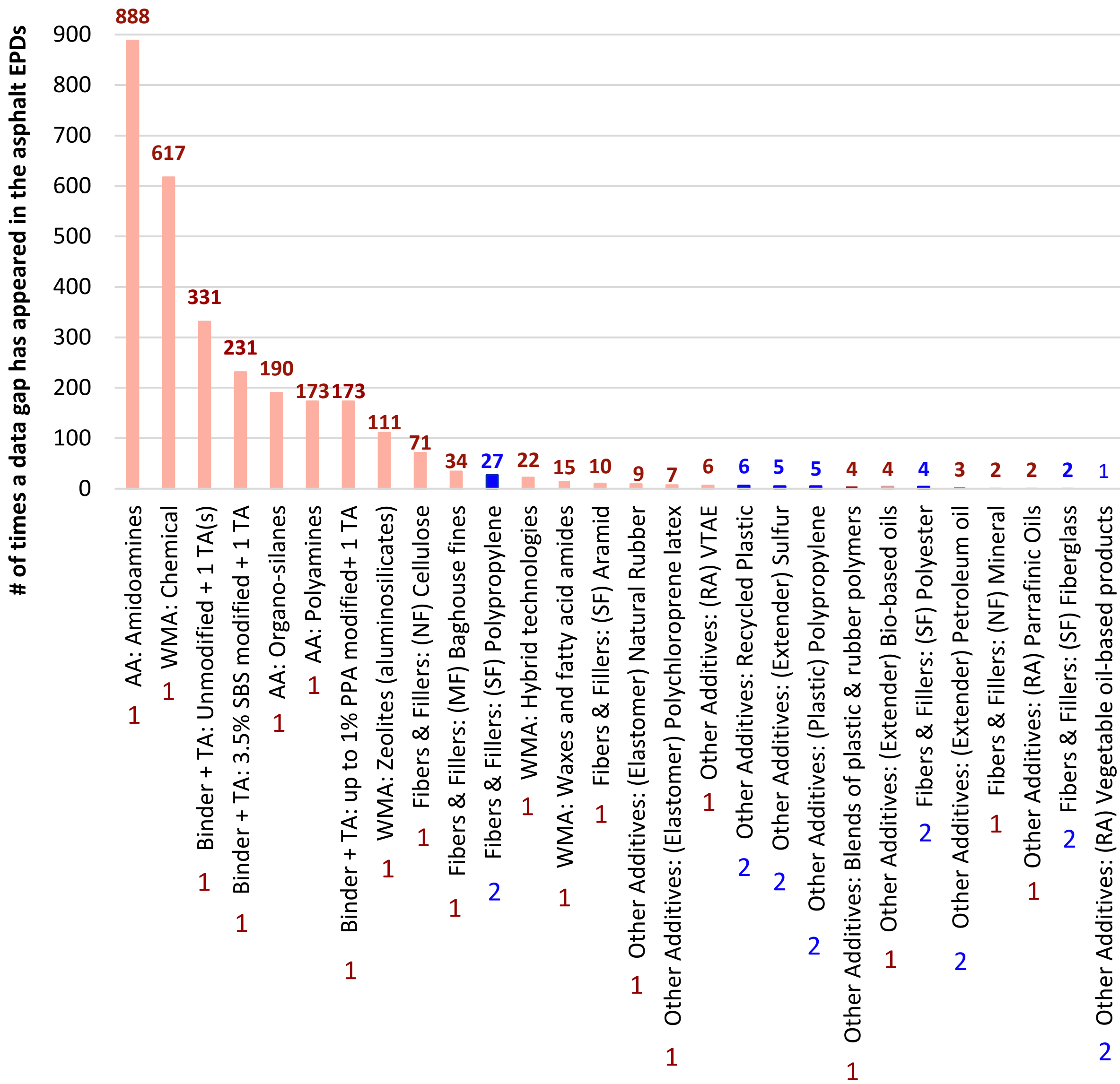

Figure 11 Frequency of EPD-listed data gaps appearing in the EPDs

Note: *1 = Primary data gap; 2 = Secondary data gap*

The data gaps reported in Figure 11 are combined into five general groups in Figure 12. The most frequently reported group of data gaps is 'antistrip agents,' driven by the most common data gap – 'amidoamine' (Figure 12, Figure 11). The next largest group of data gaps is 'Warm mix additives,' of which 'WMA-chemical' is the main representative. The third largest group is modified binders added with terminal additives, for which EPDs report additional terminal additives, but do not explicitly document which additives. 'Fibers, Filler, and Other Additives' are fourth and fifth, and significantly lower in frequency or reporting.

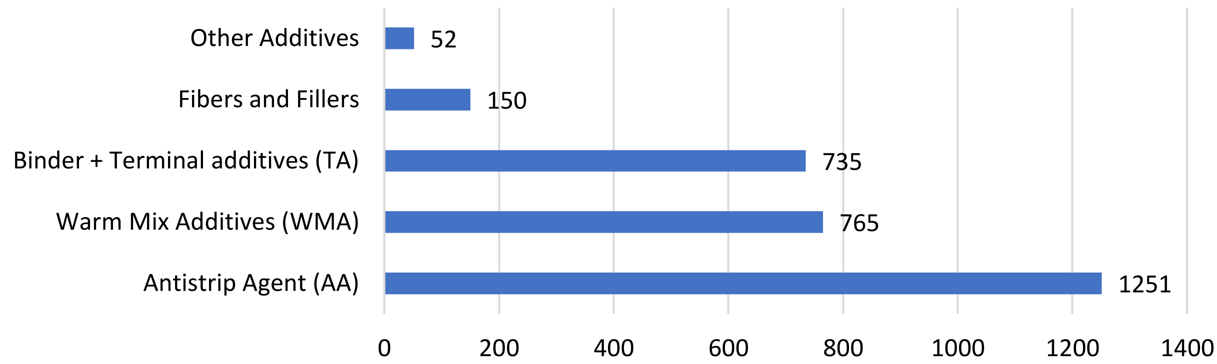

Figure 12 # of times 'commonly occurring groups' of data gaps that are reported on asphalt EPDs

The asphalt PCR states that EPDs are comparable "*if the mixtures are expected to meet similar functional and design performance criteria as specified by the customer,*" but also highlights, "*Comparability may be limited by the presence of data gaps. EPDs with data gaps should not be compared to each other unless the composition and quantity of material ingredients with data gaps are known to be the same for all products being compared.* [39]" Thus, the presence of data gaps must be considered when evaluating the comparability of EPDs. The asphalt EPD database allows for the identification of EPDs with the same data gaps. The detailed reporting of data gaps within all asphalt EPDs provides a unique opportunity to evaluate the comparability of the entire product category.

Of the 4048 EPDs gathered, 2126 (~ 53 %) report no data gaps. Thus, if they meet the functional and performance criteria specified during procurement, these 2126 EPDs are comparable. For the remaining 1922 EPDs, caution should be taken in comparisons. An EPD with one data gap can be compared with another EPD with the same data gap (assuming performance and functional equivalence). The same holds for EPDs with two data gaps, three data gaps, etc. The maximum number of data gaps reported in any single EPD is four, which occurs for nine EPDs (Table 15).

Table 15 # of EPDs reporting data gaps

# of gaps	**# of EPDs reporting the gaps**	**%**
No gaps	2126	52.5 %
1 gap	1060	26.2 %
2 gaps	702	17.3 %
3 gaps	151	3.7 %
4 gaps	9	0.2 %
Total EPDs	**4048**	
Total EPDs reporting gaps	1922	47.5 %

An additional complexity in comparing asphalt EPDs is the reporting of the same process more than once in an EPD. Amidoamine is reported twice in 81 EPDs. Since amidoamine is a class of chemicals, it is likely that when an EPD reports amidoamine twice as data gaps, they are referring to two different chemicals that fall within this class. Similarly, two EPDs reported the use of amidoamines thrice. Like amidoamine, other materials reported twice in a single EPD are 'Warm mix additive (WMA)-chemical' (21 EPDs), 'Elastomer-Natural rubber' (3 EPDs), and 'WMA-hybrid technologies (1 EPD). Future research could leverage the EPD database to complete a detailed evaluation of the comparability of asphalt EPDs based on the reported data gaps to determine current limitations and the potential impact of filling existing data gaps on comparability. Section 4.4.2 will provide some additional details.

For the seven secondary data gaps identified in the EPDs, proxy background datasets have been identified, and these data gaps appeared 53 times collectively in different EPDs. If the proxies suggested in Section 3.3.1.3 were adopted by the PCR, it could reduce the number of times data gaps were reported. Although this would be less than a 2 % reduction in the number of data gaps (2953 to 2900), it would likely be straightforward for the PCR committee to implement.

3.3.4. Additional data gaps identified in the EPDs based on this study's approach

The current section highlights additional data gaps that entail any process not identified as a "data gap" by the asphalt PCR or EPDs but classified as a data gap per this study's approach.

Table 12 identifies additional data gaps in the asphalt PCR based on the current study's data gap definitions (see Section 3). Table 16 highlights all the processes reported in the asphalt EPDs, except for the already identified data gaps listed in the Section 3.3.3.

Table 16 Additional data gaps identified in the asphalt EPDs based on this study's approach

Flow	Classification (this report)	FLCAC process	Comments	Gap Priority
Natural Sone	Available	ANL/C&C: Fine & coarse aggr.		
RAP	Available	US EPA - CDDM: Processing of RAP; at processing		
RAS	Available	US EPA - CDDM: Processing of asphalt shingles, ground, at processing plant		
MF: Milled limestone	Available	USLCI: Limestone, at mine		
MF: PC	Available	USLCI: Portland cement, at plant		
MF: Hydrated lime/ MF: lime / AA: Hydrated lime [a]	Secondary*	USLCI: Quicklime, at plant		2*
MF: Crusher fines	Secondary			2
MF: Baghouse fines [b]	Available		Same process as the production of asphalt mix modeled for the EPD	
Glass cullet	Production: Tertiary			3
	Treatment: Primary			1
	Disposal: Tertiary			3

Flow	Classification (this report)	FLCAC process	Comments	Gap Priority
Ingevity Evotherm M1	Secondary		Chemical WMA [c]; EPD available at [51]	2
Cargill AnovaÂ® 1501 WMA	Secondary		Bio-based WMA [c]; EPD available at [52]	2
Cargill AnovaÂ® 1815 Rejuvenator	Secondary		"Rejuvenators" - special class of RA [c]; EPD available at [53]	2
Ingevity Evotherm P25	Secondary		Liquid AA/WMA [c]; EPD available at [51]	2
Aramid Fiber - SurfaceTech AceXP	Secondary		Aramid fiber [c]; EPD available at [54]	2
PPA Modified - up to 1 % PPA	Secondary*	USLCI: Asphalt binder, 0.5 % PPA, consumption mix, at terminal, from crude oil		2*
PPA Modified - up to 1 % PPA + 1 TA(s)	Primary			1
SBS modified - 3.5 % SBS	Secondary*	ULSCI: Asphalt binder, 3.5 % SBS, consumption mix, at terminal, from crude oil [d]		2*
SBS modified - 3.5 % SBS + 1 TA(s)	Primary			1
Unmodified binder	Available	USLCI: Asphalt binder, no additives, consumption mix, at terminal, from crude oil		
Unmodified + 1 TA(s)	Primary			1
Unmodified + 2 TA(s)	Primary			1
GRT modified - up to 10% GRT	Available	USLCI: Asphalt binder, x% GRT, consumption mix, at terminal, from crude oil, x% GRT	A parameterized model is available on [55]	

Legend	Description	Priority
	Public dataset available and used in all the EPDs	
Primary data gap	Identified as a data gap per this report's classification	1
Secondary Data Gap	Data source prescribed by the PCR, but is classified as a secondary data gap per report, as the prescribed PCR data source is considered a proxy (marked as "**Secondary*** ") or	2*
	Data source prescribed by the PCR, but classified as a secondary data gap per the report because no model is available in a public LCA database (Marked as "**Secondary**")	2
Tertiary data gap	Processes whose results do not impact the LCA/EPD results	3

[a] The production of these raw materials constitutes only one data gap, as the production process is assumed to be the same despite different names/functions ('AA- hydrated lime,' 'MF: Hydrated lime,' and 'MF: Lime')

[b] the baghouse fines used as input material are considered closed-loop recycled within the facility, and hence, there is no subsequent waste treatment process

[c] Although the production processes of the broader groups of additives (WMA, LASA, RA, Aramid fibers) are considered a primary data gap by the PCR (Table 11), some of the asphalt producers use MSP EPDs to model the impacts of the production of some manufacturer-specific additives that have been used as inputs in the asphalt

production. Thus, per the data gap definitions mentioned in section 2.5, the production processes of these manufacturer-specific additives are secondary data gaps.
[d] *Asphalt binder can be modified using polymers such as SBS, Solution SBR, PPA, and GRT* [42]. *But to model the modified asphalt binder with SBS as an additive, SBR is used as a proxy* [41].

In addition to the 27 data gaps highlighted in the Section 3.3.3, 22 additional input materials have been identified in the asphalt EPDs and have been illustrated in Figure 13 [16]. Of these 22 input materials, the production processes for eight are available as background datasets on a public database and are not classified as data gaps (highlighted in gray in Figure 13). These include the production of 'Natural stone,' 'RAS,' 'RAP,' 'Milled/Crushed Limestone,' 'Baghouse fines,' 'Portland cement,' 'GRT modified - up to 10 % GRT' and 'unmodified binder.'

Based on the approach used in this study, the production processes of four terminally modified binders are identified as primary data gaps (labeled with "1" and highlighted in light pink in Figure 13). The rationale for this classification is detailed in Section 3.3.3. These terminally modified binders include:

- PPA Modified - up to 1 % PPA + 1 TA(s)
- SBS modified - 3.5 % SBS + 1 TA(s)
- Unmodified + 1 TA(s)
- Unmodified + 2 TA(s)

Of these four terminally modified binders, the production process for "SBS Modified – 3.5 % SBS + 1 TA(s)" was identified as a data gap 231 times (Figure 11), of which 95 times this input material was not identified as a data gap by the EPDs even though it used the identical nomenclature as times of which it was identified as a data gap by the EPDs. This variation may be a result of use the same nomenclature for different terminal additives used across different EPDs, but not enough information is available to make a determination due to the lack of specificity in reporting. Given this ambiguity, and consistent with this study's methodology, the production process for SBS Modified – 3.5 % SBS + 1 TA(s) will be classified as a primary data gap, regardless of whether all EPDs explicitly reported it as such. Of the remaining twelve input materials, to model the 'production process of crusher fines,' the PCR prescribes the usage of public reports (

Table 12). However, a public dataset modeling the production of crusher fines is not available. Thus, this production process is classified as a secondary data gap.

Similarly, some asphalt EPDs report the usage of manufacturer-specific additives; these inputs are:

- Ingevity Evotherm M1
- Cargill AnovaÂ 1501 WMA
- Cargill AnovaÂ 1815 Rejuvenator
- Ingevity Evotherm P25
- Aramid Fiber - SurfaceTech AceXP

[16] Lime, Antistrip agent Hydrated Lime, and Mineral Fillers Hydrated lime are considered as the same input material performing different function or being addressed differently by different EPDs

Asphalt producers account for the upstream impacts of these materials by using MSP EPDs. Consequently, the production processes of the above-mentioned input materials are not identified as data gaps within the asphalt EPDs themselves. However, under this study's approach, the lack of availability of these production processes as a dataset in a public database classifies them as secondary data gaps. The production processes of these six input materials are identified as Secondary data gaps on Table 16.

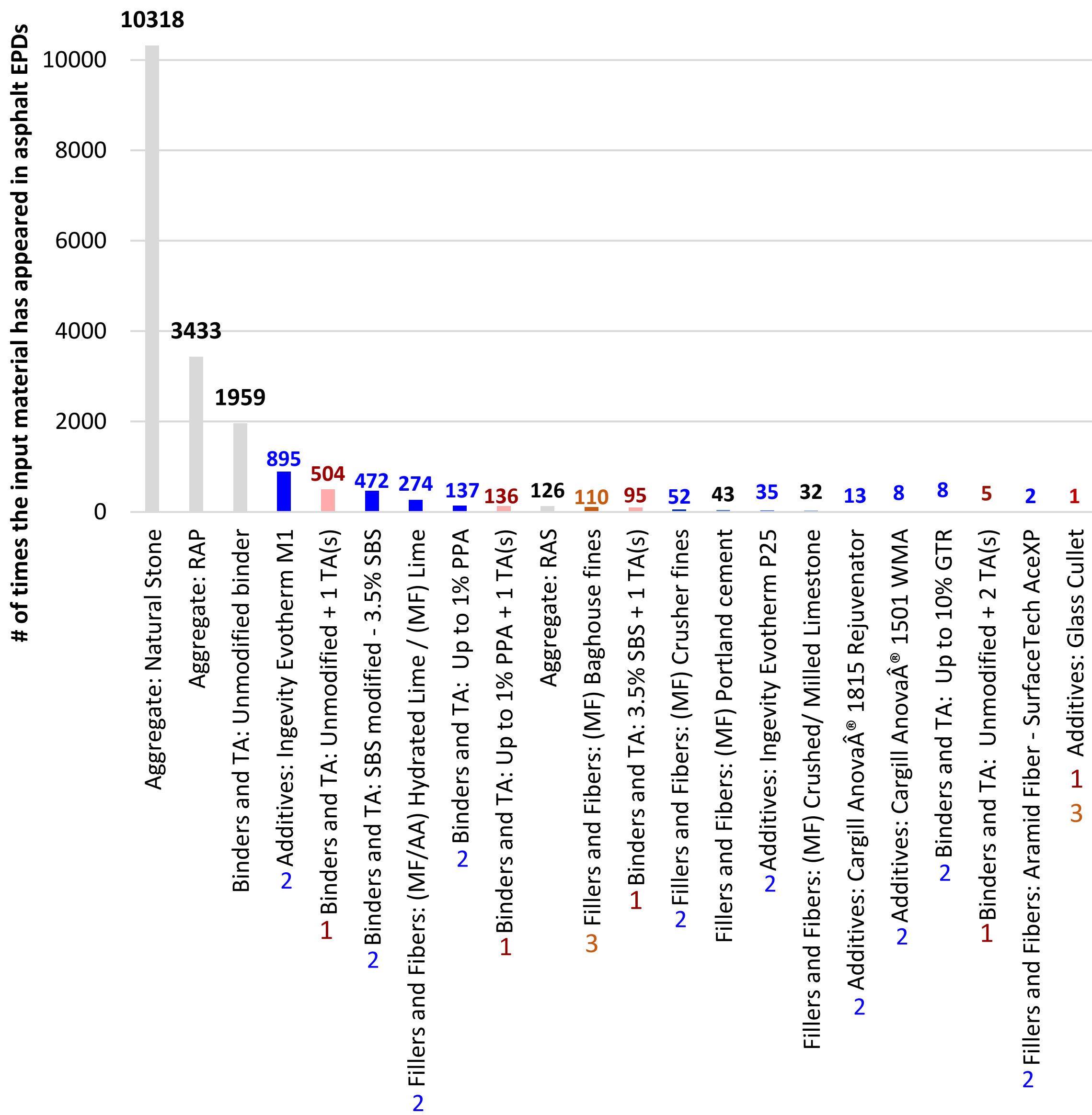

Figure 13 # of times an input material has appeared in the EPD database and its data gap classification

Note: 1 = Primary data gap; 2 = Secondary data gap; 3 = tertiary data gap

Additionally, the production of four input materials has been marked as Secondary* on Table 16. For production of hydrated lime, the PCR prescribes using a USLCI dataset that represents the production of 'quicklime.' Since Quicklime, CaO, is chemically different from hydrated lime, Ca(OH) (Table 14), the prescribed dataset is considered a proxy, and the 'production of hydrated lime' (aka 'lime') is a secondary data gap.

Further, one binder reported in the EPDs—'PPA Modified binder (up to 1 % PPA),'—also presents a secondary data gap. Although public datasets for the production of binders where the concentration of SBS and GRT can be modified exist, there are none for PPA-modified binders. As the percentage of PPA added differs between the publicly available model (0.5 %) and the input binders reported by the EPDs (up to 1.0 %). This existing dataset is considered a proxy, and the production processes for PPA Modified – up to 1% PPA is classified as secondary data gaps (Secondary* on Table 16).

To model the 'production of the SBS Modified – 3.5 % SBS binder,' although a parameterized model exists, SBR is used as a proxy for the SBS additive, as documented in (Table 14) [41, 42]. Thus, the production of the SBS Modified – 3.5 % SBS binder is a Secondary data gap, and the secondary data gaps are labeled with "2" and highlighted in blue in Figure 13.

The last input material, glass cullet or broken glass, is also considered a recovered material [56]. Three additional processes are linked to this input material per this study's methodology (Table 16). No processes account for 'production of glass rejects' in a public database, and since this production does not impact the LCA results, it is considered a tertiary data gap. There are also no 'disposal processes for waste glass'; this is considered a tertiary data gap because its absence does not impact the EPD results. However, the potential 'treatment turning broken glass into usable input' for asphalt manufacturing is missing from the public database, and since this affects the LCA results, it is considered a primary data gap. The tertiary data gaps are labeled with "3" and highlighted in orange in Figure 13.

The 22 input materials identified in Figure 13 and Table 16 are grouped into 4 general groups in Figure 14. The most frequently reported group of input material is 'aggregates,' and the next largest group is 'modified binders added with or without terminal additives'. The third and fourth largest groups are 'Fibers & Fillers' and 'Additives' (Figure 14).

Although

Table 12 identifies data gaps in the fuel combustion and waste management processes, the current assessment of data gap classification for processes reported in EPDs is limited to input materials only because the EPDs only report the material composition of the asphalt mix and not the fuel combustion or waste management processes employed during asphalt manufacturing.

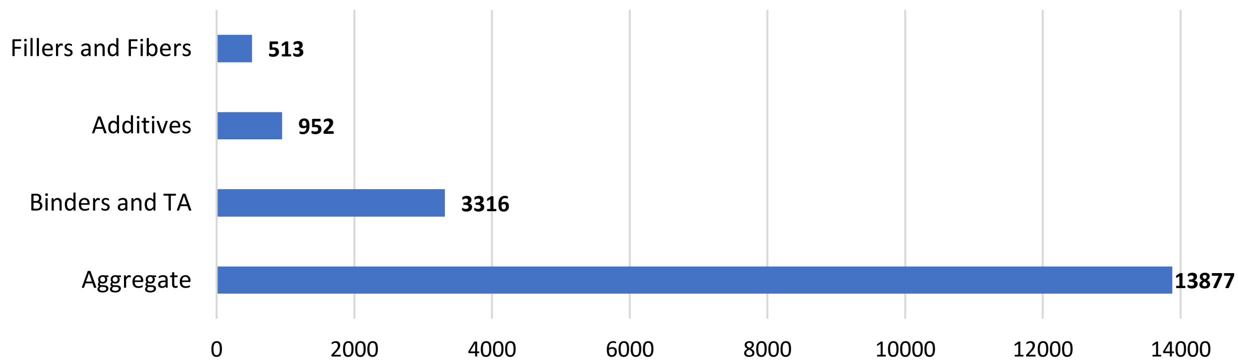

Figure 14 # of times 'commonly occurring groups' of input materials that are reported on asphalt EPDs

In all, 24 processes have been identified (Table 16). Eight production processes are not considered data gaps. Five processes are considered primary data gaps, including the production of four terminally modified binders and one treatment process for glass cullet. The production processes of nine raw materials are considered secondary data gaps. Finally, two processes are considered tertiary data gaps**,** including the production and treatment processes of glass cullet. Figure 15 summarizes the data gaps in all the processes reported in asphalt EPDs (Includes processes from sections 3.3.3 and 3.3.4.)

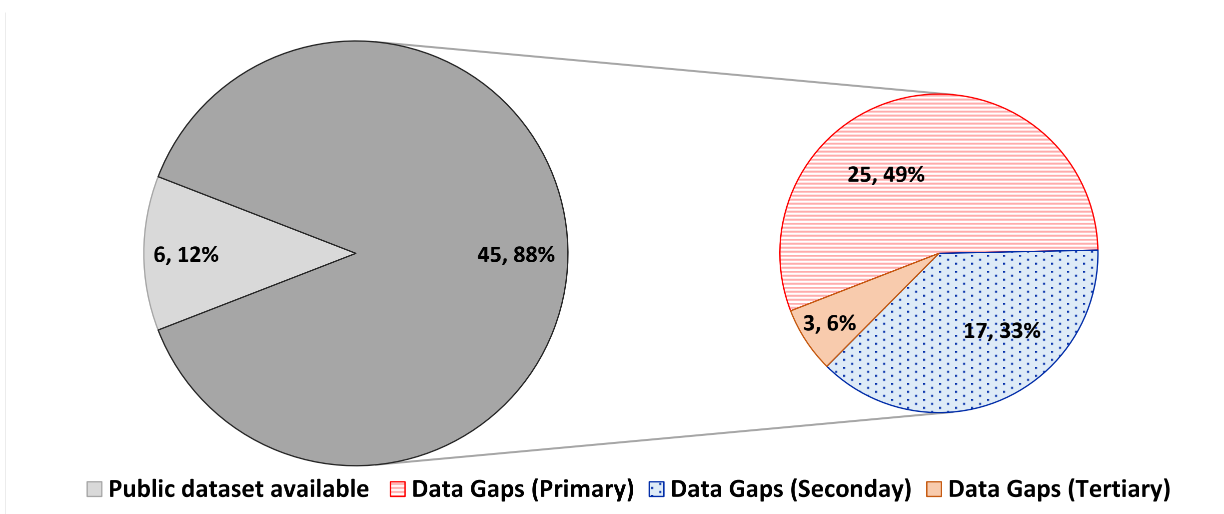

Figure 15 Asphalt EPD data gap assessment - per NIST methodology

3.3.5. Public Data Gap Inferences

Concrete PCR highlights that "EPD comparison based on LCA A1-A3 data shall be made only if the same secondary data sets and all subsequent life cycle stages are equivalent for both EPDs." This method has been adopted by asphalt PCR, bringing asphalt EPDs a step closer to creating "comparable EPDs." The asphalt PCR and EPD generation could serve as a potential example and pilot for other construction product categories that currently do not have any data prescriptions or are trying to improve their existing data prescriptions. This way, EPDs designed in conformance with specific programs are comparable if the products are expected to meet similar functional and design performance criteria as specified by the customer, and they either have no data gaps or have the same data gaps across the EPDs being compared.

3.4. Gypsum

Gypsum EPDs are required to conform to the PCR for Gypsum Panel Products v1e [57]. The products covered under the scope of the PCR include interior and exterior gypsum panel products. In June 2024, a total of 104 EPDs were manually downloaded from the EC3 website. In August 2024, an additional dataset was created using openEPD files, which were retrieved through EC3 API calls, resulting in the extraction of 107 openEPD records into a CSV file. Although capable of including such information, openEPD files typically do not currently contain information from qualitative fields such as the LCA data sources, cut-off criteria, and system boundary definitions. This missing information was directly extracted from the 104 EPD PDFs (both through manual and automated approaches) and compiled into a separate CSV file. The two datasets were merged into a single combined CSV file using 'declaration number' as the primary key variable. This merged dataset contains products from 84 EPDs that were present in both the directly extracted and API-based openEPD datasets. The 84 EPDs were analyzed for further analysis and will be detailed in the upcoming sections. Further information about the data download and database creation can be found in Section 2.1.3 and Section 2.2.

3.4.1. PCR public data gap assessment

The Gypsum PCR does not prescribe or recommend datasets for LCA modeling beyond requiring the EPDs to specify the data sources, including the source database and the year of publication[17]. Thus, a data gap analysis on the PCR prescribed/recommended data has not been conducted. The PCR requires the usage of specific data derived from specific production processes to be the first choice for LCA modeling. ISO 21930 defines specific data as the "*data representative of a construction product or service provided by one supplier, either from multiple plants or based on multiple similar construction products from the supplier* [13]."

3.4.2. EPD LCI data reporting

Three POs have produced EPDs for the gypsum product category, each with different EPD look, formatting, and typesetting: ASTM International, UL Environment, and NSF. Table 17 shows that 40 % of EPDs are developed by ASTM International, 8 % by NSF, and 52 % by UL Environment. 78 EPDs disclosed the sources of data used in the EPD preparation in the LCA report; these were mostly prepared by UL Environment and ASTM (although the LCAs are not publicly available).

Databases used in gypsum EPD LCA modeling included GaBi Thinkstep, USLCI, and EI (including US-EI). All EPDs except one disclosed the databases used in the LCA modeling (Table 17). The version of the database used varied across EPDs. 35 EPDs relied on a single database (GaBi or

[17] Section 7.1.9 of the gypsum PCR says **"All data sources shall be specified, including source database and year of publication."**[57]

Sphera LCA for Experts databases), while the remainder reported using multiple databases (Table 17). All the EPDs reported the year of specific data used in the LCA modeling.

Table 17 Database usage in different gypsum EPDs

PO	Total EPDs		Database Usage		Disclose	Data sources
		Multiple	Single	Undisclosed	Yes	No
ASTM	34	0	34	0	34	0
NSF	7	6	1	0	1	6
UL	43	42	0	1	43	0
Total	**84**	**48**	**35**	**1**	**78**	**6**

3.4.2.1. Other reporting details in EPDs

The PCR encourages "*the creation of cradle-to-grave LCA when feasible, but also recognizes Cradle-to-Gate (C2Gate) LCA with other optional life cycle stages*"[57]. All EPDs produced by NSF have a C2Gate boundary, while all but one EPD developed by UL Environment have a cradle-to-grave (C2Grave) system boundary. For EPDs developed by ASTM International, four are C2Gate EPDs while the remainder are C2Grave EPDs (Figure 16).

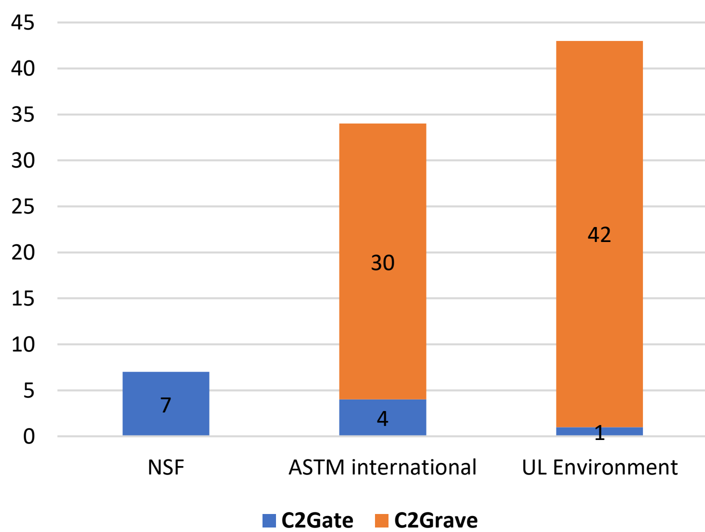

Figure 16 # of EPDs developed by each POs and their respective system boundaries

Twelve of the 84 EPDs (ten EPDs developed by ASTM International and two EPDs developed by NSF) have declared information for multiple products. Of the twelve, six are C2Grave, all from ASTM International, and six EPDs are C2Gate, of which four are from NSF and two from ASTM International (Table 18).

Table 18 Reporting types in Gypsum EPDs

PO	Total		System	Boundary	
	EPDs	Cradle	to Grave	Cradle	to Gate

		Multiple products	Single products	Multiple products	Single products
ASTM	34	6	24	4	0
NSF	7	0	0	2	5
UL	43	0	42	0	1
Total	**84**	**6**	**66**	**6**	**6**

3.4.3. EPD Data Gap Assessment

Although the datasets used in the LCA modeling of the gypsum EPDs are not publicly disclosed within the EPDs, all EPDs disclose the constituents of the gypsum board. Table 19 collates the material composition information available from the gypsum EPDs, and a data gap assessment is conducted based on the information gathered. Energy usage and waste management are excluded from this assessment due to the lack of information. For the same reason, no evaluation of the use of public datasets could be completed.

Table 19 Gypsum Data gap assessment based on the EPD material content disclosure

	Flow	FLCAC equivalent	Additional processes	Comments	Priority
Raw Material: Gypsum	*Gypsum /virgin natural gypsum)/ Natural gypsum ore*	ANL C&C Gypsum; at quarry			
	Internally recycled Gypsum	Available	Production of gypsum board	rejects and off-spec materials generated during the manufacturing of gypsum products that is internally recycled	
		Available	Processing: larger part of gypsum board production		
	Synthetic Gypsum / FGD gypsum	Production			3
		Processing			1
		Disposal			3
	Post-consumer gypsum/Natural Gypsum, reclaimed	EPA CDDM: Processing of gypsum; milled; at drywall processing facility			
Raw Material: Paper	*Paper*	USLCI: Paper, freesheet, coated, avg production			
	Paper Fiber	Data gap			1
	Gypsum paper / paper facings/ Backing Paper / Facing Paper	Data gap			1
	Recycled Paper (100 % recycled content)	Recycled paper production [a]	Production		3
			Processing [a]		1
			Disposal		3
	Adhesive, tape	Data gap			1
	Adhesive, glue	Data gap			1

	Flow	FLCAC equivalent	Additional processes	Comments	Priority
Water	*Water*	Data gap			1
Raw Material: Additives	*Additives, dry*	Data gap			1
	Additives, wet	Data gap			1
	Fungicide	Data gap			1
	Coatings	Data gap			1
	Starch	Data gap			1
	Silicone	Data gap			1
	Vinyl	Data gap			1
	Wax	USLCI: Slack wax			
	Vermiculite	Data gap			1
	Dispersant	Data gap			1
	Retarder	Data gap			1
Boards	*MgO board*	Data gap			1
	Gypsum core board	USLCI: Gypsum wallboard product, type X, 0.625 inch (15.875 mm)			
Reinforce-ment	*Thin layer of steel*	USLCI: Cold rolled sheet. steel			
	Continuous filament glass fiber / glass mat / fiberglass	USLCI: Glass fiber; at plant			
	Scrim	USLCI: Glass fiber; at plant		At least a proxy	2*
Others	*Other materials*	Data gap			1

Legend	Priority
A public production process is available	
Primary Data gaps: Processes not available as a public data set	1
Secondary data gaps*: Data source identified as FLCAC equivalent is considered a proxy	2*
Tertiary data gaps: Processes whose absence does not affect the results of the EPD	3

[a] *the production process of recycled paper is equivalent to the treatment process of turning waste paper into gypsum paper and this process is counted once.*

Based on the information available on material composition, gypsum EPDs have reported 28 input materials or material groups. Some input materials have been named differently across EPDs, but use the same production processes and can be grouped together. For example, Natural gypsum ore and Virgin Natural gypsum both represent Natural gypsum. Synthetic gypsum and FGD gypsum both represent byproducts of flue gas desulfurization (FGD) systems used in coal-fired power plants to remove sulfur dioxide emissions.

The production processes for seven material inputs are publicly available and are not identified as data gaps: 'Gypsum/ Natural gypsum ore/ virgin natural gypsum,' 'post-consumer gypsum/

Reclaimed natural gypsum,' 'Paper,' 'Wax,' 'Gypsum core board,[18]' 'Thin layer of steel,' and 'Continuous filament glass fiber/ glass mat/ fiberglass/ glass matting.'

Of the remaining 21 raw materials, the production processes of seventeen input materials are not available on a public database and are classified as primary data gaps. These seventeen raw materials are 'Paper fiber,' 'Gypsum paper/paper facings/ Backing Paper / Facing Paper,' 'Adhesive – tape,' 'Adhesive – glue,' 'Water,' 'Additives – dry,' 'Additives – wet,' 'Fungicide,' 'Coatings,' 'Starch,' 'Silicone,' 'Vinyl,' 'Vermiculite,' 'Dispersant,' 'Retarder,' 'MgO board,' and 'Other materials.'

For several raw materials, classifying their production process as a primary data gap is a conservative approach because little or no information is known about the process itself. "Other materials" is the least well-defined input because the number of raw materials, function, composition, and production process are not specified. The same could be said of (dry and wet) "additives." Although being labeled as additives provides some general aspects of their physical properties, it suggests these inputs are used in small quantities. In the case of adhesive (tape/ glue), fungicide, coatings, dispersant, and retarder, only their function is provided. Therefore, EPDs may use different raw materials with different production processes for the same function. Finally, there are gaps such as "starch," whose general composition may be known, but not its production process (e.g., corn starch, potato starch, etc.). While any information on material inputs is a starting point for a data gap analysis, without detailed information on the material and its source, results will remain general approximations.

Of the remaining four raw materials, the production of one input material, 'scrim,' is considered a secondary data gap* because a proxy production process is available on the public database. The remaining three input materials—'recycled paper (100 % recycled content),' 'FGD gypsum/synthetic gypsum,' and 'internally recycled gypsum'—are considered recovered materials. Each of these materials is associated with two to three process types: production, treatment, and disposal.

Internally recycled gypsum is the rejects and off-spec materials generated during the manufacturing of gypsum products that are recycled within the producing facility (analogous to baghouse fines in asphalt). Therefore, their production is not a data gap, as it is the same process that is being evaluated in the EPDs. The treatment process required to turn the gypsum retrieved into usable input material for gypsum boards is also part of the larger process of gypsum board production. Therefore, it is also not a data gap. Thus, these nine processes are not considered data gaps.

Similarly, there are no processes in a public database that account for production of FGD gypsum. Thus, the 'production process for FGD gypsum' is considered tertiary data because its absence from the public database does not affect the results of the EPD. There are no 'disposal processes for FGD Gypsum'; the absence of the disposal process for FGD Gypsum from the public database does not affect the results of the EPD. The potential 'treatment process turning waste FGD

[18] Six products, all reported in the same EPD, include gypsum core board as an input because the EPD covers the production of sound-reducing drywalls that consider gypsum board a raw material to which a series of layers—steel plate, MgO board, etc.—are added.

gypsum' into usable input material (i.e., dewatering of FGD slurry and any additional drying) for gypsum board is also missing from the public database and is considered a primary data gap, as it impacts the results of the LCA. Although this treatment process is missing from public databases, it should be included in the EPDs of gypsum per the PCR.

Recycled paper (100 % recycled content) could be considered a recovered material if the gypsum manufacturer recovers wastepaper that becomes an input in gypsum manufacturing. In this case, the 'production process of [waste] paper' that is used as input material is absent and considered a tertiary data gap because its production does not impact the LCA results in the EPD. The 'disposal process of the wastepaper' is also considered a tertiary data gap because its absence from the public database does not affect the results of the EPD. 'The treatment turning wastepaper into usable input' for the gypsum board is also missing from the public database and considered a primary data gap because it impacts the LCA results. Although this treatment process is missing from public databases, it should be included in the EPDs of gypsum per the PCR. In the case where the gypsum producer procures the recycled paper instead of producing the required paper using waste, only the production process is required to be considered in the LCA. This production process is equivalent to the treatment process that converts the wastepaper into usable gypsum paper, and thus, is a primary data gap. This production process of recycled paper is not counted as a separate data gap because the treatment process of wastepaper is an equivalent process to the production of recycled paper, and the treatment process of wastepaper already accounts for the primary data gap.

Figure 17 Illustrates the data gap assessment for gypsum EPDs. In all, 27 data gaps have been identified. Nine processes, seven production processes for raw materials, and two production and treatment processes for internally recovered gypsum, for which datasets are publicly available, are not considered data gaps. There are 19 primary data gaps, including the seventeen production processes of raw materials mentioned above and two treatment processes required to turn recovered material into usable input. One production process is considered a secondary data gap because of the availability of proxy processes on the public database. Four processes are considered tertiary data gaps: production processes of FGD gypsum and recycled paper, and disposal processes of FGD gypsum and recycled paper.

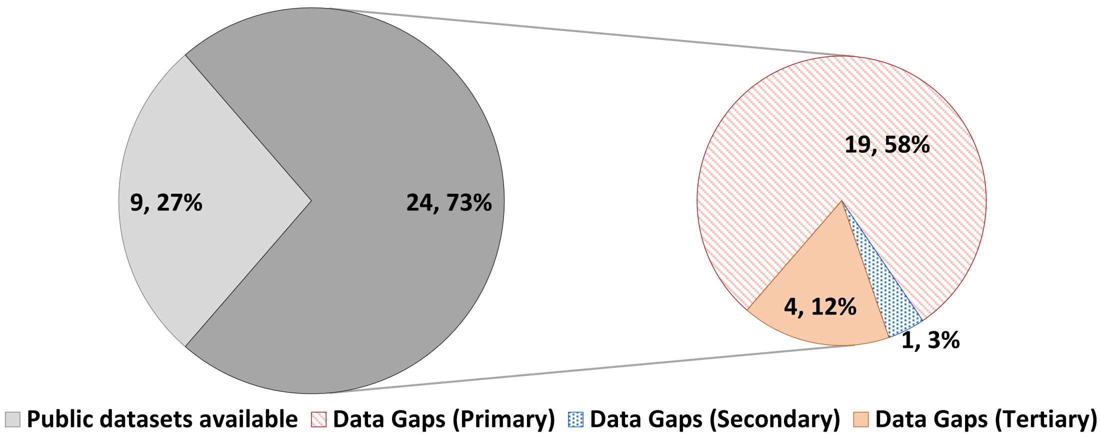

Figure 17 Gypsum EPD data gap assessment

3.4.4. Public Data Gap Inferences

All EPDs except one disclose the databases used in the LCA modeling, but none disclose specific background datasets in the public-facing EPDs, limiting the ability to generalize the assessment conducted in this study.

Currently, 76 % of the processes identified in Section 3.4.3 are not available as a public dataset. For raw materials, 21 % of production processes are publicly available, but because EPDs do not disclose the LCI, it is infeasible to determine whether the identified public datasets were actually used in the LCA modeling. To further the background data harmonization in gypsum EPDs, a critical first step would be to ensure that the background data being used is reported. Additionally, separating multiple products currently grouped together under a single EPD into multiple EPDs, each covering a single product from a single plant, will improve machine-readability for future analysis.

4. Conclusions

This study developed and implemented a process to develop machine-readable databases of information reported in EPDs validated and applicable to the U.S. for four product categories: concrete, cement, asphalt, and gypsum. The databases were used to analyze the reported LCI data used for all EPDs in a product category to complete a public LCI data gap assessment. The results of the LCI data analysis were also used to develop some representative models in NIST TN 2338, following common practices in current EPDs [24].

4.1. PCR Requirements and EPD Database Creation

The approach implemented to develop these databases varied by product category due to differences in the number of available EPDs, sources, and file formats of EPDs, and the level of standardization in information reporting and formatting in EPDs. Asphalt is the most standardized, both in formatting (using HTML files) and LCI data source selection through an automated EPD generator tool. Additionally, identified data gaps are reported in each asphalt EPD. Similarly, concrete EPDs use a standardized format (using PDF files), report LCI data sources, and the PCR also prescribes background data sources for various processes. Meanwhile, cement and gypsum EPDs are not standardized in format and provide less, or even no, information on LCI data sources. The respective PCRs also do not prescribe background data sources for processes. Additionally, cement EPDs and a handful of gypsum EPDs report multiple products for a given facility in a single EPD, and this poses a significant challenge in automating the generation of machine-readable documents using the methods mentioned in Section 2.

Due to these differences in PCRs and published EPDs, a different approach was implemented to create the EPD database for each product category. The concrete database is based on an existing database created using scripts. The existing database was (minimally) validated manually by NIST, and any modifications completed are directly updated in the EPD database spreadsheet. The cement database was developed through manual entry of information directly from the EPDs into the spreadsheet and validated during analysis. The asphalt database used scripts to automate the creation of the spreadsheet by hard-coding the extraction of each line of information in the EPD and populating the spreadsheet. The gypsum database was developed using scripts to convert the openEPD file based on an EPD to automate the creation of the spreadsheet. Then, a combination of Azure's Document Intelligence and manual processes was used to extract material content from the EPD PDFs and appended it to the spreadsheet using scripts. Additional manual information extraction was also used in finishing the gypsum EPD database to have all the information required for the analysis. This approach provided key insights into the machine-readability of current EPDs and how to automate the extraction of information from non-standardized EPDs, including those reporting multiple products.

Given the lack of standardization both within and across most product categories and the limitations in information currently reported in openEPD files, it was determined that developing and updating machine-readable EPD databases may require the leveraging of machine learning (ML) to extract information directly from EPD source files. This will require developing prompt engineering and training an LLM model using a range of EPD formats across a range of product categories.

4.2. Data Source and Gap Assessment Results

Given the variability in EPD formats and LCI data disclosure, the data source, and data gap assessment results, interpretations and next steps vary by product category. The product categories that are further along in standardization can provide a path forward for those that have not progressed as far to date.

Concrete EPDs report consistent selection of data sources for nearly all processes, in part driven by prescriptive data source requirements. The PCR also includes fifteen processes for which public datasets were determined to be appropriate. However, the EPDs have also reported input materials/energy sources whose production/generation processes are modeled using proprietary data sources. These input material/energy sources include: 'recycled concrete aggregates,' 'crushed aggregates,' 'lightweight aggregates,' 'tap water,' 'natural aggregates,' 'carbon black pigments,' 'dolomite,' 'microfibers,' 'citric acid,' 'phosphoric acid,' 'non-hazardous municipal solid waste disposal,' 'oil/lubricants/grease,' 'limestone' 'metakaolin,' 'blend of aggregates' and 'medium voltage electricity'. Some of these raw materials' production processes are already available on the public database (lubricant feedstock, limestone). Additionally, newer versions of the public database have started including datasets for raw materials production processes (e.g., aggregates) and energy sources (e.g., electricity).

Production processes for some raw materials (e.g., binder and admixtures) are mostly modeled using supplier-specific data or industry average data by most EPDs. However, a representative model of these processes is not available on the public database (except for Portland cement). These representative models (like an industry average LCA report converted to an LCA dataset), could be made available and prescribed as a default data source when a supplier-specific data source is not available.

One potential next step toward background data harmonization for concrete EPDs could focus on two priorities. First, clarifying the precedence of data sources where there is an inconsistent or unclear ranking of data sources. Second, providing data source prescriptions for processes that are extensively used in EPD generation but currently lack a data source prescription in the concrete PCR. Efforts to fill the data gaps could begin with creating publicly available datasets for production or treatment processes currently unavailable in a public database. Next, efforts could focus on converting reports into models within a public database. Consolidating all the prescribed data sources within the public database(s) would provide a unified and transparent repository. Additionally, improving the interoperability of public databases would increase the ability to prescribe public data sources where available.

Cement EPDs report significantly more variability in data source selection and reporting. At least six reported input materials' production processes are currently unavailable on a public database: 'high-voltage electricity,' 'glycine,' 'clay,' 'silica sand,' 'iron ore,' and 'tap water'. A few other input materials' production processes are also modeled using supplier-specific data sources (e.g., grinding aids and clinker), whose representative models are not available as public datasets. Also, some processes in cement EPDs have been reported as modeled using different datasets: ‘train transport,’ ‘barge transport,’ ‘high-voltage electricity,’ ‘natural gas,’ ‘gypsum,’ ‘limestone,’ and ‘hard coal/anthracite’. To further the background data harmonization in cement EPDs, the next initial steps could be (1) for the PCR to provide background data prescriptions to reduce variability in data source selection, and (2) require EPDs to report background data sources being used. Additionally, to ease machine-readable digitized EPDs, the PCRs could encourage separating multiple products into separate EPDs. As with concrete, efforts to fill the data gaps could begin with creating publicly available datasets for production or treatment processes currently unavailable in a public database.

Of the product categories considered in this study, asphalt EPDs provide the greatest transparency and comparability of EPDs. This is enabled by prescribing publicly available free LCI data sources where feasible. The asphalt PCR also identifies LCI data gaps and proxies, most of which are related to asphalt additives (e.g., antistrip agents, warm mix additives, asphalt binders, and terminal additives). The current study identifies additional data gaps due to the differences in the applied methodology. For example, some fuel sources identified in the PCR that are not listed as proxies by the PCR have been categorized as proxies for various reasons. This study identifies proxy datasets for some of the processes that are identified as data gaps by the asphalt PCR. If the proxies suggested are adopted by the PCR, it could reduce the number of times data gaps were reported (although minimally). As with concrete and cement, efforts to fill the data gaps could begin with creating publicly available datasets for production or treatment processes identified as data gaps either by the PCR or this study. The asphalt PCR and EPD generation implementation can serve as the basis for best practice for other product categories to replicate.

Gypsum EPDs do not report data source selection, but the material content is reported, which can be used as the basis for the LCI data gap assessment. The production processes for seventeen input materials are not available in public LCI databases, the most noteworthy can be grouped into paper, adhesives, and additives. However, it is not feasible to identify the actual process selected or what variability there was in data source selection across EPDs due to the lack of data source reporting. To further the background data harmonization in gypsum EPDs, a first step could be to require reporting of LCI data sources. Additionally, creating separate EPDs for each product could also ease the machine readability. Efforts could begin to fill the data gaps identified in this study until LCI data source information becomes available, at which time an enhanced data gap assessment could be completed to further refine the priority data gaps.

4.3. Limitations

There are limitations associated with both the EPD database development implemented and the data gap assessment completed in this study. The most prominent limitations are highlighted and generalized below. This is not a comprehensive list, and other limitations have been mentioned throughout the document where appropriate.

There are several limitations relative to the data gap assessment. First, the data gap priority categorization is just one of many potential methods to identify and rank LCI data gaps and includes subjective decisions (e.g., recovered materials production, treatment, and disposal) that were not always unanimous amongst the authors, but necessary to create the groupings. Second, this study does not quantify the impact of data gaps on the LCA modeling results or evaluate the data quality (e.g., vintage, representativeness), limiting the comprehensiveness of the assessment. Third, the lack of and variability in disclosure of LCI data sources limit the completeness of the assessment. Fourth, this data gap assessment is a snapshot in time. Any new EPDs, updates to existing EPDs, or updates to the PCR since that date would not be included in the data gap assessment. For example, the new concrete PCR is scheduled for release in 2025 and could address some of the public data gaps discussed in this study. Finally, this study does not include NIST efforts associated with NIST TN 2338 [24] that has begun to address some of these data gaps.

Limitations related to the EPD databases include the following. First, the databases were not released in a standardized format due to varying amounts and types of information provided in EPDs both within and across product categories, as well as the varying approaches implemented to create each database. Second, although some validation was completed for the databases, errors could remain, which could hinder the associated data gap assessment. The analysis for concrete assumes that the existing concrete EPD database from Ref. [25] has been previously validated and is used as-is except for the manual adjustments discussed in Section 2.2. Similarly, the manual steps in the process could have introduced mistakes in the data. Third, as was mentioned under the data gap assessment limitations, the databases were developed at a moment in time. Any new EPDs or updates to existing EPDs since that date would not be included in the data gap assessment. For example, asphalt EPDs are updated frequently, potentially leading to changes in some or all EPD information.

4.4. Future Work

This study provides both a high-level analysis focused on completing a public data gap assessment and publicly available EPD databases that can be used to complete additional detailed analysis and increase transparency and comparability of construction material LCA modeling. The data gap assessment could be used in combination with NIST TN 2338 [24] and other LCI data gap assessments to assist in prioritizing public LCI datasets and model development, targeting those that would be the most impactful to the LCA community, the construction sector, and the U.S. economy at large. The EPD databases could be used to evaluate a range of potential research questions beyond the LCI data source analysis completed in this study, whether it's comparing EPDs within a product category (e.g., comparability of and variation in reported impacts across EPDs for a specific location) or across product categories (e.g., EPDs throughout the concrete supply chain). Fully standardizing the database creation process and analyzing the information could enhance the quality of those databases. Some of this future work is discussed in more detail below.

4.4.1. Standardized, Automated, Machine-Readable EPDs and Databases

Given the shift from PDFs towards machine-readable digital formats, as exemplified by the asphalt industry's EPD generation process [26]; there will be a transition away from PDFs as the single point of truth to alternative machine-readable formats (e.g., JSON and CSV). However, during this transition period, existing (valid-as-of-date) EPDs currently in PDF format need to be converted into machine-readable formats to make the information in the EPDs easier to compare.

Therefore, NIST is currently developing an approach to fully automate (1) the machine-readable digitization of EPDs and (2) the creation of EPD databases by leveraging LLMs and an existing standardized data format (openEPD). This study identified issues with converting information in non-standardized PDF files into a standardized machine-readable format. The approach will use prompt engineering and openEPD templates to convert PDFs into an openEPD file that includes all information in the EPD, and then compile those openEPD files into a database to allow for analysis. Validation will be completed using the existing databases. The asphalt EPD database would be the initial database recreated because of its standardized EPD format and availability of openEPD files for all new EPDs. The approach would be further refined and applied to the remaining databases in the following order based on expected complexity: concrete, cement, and gypsum. The end goal is to have a process that can be implemented to convert any existing EPD into the openEPD format and append that information into the associated EPD database.

4.4.2. Potential Future Analysis using EPD Databases

The EPD databases could be used for a wide range of analysis beyond the focus of this study – LCI data sources and public LCI data gap assessment. A few potential uses include the following:

- Calculate distributions of impact category values
- Filtering and comparing material content
- Determine how often specific processes are reported
- Compare modeling assumptions to ensure consistency
- Evaluate the comparability of EPDs based on identified data gaps
- Assess the impact of specific requirements in a PCR

Some decision-support tools (e.g., EC3) provide some of these capabilities, such as filtering and comparing products valid in the same region. However, none provide access to all information published in the EPDs. By publishing the EPD databases, researchers can pursue these and other not yet identified analyses both within and across product categories. For context, two of these examples are discussed below, one for analysis within a product category and another for analysis across product categories.

As mentioned in Section 3.3.3, the asphalt PCR states that EPDs are comparable "*if the mixtures are expected to meet similar functional and design performance criteria as specified by the customer,*" but also highlights, "*Comparability may be limited by the presence of data gaps. EPDs with data gaps should not be compared to each other unless the composition and quantity of* *material ingredients with data gaps are known to be the same for all products being compared.*" The asphalt EPD database could be used to assess comparability based on the identified data gaps as reported in the EPDs.

One type of analysis across product categories is to evaluate the consistency of LCA modeling choices throughout a product's supply chain. For example, cement is the primary input and source of embodied emissions in concrete, and the concrete PCR prescribes the use of cement EPDs, which were shown to have significant variability in this study. Thus, increasing consistency in LCA modeling and data selection between the two product categories could align methodologies, improve transparency, and foster standardization in EPD preparation processes.

References

[1] Frey S, Bar Am J, Doshi V, Malik A, Nobel S (2023) Consumers care about sustainability—and back it up with their wallets. Available at https://www.mckinsey.com/industries/consumer-packaged-goods/our-insights/consumers-care-about-sustainability-and-back-it-up-with-their-wallets#/

[2] AzariJafari H, Rangelov M, Gregory J, Kirchain R (2023) Suitability of EPDs for Supporting Life Cycle and Comparative Analysis of Concrete Mixtures. *Environmental Science & Technology* 57(19):7321–7327. https://doi.org/10.1021/acs.est.2c05697

[3] City and County of Denver (2018) *Green Buildings*. Available at https://www.denvergov.org/Government/Agencies-Departments-Offices/Agencies-Departments-Offices-Directory/Community-Planning-and-Development/Plan-Review-Permits-and-Inspections/Commercial-and-Multifamily-Projects/Green-Buildings-Ordinance

[4] Commonwealth of Pennsylvania Part I Chapter 22 “Green” Procurement. Available at https://www.pa.gov/content/dam/copapwp-pagov/en/dgs/documents/documents/procurement-forms/handbook/pt1/pt%20i%20ch%2022%20green%20procurement.pdf

[5] Environmental Product Declaration (EPD) Planning for the Future: Key Trends 2025-2033 (2025) Available at https://www.marketreportanalytics.com/reports/environmental-product-declaration-epd-54085

[6] ISO 14040 (2006) ISO 14040: Environmental Management Life Cycle Assessment – Principles and Framework. (International Organization for Standardization (ISO), European Standard, Geneva. Switzerland). Available at https://www.iso.org/

[7] ISO 14044 (2006) ISO 14044: Environmental management: life cycle assessment; requirements and guidelines. (International Organization for Standardization (ISO), European Standard, Geneva. Switzerland). Available at https://www.iso.org/

[8] International Organization for Standardization (2022) ISO 14020:2022. Available at https://www.iso.org/standard/79479.html

[9] ISO (2010) ISO 14025:2006 Environmental labels and declarations — Type III environmental declarations — Principles and procedures. Available at https://www.iso.org/standard/38131.html

[10] International Organization for Standardization (2017) ISO 14027:2017. Available at https://www.iso.org/standard/66123.html

[11] International Organization for Standardization (2022) ISO 14029:2022. Available at https://www.iso.org/standard/78513.html

[12] Bhat CG, Adhikari T, Mellentine J, Feraldi R, Lasso A, Swack T, Mukherjee A, Dylla H, Rangelov M (2022) 2022 ACLCA PCR Guidance – Process and Methods Toolkit Creating standardized, consistent, and reliable PCRs & EPDs for transparency, procurement, and supply chain data. Available at https://cdn.prod.website-files.com/6706f3c8022e66be4da4c0be/6717f4d22f1b08589aac43bd_2022-ACLCA-PCR-Guidance_v1_Introduction_05252022.pdfhttps://cdn.prod.website-files.com/6706f3c8022e66be4da4c0be/6717f4d22f1b08589aac43bd_2022-ACLCA-PCR-Guidance_v1_Introduction_05252022.pdf

[13] ISO (2017) ISO 21930:2017 Sustainability in buildings and civil engineering works — Core rules for environmental product declarations of construction products and services. Available at https://www.iso.org/standard/61694.html

[14] Ingwersen WW, Stevenson MJ (2012) Can we compare the environmental performance of this product to that one? An update on the development of product category rules and future challenges toward alignment. *Journal of Cleaner Production* 24:102–108. https://doi.org/10.1016/j.jclepro.2011.10.040

[15] Ranganath S, McCord S, Sick V (2024) Assessing the maturity of alternative construction materials and their potential impact on embodied carbon for single-family homes in the American Midwest. *Frontiers in Built Environment* 10:1384191. https://doi.org/10.3389/fbuil.2024.1384191

[16] National Center for Sustainability Standards (2024) PCR for Concrete v2.3 – 2024 Extension.

[17] Federal LCA Commons (2024) Memorandum of Understanding among the U.S. Department of Energy (DOE), U.S. Environmental Protection Agency (EPA), U.S. Department of Agriculture (USDA), U.S. Department of Transportation (DOT), and the National Institute of Standards and Technology (NIST) on the Federal Life Cycle Assessment (LCA) Commons. Available at https://www.lcacommons.gov/memorandum-understanding-0

[18] EPA (2022) Interim Determination on "Substantially Lower Embodied Carbon" Construction Materials. Available at https://www.epa.gov/system/files/documents/2023-01/2022.12.22%20Interim%20Determination%20on%20Low%20Carbon%20Materials%20under%20IRA%2060503%20and%2060506_508.pdf

[19] EPA (2024) Life Cycle Inventory Data Gap Assessment. (Environmental Protection Agency, Washington D.C.), p 40. Available at https://www.epa.gov/greenerproducts/label-program-low-embodied-carbon-construction-materials

[20] EPA (2024) A Vision and Plan to Improve Secondary Life Cycle Assessment Data Used in Environmental Product Declarations. (Environmental Protection Agency, Washington D.C.), p 35. Available at https://www.epa.gov/greenerproducts/label-program-low-embodied-carbon-construction-materials

[21] EPA (2024) U.S. EPA Criteria for Product Category Rules (PCRs) to Support the Label Program for Low Embodied Carbon Construction Materials (EPA's PCR Criteria) (Version 1—2024). (Environmental Protection Agency, Washington D.C.), EPA-740-R-24-009, p 65. Available at https://www.epa.gov/greenerproducts/label-program-low-embodied-carbon-construction-materials

[22] Bacas P, Carpenter A, Cashman S, Chen J, Hottle T, Ingwersen W, Kneifel J, Oiwa L, Ranganath S, Rodriguez-Garcia G, Young B (2024) Federal Life Cycle Inventory Data Gap and Data Quality Assessments. Available at https://virtual.oxfordabstracts.com/event/public/24956/session/121528

[23] Ingwersen WW, Li M, Young B, Vendries J, Birney C (2022) USEEIO v2.0, The US Environmentally-Extended Input-Output Model v2.0. *Scientific Data* 9(1):194. https://doi.org/10.1038/s41597-022-01293-7

[24] Rodriguez Garcia G, Ranganath S, Kneifel J (2025) Public Life Cycle Inventory Data Gap Analysis through Process Modeling. (National Institute of Standards and Technology, Gaithersburg, MD), NIST TN 2338, p NIST TN 2338. https://doi.org/10.6028/NIST.TN.2338

[25] Broyles JM, Gevaudan JP, Brown NC (2024) A compiled dataset of ready-mix concrete Environmental Product Declarations for life cycle assessment. *Data in Brief* 52:109852. https://doi.org/10.1016/j.dib.2023.109852

[26] AsphaltEPD - Find an EPD Available at https://asphaltepd.org/published/

[27] DownThemAll! Available at https://www.downthemall.net/

[28] EC3 - Find & Compare Materials Available at https://buildingtransparency.org/ec3/material-search

[29] openEPD: Home - An Open Format for the world's Digital EPDs Available at https://www.open-epd-forum.org/

[30] Home | ecoQuery Available at https://ecoquery.ecoinvent.org/3.10/cutoff

[31] National Renewable Energy Laboratory/USLCI | LCA Collaboration Server Available at https://www.lcacommons.gov/lca-collaboration/National_Renewable_Energy_Laboratory/USLCI_Database_Public/datasets

[32] LCA Collaboration Server Available at https://www.lcacommons.gov/lca-collaboration/

[33] Long Trail Sustainability (2023) Datasmart LCI Package Manual. Available at https://longtrailsustainability.com/wp-content/uploads/2024/05/LTS-DATASMART-LCI-Package-Manual_2023.pdf

[34] LCA Database *Sphera*. Available at https://lcadatabase.sphera.com/

[35] GREET *Energy.gov*. Available at https://www.energy.gov/eere/greet

[36] National Center for Sustainability Standards (2020) PCR for Portland, Blended, Masonry, Mortar, and Plastic (Stucco) Cements. Available at https://www.cement.org/docs/default-source/pcr-2020/pcr-portland-cement-2020.pdf

[37] Foundry Industry Recycling Starts Today (FIRST) (2004) Foundry Sand Facts for Civil Engineers. (US DOT - FHWA, Springfield, VA), FHWA-IF-04-004, p 69. Available at https://www.fhwa.dot.gov/pavement/pubs/013791.pdf

[38] Foundry Sand - Material Description - User Guidelines for Waste and Byproduct Materials in Pavement Construction - FHWA-RD-97-148 Available at https://www.fhwa.dot.gov/publications/research/infrastructure/structures/97148/fs1.cfm

[39] National Asphalt Pavement Association (NAPA) (2022) Product Category Rules (PCR) For Asphalt Mixtures v2.0. Available at https://www.asphaltpavement.org/uploads/documents/EPD_Program/NAPA_PCR_AsphaltMixtures_v2.pdf

[40] National Asphalt Pavement Association (NAPA) (2022) Product Category Rules (PCR) For Asphalt Mixtures v2.0 - Annex 1 Prescribed Upstream (Secondary) Data Sources. Available at https://www.asphaltpavement.org/uploads/documents/EPD_Program/NAPA_PCR_AsphaltMixtures_v2.pdf

[41] Wildnauer M, Mulholland E, Liddie J (2019) Life Cycle Assessment of Asphalt Binder. (Asphalt Institute), 978-1-934154-76–2, p 81. Available at https://www.asphaltinstitute.org/wp-content/uploads/Asphalt-Institute-LCA-Report_2019-03-14_final.pdf

[42] Mukherjee A (2021) Update to the Life Cycle Assessment for Asphalt Mixtures in Support of the Emerald Eco Label Environmental Product Declaration Program. (National Asphalt Pavement Association, Greenbelt, MD), p 68.

[43] Marceau ML, Nisbet MA, VanGeem MG (2006) Life Cycle Inventory of Portland Cement. PCA R&D. (Portland Cement Association, Skokie, IL), 2095, p 69. Available at http://large.stanford.edu/courses/2016/ph240/pourshafeie2/docs/marceau-2007.pdf

[44] Portland Cement Association (2023) Environmental Product Declaration: Portland Cement. (Portland Cement Association, Skokie, IL), p 13. Available at https://www.cement.org/a-sustainable-future/reaching-our-goal/quantifying-environmental-impact/#chapter=chapter4

[45] Slag Cement Association (2021) An Industry Average Environmental Product Declaration for Slag Cement. (Slag Cement Association, United States and Canada), p 13. Available at https://pcr-epd.s3.us-east-2.amazonaws.com/707.EPD_for_Slag_Cement_Industry_Wide_EPD.pdf

[46] Miller CA (2008) Characterizing Emissions from the Combustion of Biofuels. (U.S. Environmental Protection Agency, Washington, DC), EPA/600/R-08/069. Available at https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NRMRL&dirEntryId=191572

[47] Hottle T, Hawkins TR, Chiquelin C, Lange B, Young B, Sun P, Elgowainy A, Wang M (2022) Environmental life-cycle assessment of concrete produced in the United States. *Journal of Cleaner Production* 363:131834. https://doi.org/10.1016/j.jclepro.2022.131834

[48] Argonne National Laboratory/Cement and Concrete | LCA Collaboration Server Available at https://www.lcacommons.gov/lca-collaboration/Argonne_National_Lab/Concrete/dataset/PROCESS/dea78ba6-49e9-061c-527a-d7b5891ba61a

[49] Baghouse Fines - Material Description - User Guidelines for Waste and Byproduct Materials in Pavement Construction - FHWA-RD-97-148 Available at https://www.fhwa.dot.gov/publications/research/infrastructure/structures/97148/bd.cfm

[50] Kopman J (2020) Brown Grease or Yellow Grease? Knowing the Difference Can Help Keep Your Grease Trap Clean and Your Restaurant in Top Working Condition. *Environmental Remedies*. Available at https://envremedies.com/envremblog/brown-grease-or-yellow-grease

[51] Sustainability | Ingevity - Ingevity Available at https://www.ingevity.com/sustainability/sustainability-resource-library/

[52] Ieke Bak from Ecochain Technologies B.V (2023) Environmental Product Declaration (EPD) Anova® 1501 Warm Mix Additive by Cargill Inc. Available at https://asphaltepd.org/static/core/upstream_epds/cargill/20230125_Verified_EPD_Anova_1501_Warm_Mix_Additive.pdf

[53] Ieke Bak from Ecochain Technologies B.V (2023) Environmental Product Declaration (EPD) Anova® 1815 Rejuvenator by Cargill Inc. Available at https://asphaltepd.org/static/core/upstream_epds/cargill/20230125_Verified_EPD_Anova_1815_Rejuvenator.pdf

[54] Ieke Bak from Ecochain Technologies B.V (2022) Environmental Product Declaration (EPD) Ace XP by SurfaceTech. Available at https://surface-tech.com/wp-content/uploads/2022/06/SCS-EPD-07560_SurfaceTech_AceXP_082622.pdf

[55] Federal Highway Administration/MTU Asphalt Pavement Framework | LCA Collaboration Server Available at https://www.lcacommons.gov/lca-collaboration/Federal_Highway_Administration/mtu_pavement/dataset/PROCESS/78f5734a-b99e-4dd7-a05d-715ad4c55a7c

[56] Kalampokis S, Kalama D, Kesikidou F, Stefanidou M, Manthos E (2023) Assessment of Waste Glass Incorporation in Asphalt Concrete for Surface Layer Construction. *Materials* 16(14):4938. https://doi.org/10.3390/ma16144938

[57] National Center for Sustainability Standards (2020) PCR for Gypsum Panel Products. Available at https://d2evkimvhatqav.cloudfront.net/documents/pcr_gypsum_panel_products_2019_1.1.pdf?v=1594929936

List of Symbols, Abbreviations, and Acronyms

AA
Antistrip Agents

ACLCA
American Center for Life Cycle Assessment

Admx
Concrete Admixtures

Aggr.
Construction aggregates

ANL
Argonne National Laboratory

ANL C&C
Argonne National Laboratory Concrete and Cement Database

API
Application Programming Interface

BC
Boiler Combustion (Fuel Combusted in Industrial Boiler)

BHC
Blended Hydraulic Cement

BMEx
Binder Modifier, Extender

BMNA
Binder Modifiers, Natural Asphalt

BMO
Binder Modifiers, Other

BMP
Binder Modifier, Plastics

BMRA
Binder Modifier, Recycling Agent

C2Gate
Cradle-to-Gate

C2Grave
Cradle-to-Grave

CDD Management
Construction and Demolition Debris Management

CNG
Compressed Natural Gas

CO_2
Carbon Dioxide

CSV
Comma-Separated Values

DGA
Data Gap Assessment

DOE
Department of Energy

EC3
Embodied Construction Carbon Calculator

EI
ecoinvent

eLCI
Electricity Baseline LCI Data

EOL
End of Life

EPA
Environmental Protection Agency

EPD
Environmental Product Declaration

EPM
Ethylene Propylene Copolymers

EPDM
Ethylene Propylene Diene

EVA
Ethylene-Vinyl Acetate

FLCAC
Federal LCA Commons

GA
Grinding Aids

GGBFS
Ground Granulated Blast Furnace Slag

GREET
Greenhouse Gases, Regulated Emissions, and Energy use in Technologies

GRT
Ground Rubber Tire

HD
Heavy Duty (Truck)

HEO
Heavy Equipment Operations

HTML
Hypertext Markup Language

IA EPD
Industry Average Environmental Product Declaration

IC
Internal Combustion

ICE
Internal Combustion Engine

IE
Industrial Equipment

ISO
International Organization for Standardization

LA
Liquid Antistrip

LCA
Life Cycle Assessment

LCI
Life Cycle Inventory

LLM
Large Language Model

LNG
Liquefied Natural Gas

LPG
Liquefied Petroleum Gas (Propane)

MF
Mineral Fillers

ML
Machine Learning

MS
Manufacturer Specific

MSP EPD
Manufacturer-Specific Product Environmental Product Declaration

MSW
Municipal Solid Waste

NAPA
National Asphalt and Pavement Association

NERC
North American Electric Reliability Corporation

NIST
National Institute of Standards and Technology

NG
Natural Gas

NREL
National Renewable Energy Laboratory

NRMCA
National Ready Mix Concrete Association

PBHC
Performance-Based Hydraulic Cement

PCA
Portland Cement Association

PC
Portland Cement

PPA
Polyphosphoric Acid

PO
Program Operator

Prod.
Production

RA
Recycling Agents

RAP
Recycled Asphalt Pavement

RAS
Recycled Asphalt Shingles

RD
Renewable Diesel

REOB
Re-Refined Engine Oil Bottom

RET
Reactive Ethylene Terpolymers

RNA
Rest of North America

RoW
Rest of the World

RCA
Recycled Concrete Aggregates

RFO
Residual Fuel Oil

SBR
Styrene-Butadiene-Rubber

SBS
Styrene-Butadiene-Styrene

SCA
Slag Cement Association

SF
Synthetic Fibers

SPA
Structural Path Analysis

SW
Solid Waste

TA
Terminal Additives

TiO_2
Titanium Dioxide

US-EI
Datasmart LCI Database

USLCI
United States Life Cycle Inventory

USEEIO
US Environmentally Extended Input-Output

UUID
Unique Identifiers

VTAE
Vacuum Tower Asphalt Extender

WARM
Waste Reduction Model

WMA
Warm Mix Additives

Appendix A. Supplemental Material

This appendix provides brief summaries for each of the supplemental materials associated with this study. The "Supplemental Material – Data Gap List, EPD Databases, and Code" ZIP file is available at https://doi.org/10.6028/NIST.TN.2353sup1, and includes the lists of data gaps, each of the EPD databases, and the programming code scripts used for extracting data from EPDs. Additionally, there is a ReadMe file and Definition file to assist in using the EPD databases. Finally, there is a summary of the parsing pipeline for extract the data that provides software developer-relevant details about the approach.

A.1. Data Gap Lists and EPD Databases

The data gap lists are available in 'datagap assessment.xlsx," with a tab for each product category separated by PCR and or EPD level assessments whenever applicable.

The EPD database files are as follows:

- Concrete – ConcreteDB_adoptedFromBroylesEtAl.csv
- Cement – CementDB_manuallyExtracted.csv
- Asphalt – AsphaltDB_hardCoded.csv
- Gypsum – GypsumDB_hybrid.csv

The "readMe.txt" file summarizes the suite of files provided as supplemental material for this study. The "definition.xlsx" file includes a tab for each EPD database with definitions for each column variable.

A.2. Scripts and LLM Prompting

The supplement material includes programming code script files used to complete the following tasks in this study. First, the "asphalt_db.py" file provides the python scripts to extract data from asphalt EPDs and develop the asphalt database. Second, the "openEPD_api_call.py" and "gypsum_uuids_script_db.py" files are the scripts used to complete the EC3 API call to compile all unique IDs for the U.S. valid gypsum EPDs available in EC3 and an EC3 API call loop to download the openEPD file for each unique ID and convert and append the openEPD file to a CSV file, respectively. Third, the "azure_gypsum_script.js" file is the code used to extract information on material content from the EPDs with single products using Azure Document Intelligence.

Additionally, as discussed in Section 2.2, an example set of LLM prompts were developed for extracting information on material content from the gypsum EPDs with multiple products. These prompts were:

- Prompt #1: What is the declaration number in this EPD?
- Prompt #2: Can you give me the product specifications that you find in the pdf?

- Prompt #3: Can you give me the product specifications and declaration number from the PDF in a CSV file?
- Prompt #4: Can you give me the material composition and declaration number in the PDF?
- Prompt #5: Can you give me the declaration number and the material composition or product composition in the PDF?
- Prompt #6: Can you give me the declaration number and either material composition or product composition in percentages that you find in the PDF?
- Prompt #6 returned the material content information, and the same prompt was used with the rest of the EPDs.

The prompts were generally effective, but the approach requires more analysis to refine and improve its efficiency and robustness to varying data reporting and formats.

The development of the parsing pipeline to extract relevant data from the semi-structured EPDs is summarized in the supplement document titled “Parsing Pipeline.docx,” which is structured for future submission to a topic-relevant blog for software develops and computer programmers.