



AALBORG UNIVERSITY
DENMARK

Aalborg Universitet

Can LCA be FAIR? Assessing the status quo and opportunities for FAIR data sharing

Ghose, Agneta

Published in:
International Journal of Life Cycle Assessment

DOI (link to publication from Publisher):
[10.1007/s11367-024-02280-3](https://doi.org/10.1007/s11367-024-02280-3)

Creative Commons License
CC BY 4.0

Publication date:
2024

Document Version
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

Citation for published version (APA):
Ghose, A. (2024). Can LCA be FAIR? Assessing the status quo and opportunities for FAIR data sharing. *International Journal of Life Cycle Assessment*, 29(4), 733-744. <https://doi.org/10.1007/s11367-024-02280-3>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.



Can LCA be FAIR? Assessing the status quo and opportunities for FAIR data sharing

Agneta Ghose¹

Received: 28 July 2023 / Accepted: 8 January 2024 / Published online: 22 January 2024
© The Author(s) 2024

Abstract

Aim The purpose of this study is to assess the status quo of data sharing in LCA in relation to the FAIR (findability, accessibility, interoperability, and reuse) data principles.

Methods This study investigates how is LCA data from publicly funded research currently shared. Firstly, the focus is on life cycle inventory data shared in journal articles. Given that FAIR data sharing is not only the responsibility of the LCA practitioner, this study further investigates guidelines (e.g., data sharing standards and data management plans) and infrastructure (repositories, data formats, and nomenclature) to identify the tools and services available to LCA community which are essential to enable FAIR data sharing.

Results The study identifies that although there is growing awareness to improve data sharing practices, implementation of FAIR guidelines for data sharing is seldom seen in practice. LCA studies that adhere to FAIR principles are primarily due to use of generic data repositories which provide tools to support data sharing. However, there is no guidance on how LCA specific data should be shared to ensure its findability, accessibility, interoperability, and reusability. This study suggests a workflow to enable FAIRification of LCA data. In addition, the study recommends further efforts within the LCA community on skill and technology development, strategic funding, and recognition of the best practices in relation to data sharing.

Conclusion In conclusion, this study highlights the necessity of data sharing incentives, guidelines, and platforms/repositories specific for the LCA community.

Keywords LCA · Data sharing · Data management plan · DMP · FAIR data · Data repositories · LCA data curation

1 Introduction

FAIR, an acronym for findable, accessible, interoperable, and reusable, is a core set of principles introduced in 2016 (Wilkinson et al. 2016). The adoption of these principles can improve data sharing and reuse in all disciplines, hence rapidly achieving widespread agreement in all scientific communities. The implementation of FAIR in the European Commission and its member states is supported by the European Open Science Cloud (EOSC 2020). Parallel initiatives such as NIH Data Commons (US), the Australian Research Data Commons, and the African Open Science Platform are

developed to ensure that data are FAIR across disciplines and geographic boundaries (Collins et al. 2018).

The use of Life Cycle Assessment (LCA) as an environmental assessment tool has grown extensively in all industrial sectors. LCA is a data-driven tool (Ghose et al. 2022). Creating a robust LCA requires data on raw material, energy, and emission flows for all or several processes in a product life cycle. The quality of inventory data determines the credibility of the LCA results. Cost and time required to develop such robust models are general issues in the development of LCA (Brose 2011). Practitioners need to extract data from several sources. The extracted data must often be transformed to different formats to make LCA calculations. Datasets must be reviewed for quality control before a formal analysis is conducted. A key obstacle in transparency for LCA data is related to the need for confidentiality from industries providing primary data (Stenzel and Waichman 2023). Primary data-related production processes could be highly aggregated or even incomplete. Sometimes complex

Communicated by Martin Baitz.

✉ Agneta Ghose
agneta@plan.aau.dk

¹ Department of Sustainability and Planning, Aalborg University, Rendsburggade 14, 1-432, 9000 Aalborg, Denmark

simulations are developed to fill large data gaps (Parvatker and Eckelman 2019). Besides the financial and time constraints, access to relevant data is hindered by poor data management. One of the primary challenges of data sharing is that the datasets generated by each research group are independent and siloed from one another, leading to limited transparency, interoperability, and the lack of data sharing protocols. As the demand for credible and transparent sustainability assessment increases, it is pertinent that data used for LCA has high integrity and quality and is well organized for accessibility and interoperability (Ghose et al. 2022).

Shared data can be of several types (Michener 2015; Venkataraman 2022). In the LCA domain, there is no consensus on the definition of the term “LCA data.” LCA data typically refers to complex information including quantitative data and a wide range of metadata including LCA-specific modeling choices, assumptions, and inputs from subject matter experts (Kahn et al. 2022). Generally, research data can refer to several data types. Data can be primary data also referred to as raw data. This could be referred to as process or site-specific data, bill of materials, etc. Primary data is valuable for LCA. It makes the model credible through accurate raw data that is specific to a study—instead of relying on averages. It is pertinent to understand that often data suppliers from industries are cautious or hesitant about disclosing their primary data. Secondary data refers to data that has been collated, summarized, and undergoes a formatting treatment. For example, when primary data from different suppliers is linked together in the form of a foreground matrix in LCA, it can be referred to as secondary data. Other data types include data products and non-standard data outputs. Data products are data that have undergone significant processing and calculations. Often an LCA data resulting from a study is a highly engineered information that functions more as a model, which includes the life cycle inventory, impact assessment results, results from sensitivity or scenario analysis. Non-standard data outputs refer to protocols, tools, or codes developed for analysis.

The emergence of FAIR and open data as the best practice for scientific communities highlights the challenges in the LCA domain related to data sharing to ensure transparency and reproducibility of LCA data (Kahn et al. 2022). Currently, there are no clear specifications on LCA data sharing to ensure findability, accessibility, interoperability, and reuse (FAIR).

1.1 Implementing FAIR in research

Research articles and data products are key outcomes of scientific research (Michener 2015; Shanahan and Bezuidenhout 2022). Data generated in academic work should be discoverable and usable by a variety of potential reusers for effective and efficient development of research and

innovation. The FAIR guiding principles were developed such that they could be applied broadly to all categories of research outputs; this includes data, code, and software (Wilkinson et al. 2016; Bast 2019). The broad category of research outputs is also referred to as **digital objects**. The guidelines to implement FAIR principles emphasize that it applies both to the data and the metadata of the digital object. In addition, the principles also emphasize on machine actionability of the digital objects such that computational system can find, access, interoperate, and reuse data (Go FAIR 2022).

To ensure findability, each digital object must be accompanied by a persistent and globally unique identifier (PIDs) with essential metadata. A PID is a unique identifier that enables stable links to the digital objects. Examples of PIDs are digital object identifiers (DOI) used for published articles and data, ORCIDs for researchers, Cordis ID or RAIDs for research projects (Australian Research Data Commons 2021), and RRIDs for associated research resources (Bandrowski and Martone 2016). To ensure that digital objects are accessible, they should be provided with rich metadata (Shanahan and Bezuidenhout 2022). Rich metadata refers to how, why, when, and by whom the objects were created, in addition to all relevant attributes that can support reproducibility, replicability, and an accessible license for data reusers. Digital objects must be stored in trusted repositories or cloud services that have widespread use. To ensure interoperability, digital objects should be represented in formats that are widespread and easy to use. For example, data stored in.csv format are easier to reuse compared to data stored in a table in a.pdf or.docx format (Collins et al. 2018; Venkataraman 2022). Research communities need to support common interoperability frameworks, such as data sharing, data formats, metadata standard, tools, and infrastructure. The ultimate objective of FAIR is to optimize the reuse of data. To ensure data is reusable, data and metadata should be well described using domain-relevant community standard, clear usage license, and detailed provenance (Collins et al. 2018; Venkataraman 2022).

1.2 A FAIR ecosystem

The implementation of common research data management can be complex for researchers. Funding institutions have recognized that research data sharing is beyond the sole responsibility of individual researchers (Collins et al. 2018). The transition to FAIR data requires the integration of FAIR principles in the entire research workflow. The EU commission report on the integration of FAIR guidelines suggests that the research workflow can be assisted by creating a FAIR ecosystem that includes several elements each of which are managed using the FAIR principles (Collins et al. 2018). Key components of a FAIR ecosystem are data

management plans, data repositories, technological support, data policies or standards, and data producers/users.

For example, the function of a **data management plan (DMP)** goes beyond a record of information on basic data storage and backups. DMP is a valuable resource as it holds information on all data and related outputs, hence broadly applicable to all research data outputs such as software, workflows, and protocols. Thus, the DMP is a living document and must be regularly updated to provide the hub of information on digital objects related to a research project (Collins et al. 2018). Therefore, systems containing information about research projects should explicitly refer the project's DMP. DMP should support the implementation and themselves confirm to FAIR principles and be open where possible, i.e., every DMP should have a PID with essential metadata, including an accessible license, stored in interoperable format in a trusted repository (Whyte 2019). There are several DMP tools¹ available that support the development of comprehensive DMPs and align them with the FAIR principles.

Increasingly, funding authorities demand a strategy for continued access to data produced by the projects they fund. Storing all digital outputs in trusted **digital repositories** provides reliable, long-term access to manage digital resources. Digital repositories may be generic, such as Zenodo and Figshare (Nielsen and Smith 2014; Figshare Team 2022), or subject specific, such as PANGAEA (a data repository for earth and environmental science) (PANGAEA 2023). Such digital repositories play an important role to enable data sharing based on FAIR principles. Longevity and security of data are dependent on various issues such as technical, financial, legal, and organizational support. It is often challenging for a single institution or research body to provide all the necessary support for domain-specific data over long term. Thus, trusted and certified repositories can fulfill the needs of data producers and users. It is therefore vital to ensure the development of domain-relevant repositories.

Technological tools to facilitate data management workflows (ARGOS OpenAIRE 2022) and automated processing are essential to support data providers. These can enable data upload and improve data discoverability. **Data policies** define and regulate data management and data interoperability, standards, data quality, data protection, and information security (European Commission 2020). Hence, data policies play a key role in the FAIR ecosystem to incentivize the adoption of FAIR principles in data management (Musen et al. 2022). For example, data policies can direct the use of core metadata standards such as the Dublin Core metadata element (DublinCore 2022). The Dublin Core is a standard (ISO 15836) that defines semantically linked key metadata

elements used for a broad number of resources including digital and physical. Such a metadata standard when adopted and consistently can improve data discovery and the possibility to assess and utilize data at scale. Finally, the key components of the FAIR ecosystem are the **data producers/users**. Investing in skill development, acknowledgement by merit to good data sharing in practice and investment in creation and maintenance of data-sharing platforms within community can enhance collaboration and the research ecosystem (Collins et al. 2018).

In the light of rising awareness and growing need of better data management across disciplines, it is pertinent to investigate how the FAIR principles are considered in the LCA or the sustainability assessment domain. The purpose of this research is to investigate tools and infrastructure currently available to support a FAIR data sharing in LCA domain.

2 Status quo of data sharing in LCA

This study attempts to present the status quo of data curation methods adopted by LCA practitioners to share the life cycle inventory (LCI) data developed for academic research. LCI data is a key resource that can be used to assess the replicability of an LCA model or reused to develop other models. First, the study assesses the status quo of LCI data shared in peer-reviewed articles which are key research outputs. Furthermore, the study aims to assess key components of the FAIR ecosystem (see Fig. 1) available to support data sharing in the LCA domain to identify and discuss the barriers and opportunities related to the current data infrastructure. Data management plans of recent European projects that use LCA were examined to gain insights on data management protocols and stewardship. The study further assesses stand digital infrastructure such as data repositories, data format, vocabularies, and LCA software.

3 Research articles

To understand the status quo of data sharing in scientific articles, 25 peer-reviewed LCA articles were selected. These articles were published between 2018 and 2022. These studies covered a broad range of topics such as construction, food production, waste treatment, novel technologies, bio-based material production, vehicles, and batteries. No review papers were selected. All articles investigated for this study can be found in the Zenodo repository (see SI 1). There are no common guidelines on the implementation of FAIR with respect to LCA. Hence, to assess the adherence to FAIR principles, five aspects were investigated: (1) Where was the inventory shared? (2) Does the shared LCI data have a persistent identifier? (3) How was the data shared (data

¹ ARGOS, EUDAT, DMP tool, DSW.

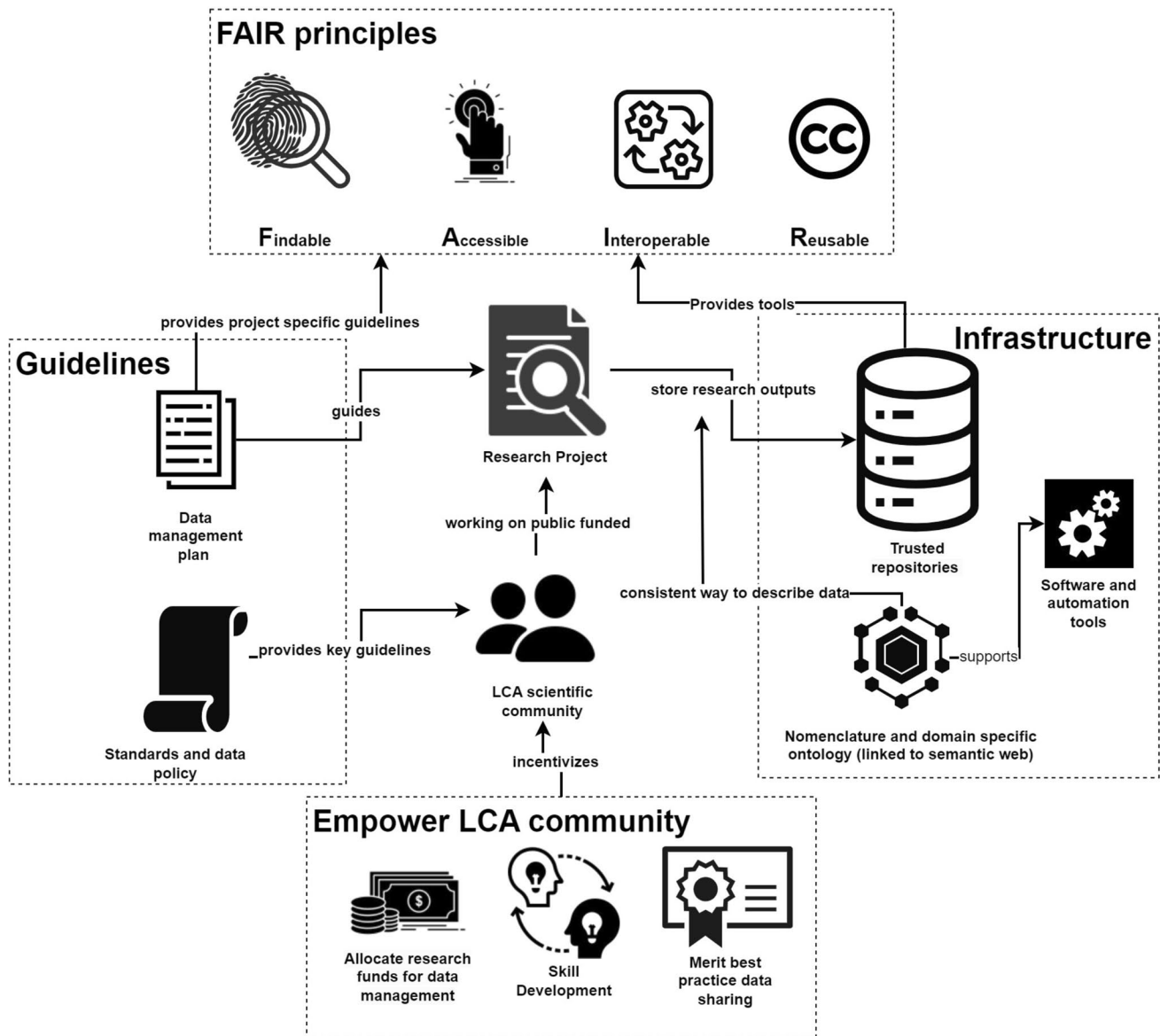


Fig. 1 An overview of the FAIR data ecosystem. Consistent implementation of the FAIR principles while sharing data by the LCA community depends on multiple components such as guidelines from data management plans and data policies; technological infrastruc-

ture such as trusted digital repositories to store data, domain-specific ontologies that support interoperability and automation in software; and incentivizing data sharing through merit, skill development, and allocating funds for data management

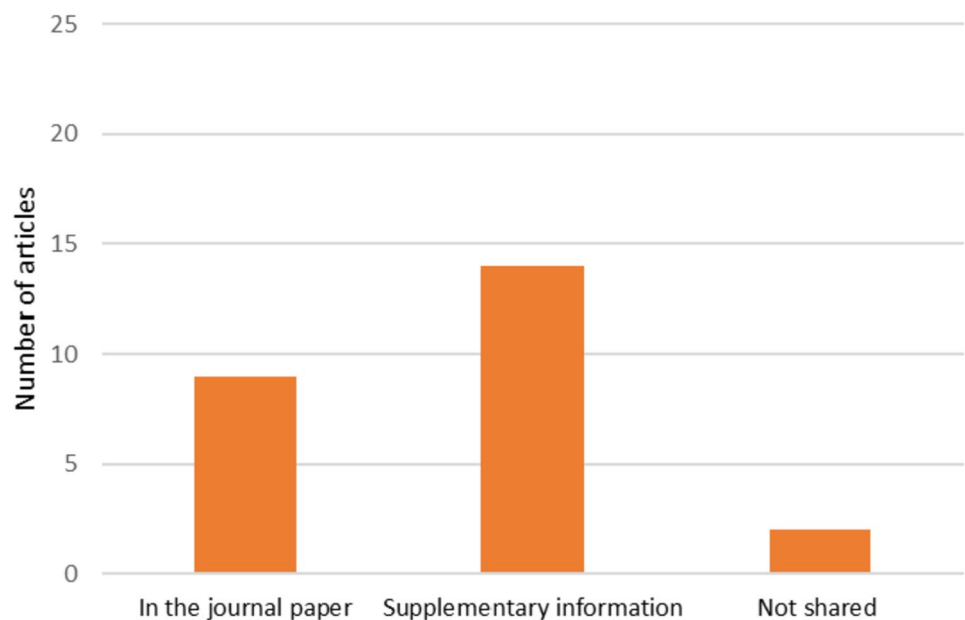
format)? (4) Is the shared format interoperable? (5) Which license was attributed to the shared data?

The status quo of LCI data shared in the LCA articles is shown in Fig. 2. While there were some studies that included the inventory in the journal paper, LCI data was commonly shared in the supplementary information. It was also interesting to note that two studies did not include the LCI or any data used to model the inventory. These studies only referred to data sources used for the LCI which were inaccessible.

Of the 23 studies that shared the LCI, only one study (Köhler and Pizzol 2019) included a separate persistent identifier attached to the dataset. A separate identifier for the

dataset ensures its findability even if the associated article is inaccessible (e.g., behind a paywall). Only seven of the 25 studies were published with an open access license (Weber et al. 2018; Muñoz et al. 2018; Asem-Hiablie et al. 2018; Thonemann and Pizzol 2019; Köhler and Pizzol 2019; Rotz et al. 2021; Keller et al. 2022). Interoperability with respect to the data format was limited. Only three studies (Thonemann and Pizzol 2019; Köhler and Pizzol 2019; Keller et al. 2022) shared the LCI in csv files, formatted such that the inventory could easily be used in common LCA software (e.g., Brightway/Simapro). Other studies shared the LCI

Fig. 2 Where the LCI data is shared in selected peer-reviewed articles reporting an LCA between 2018 and 2022



data either in a pdf or a word document. It was interesting to note that the study which shared the most reproducible and replicable inventory, model, and code was shared with the most restrictive open access license (i.e., [CC-BY-NC-ND](#)). The authors of this paper (Köhler and Pizzol 2019) mentioned their limited knowledge of open access licensing (S. Köhler and M. Pizzol, personal communication, 2022). The choice of the license was made primarily due to economic reasons and a restrictive open access license was significantly cheaper than a [CC-BY 4.0](#) international license. Increasingly, journal publishers are directing authors to start storing all supplementary information (e.g., data and methods) in established generic data repositories (Pop and Salzberg 2015; Kwon 2020; Springer 2023; ScienceDirect 2023; Frontiers 2023). These repositories provide multiple tools to ensure datasets can be shared following the FAIR guidelines. However, these guidelines are generic. The authors lack clear information on how the LCI data should be structured and the relevant metadata and file format to be provided (M. Pizzol, personal communication, 2022).

3.1 Domain standard for data sharing

Currently, there are no overarching standards that provide a procedure for LCA data sharing. The ISO standard (ISO 14048 2002) provides the technical specification to facilitate reporting of LCI data and its compliance from ISO 14040 and ISO 14044. The standard provides an elaborate data documentation format, which consists of three parts including process description (including inputs and outputs), modeling and validation (including modeling choices, quality of the data), and administrative

information. This elaborate documentation was developed to avoid any ambiguity in the data.

The specifications have been used to develop ISO-compliant data formats, most notably ILCD developed by the European Commission's Joint Research Committee, EcoSpold2 developed by Ecoinvent, and OpenLCA JSON-LD data model developed by GreenDelta (Kahn et al. 2022). The data specifications while elaborate do not necessarily support findability, accessibility, interoperability, and reusability.

For example, the FAIR principles recommend the use of globally unique and persistent identifiers for the (meta) data such as a Unique Resource Identifier (URI) that gives a namespace for the web location of a resource which is missing in the data documentation. Similarly, while the (meta)data documentation is elaborate it remains insufficient with respect to accessibility related to licensing the data for reuse. The data formats developed using the standard are key resource for data exchange and for interoperability. However, data vendors take different approaches to including ISO specifications which creates inconsistencies and challenges in interoperability during data exchange (Kahn et al. 2022). Kahn et al. (2022) also critiqued the insufficient information on intended applications of a dataset, upstream input providers, etc. that affects the reusability of the data.

The ISO standard for data documentation was developed in 2007, many years before the FAIR guidelines were introduced. Facilitating data exchange is transforming at a rapid pace. In the future, emendations must be considered in this standard to include the principles of FAIR data sharing.

3.2 Data management plans

Publicly funded research provides the advantage of enabling as well as enforcing data sharing which usually does not occur. Increasingly research funders are demanding for a **data management plan** (DMP) during the grant application process. DMP is the core element for any scientific research project and a valuable resource in the FAIR ecosystem. A detailed DMP acts as a road map that guides and explains how data are treated throughout the life of the project and after the project is completed (Michener 2015). Funding bodies provide grant-specific templates for data management plans which include sections that require consideration of the FAIR principles (Collins et al. 2018; Venkataraman 2022).

Ten publicly shared DMPs of Horizon 2020 projects published between 2019 and 2022 that have collected LCA data were examined in this study (see SI 2). With respect to the objective of the study, we will share the reflections on how these projects consider the FAIR principles in data management.

With respect to **findability**, the DMPs make suggestions to use specific folder and file naming styles and keywords to be used in the project. In addition, these projects also suggest that a data catalogue that provides an overview of all data output is prepared and shared on the final version of the DMP or the project website. Recent DMPs (five of ten) suggest publishing the data collected and generated in the projects in open and free data repositories such as Zenodo. In addition, some projects have considered sharing the data in other repositories hosted by the participating institutions (for long-term repository) or discipline-specific repositories. It was interesting to note that none of the discipline-specific repositories chosen were specific to LCA or industrial ecology.

Accessibility of data is usually determined by first taking into consideration the restrictions to access. It is common that LCA research projects may have collaborations with commercial partners that imply that some of the data remains confidential. Restrictions on data access or impossibility to share them are usually considered in the following cases: (1) collected data belonging to third party which have denied permission for sharing them on account of confidentiality and proprietary issues and (2) protection of personal data of key informants involved in surveys, events, interviews, and case studies. It was interesting to note that most projects consider data collected to perform LCA or LCC as confidential. EU-funded projects usually emphasize that all data generated from the project is available open access. Hence, most projects aim to ensure open access publication for key research results.

To ensure **interoperability**, DMPs recommend using common discipline-specific vocabularies/terminologies to

describe the metadata and data formats that support interoperability between software. Certain DMPs suggest that publication of research outputs in open repositories ensures the adoption of established metadata schemas (e.g., [DataCite](#) or [Dublin Core metadata initiative](#)) which meet the basic standards to adhere to uniform schema to share data. While use of generic schemas is an advantage, with respect to sharing LCA-specific data, there is no description of use of LCA-specific terminologies (e.g., metadata descriptors provided by GLAD) or data format (e.g., EcoSpold or ILCD). Another key feature missing was the recommendation to share data in machine-readable format.

Reusability of datasets is usually linked to clear and correct licensing of the research outputs to avoid legal ambiguities. Projects use standardized licenses for specific outputs such as [Creative Commons](#) for general outputs. While a majority of DMPs did not specify the type of license that will be linked to the research outputs, they reiterate the availability of both confidential and open access data. Two of the DMPs that share type of licensing to be used refer to restrictive open access licenses such as Creative Commons' Attribution-NonCommercial-ShareAlike 4.0 International (CC-BY-NC-SA 4.0) license or Attribution-ShareAlike 4.0 International (CC-BY-SA 4.0).

Overall, most DMPs refer to the FAIR guidelines; however, the implementation of the guidelines is still vague. None of the project DMPs examined provide the link or catalogue to the repositories where the open access datasets are deposited. With respect to LCA datasets, limited to no suggestions are provided to enable FAIR data sharing.

3.3 Data repositories

Data repositories for LCA data can be generic (e.g., Zenodo or Figshare) or industry and country-specific LCI repositories (e.g., Agri-footprint, US LCA Data Commons) or the possibility to deposit data to highly integrated databases (e.g., Ecoinvent), where data is structured in a specific format and linked to the remaining database. With respect to the FAIR guidelines, key features of generic and domain-specific data repositories were identified. This includes, if the repository provides the possibility to have a unique and globally persistent identifier, mandatory metadata requirements or license requirements.

3.3.1 Generic data repositories

Generic data repositories provide several unique features to integrate the FAIR principles (such as Zenodo or Figshare). Key advantages of using these repositories are they are usually free, with user-friendly interface, provision to link a globally unique and persistent (e.g., DOI) and license to every data upload. In addition, these repositories assist in

data management, adding version control, adding embargo for sensitive data or data under review; integration with other platforms, e.g., GitHub to share software and code; and project management by integrating data outputs to related project grants or collaborate on projects that may include working with sensitive or embargoed data. The disadvantage of using these platforms is with respect to interoperability. Given that, these repositories accept data in any format; it is likely that data could be shared in non-interoperable formats (e.g., a pdf document) and hence difficult to integrate with LCA software. This also means that LCA data stored in these repositories do not necessarily include LCA-specific metadata.

3.3.2 LCA-specific data repositories

There are several country and industry-specific LCA repositories (European Commission 2018a). In 2014, the European Commission initiative developed the Life Cycle Data Network (LCDN), a web registry where registered datasets can be searched and then browsed directly from the relative repository (referred to as node) in the network. The European Commission also hosts nodes for data sharing that are free and accessible to data providers (including research institutes, universities, and SMEs) if the data is compliant with ILCD entry level requirements or EU product environmental footprint guidelines (PEF) (European Commission 2018b). For example, the EU-funded research projects node was developed to facilitate disseminating LCA data and results from EU-funded projects. Similarly, the EU small data provider database (SDPDB) node was developed for data providers that want to share less than 10 process datasets. However, use of these repository remains limited with very few datasets (e.g., SDPDB includes only three datasets).

Similar to EU's LCDN, GLAD (Global LCA Data Access network) is an international initiative and UNEP that serves as a directory of LCA databases worldwide (Giacovelli 2020). Datasets can be linked to GLAD provided that they meet the minimum requirements for documentation (i.e., specific metadata description). GLAD allows practitioners to find datasets through their search engine and compare them based on globally agreed metadata descriptors. A specific data search on the GLAD web service leads to the landing page of the LCA data provider. The largest number of free datasets searchable on GLAD is from Agribalyse (French public database of environmental indicators for agricultural and food products based on life cycle analysis), Sphera (commercial database provider), and the US Federal LCA data commons. Unfortunately, online resources change over time and certain URLs (e.g., LCA data provided by World Steel, US Data commons) (particularly those linking to free data sources) currently point to broken links or to data query pages

that are non-functional. This highlights the need to store data in trusted repositories that maintain long-term storage and access and provide a persistent identifier that reliably points to a digital entity. GLAD also aims to support interoperability. Datasets registered on GLAD can be registered in any LCA format (ILCD, EcoSpold,.csv, JSON-LD). GLAD offers a conversion tool that allows users to convert a dataset from its native format to a format that is suitable for a specific LCA software. The GLAD initiative aims to support accessibility and interoperability; it can enhance findability and reusability of the dataset by including mandatory metadata descriptors that include persistent identifiers and license.

3.3.3 LCA database developers

Published LCI data does not stand alone but often combined with background databases developed by other sources. Commercial LCA database developers provide databases designed to be background database.² Widely used databases such as Ecoinvent cover a diverse range of industrial sectors with varying geographic resolution (Miranda et al. 2023). These database developers also provide a platform to LCA practitioners to upload data. Database developers are keen on maintaining data quality and transparency. The advantage of submitting data to database developers is that the data provided is critically reviewed and integrated to the larger database. Moreover, these databases are widely used thus increasing the reach of a dataset and ensuring interoperability with the database in various LCA software. The data provider has the copyright on the data for their own use, but it is also influenced by the restrictive licensing of the commercial database. For example, data deposited to Ecoinvent ensures that Ecoinvent has a non-exclusive use of the uploaded data which cannot be withdrawn by the provider without a compensation. Moreover, access to prominent proprietary databases is linked to a fee and the data reuse in other domains is limited due to restrictive licensing. Pauliuk et al. (2019) emphasize the need for wider data sharing due to the cross-disciplinary nature of industrial ecology tools. While the access to commercial LCA databases might be common among practitioners, it limits the data access across domains.

² Background data is the portion of the LCA study that is not specific to the system being modeled. However, it reflects the industrial economy as a whole and is drawn from reference databases. The background is made of data acquired from secondary sources which are estimated or market averages (Kuczenski et al. 2018).

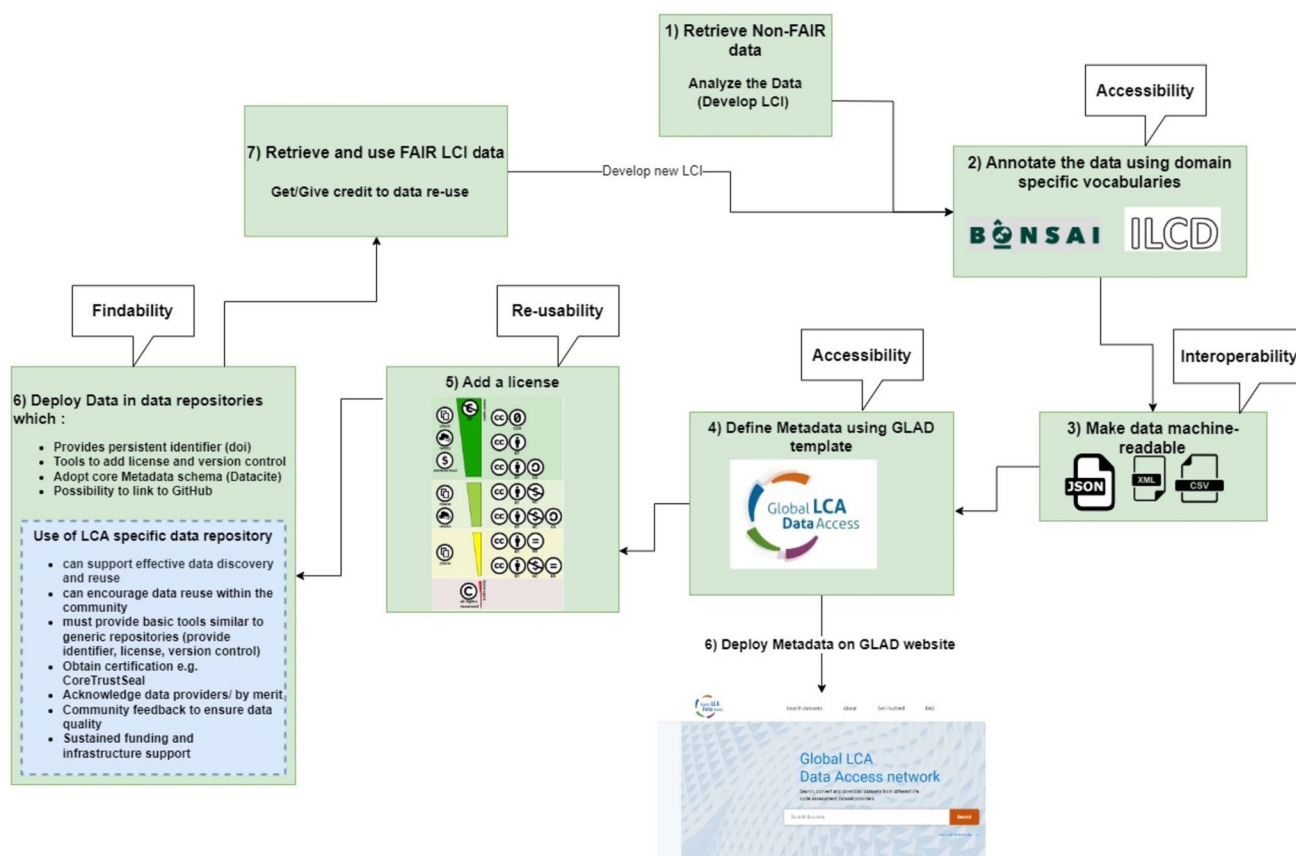


Fig. 3 FAIRification workflow for LCA data

3.4 Nomenclature, Ontologies, and Data Formats

Fritter et al. (2020) identified nomenclature and data exchange formats as core criteria for interoperability. Nomenclature refers to schemas adopted by a domain to classify, categorize, and organize datasets. In LCA, this could refer to how different elements of an LCA dataset is defined (e.g., flows, activities, units). Different nomenclatures are used by different database developers. Loss of data occurs when there are inconsistencies in the nomenclature, for example, if flows and activities are classified differently or there are inconsistencies in metadata provided it leads to a challenge in combining data (Ingwersen 2015). Developing and using semantically linked nomenclature, i.e., ontologies, could reduce the issues of developing mappings of formats between LCI data from different sources (Fritter et al. 2020). Annotating data to semantically linked ontologies is a key to FAIR data sharing (Brewster et al. 2020), as data can only be reused if it is well described and classified, available in both human and machine-readable formats. The BONSAI ontology is a semantic nomenclature meant to be used for LCI data (Ghose et al. 2022). This ontology has relatively minimal

requirements as it mainly describes the core elements of an LCI dataset; however, it gives logical meaning to each semantically linked term (e.g., a flow and its properties) allowing for machine interpretation of data. If semantically linked terms are adopted and used consistently, machines can process, store, manage, and retrieve information based on meaning and logical relationships (Ghose et al. 2022). Giving logical meaning to data using semantically linked ontologies can enable automatization, data interoperability, and efficiency.

EcoSpold2 and ILCD are the two most common data exchange formats used by database developers. Both data formats are ISO compliant. These formats ensure that a dataset within a database meets an acceptable level of documentation, review, and quality (Fritter et al. 2020). Ensuring that datasets are shared in either of these formats supports data interoperability. However, sharing data in these formats burdens the data provider as it requires a resource-intensive LCI data compilation and reporting process. Fritter et al. (2020) recommend the development of data templates that contain essential fields required for both formats. In addition, provide data in JSON-LD as this format allows for easier parsing and linking to semantic data.

4 FAIR LCA data sharing workflow

While there are a plethora of tools, templates, and repositories, the LCA community lacks a clear workflow on their implementation. Figure 3 proposes a workflow inspired by the FAIRification process (Go FAIR 2022) on how the available resources can be implemented to enhance data sharing within the LCA community. The workflow consists of the following steps:

1. **Retrieve data** to develop an LCI. This could refer to collecting data from industrial activities or existing foreground LCI data developed in existing LCA studies.
2. **Annotate the data.** Data annotation refers to labeling the data with relevant tags and the relation between the tags is defined using domain-specific nomenclature. Identify the existing structure of the data to understand the relation between the different elements. For example, data retrieved to develop an LCI may contain information on production activities with multiple input and output flows. Each flow may be reported in different units, or each activity may have a specific location. Understanding these relations might be intuitive for an experienced LCA practitioner but sharing them explicitly can support the accessibility and interoperability of the data. It is recommended that the data is annotated using a semantically linked ontology (Ghose et al. 2022).
3. **Use machine-readable data format.** Sharing data using common LCA formats such as EcoSpold2 and ILCD can alleviate common issues with interoperability (Fritter et al. 2020). However, adopting JSON-LD data format can allow for publishing semantic data. JSON-LD format has lower bandwidth requirements, can be easily parsed, and is ideal for a public API allowing for easier integration into web applications. Data conversion tool is provided by GLAD; however, LCA software providers (e.g., openLCA) can play a key role in providing data format converters, hence supporting both efficiency for LCA practitioners and data interoperability (Kahn et al. 2022). Most importantly, it is important to share data in a format that is easy to retrieve and compute. This also means that it is better to share data in a.csv format rather than a pdf document.
4. **Define metadata.** Defining the dataset using the GLAD core metadata descriptors allows the possibility to deploy the metadata on the GLAD webpage thus enhancing the accessibility and discoverability of the data (Kusche 2020). This is relatively easy to do as GLAD provides a [metadata template](#) on its website. However, the metadata format can itself be enhanced to be semantically linked by building an ontology that is

supported by established metadata schema (e.g., [Data-Cite](#))

5. **Add a license.** Licensing the dataset is vital to ensure correct reuse. The highest potential reuse of data comes when data are both FAIR and open. However, given that commercial partners imply that some of the data remains confidential, data can be licensed using a broad spectrum of licenses (Creative Commons 2016) that prevent commercial or derivative use of the dataset. It is worth knowing that using restrictive licenses may prevent reuse but still ensure reproducibility and transparency of the LCI model, thus enhancing data/model quality.
6. **Deploy data.** Publish the FAIRified LCA data, together with relevant metadata and a license in a trustworthy digital repository. Currently, generic repositories such as Zenodo and Figshare provide several tools that support FAIR compliance in research data sharing (particularly a PID and license). However, there is an increasing need for a development of LCA-specific repository with similar tools that can facilitate effective data discoverability and reuse among LCA practitioners.
7. **Use and acknowledge FAIR data.** Encourage and incentivize reuse FAIR data output. With the availability of FAIR LCA, data practitioners should be required to demonstrate that existing FAIR data resources are consulted and used to develop LCA models. For example, journals recognizing and rewarding the publishing and reuse of FAIR data are likely to encourage researchers (Neylon 2017; Collins et al. 2018; Bast 2019). Moreover, acknowledgement and credit should be given for all roles supporting FAIR data including data analysis, annotation, management, and curation.

5 Recommendations

Creating and sharing FAIR LCI datasets require some key elements such as publishing data and relevant metadata with a globally unique PID, annotated to semantically linked nomenclature, shared in machine-readable formats (e.g., JSON-LD), and providing appropriate license. However, enabling FAIR data sharing goes beyond the data. To enhance adoption of FAIR principles, additional recommendations are as follows:

Develop community-specific data repositories. Generic repositories do not have any requirements to ensure domain-specific standards, data formats, and metadata. These can be barriers to data discovery, interoperability, and reusability (Matthews 2022). These barriers can be overcome with the development of community-specific repositories. Concrete steps needed to ensure development of domain repositories include sustained funding, technical infrastructure support (for example, implement Application Programmable

Interfaces (APIs) to make data available to database developers and data users), provide basic tools like those provided generic repositories (for example, provide unique identifier, multiple licensing options, ease of access, version control), ensure trust and merit, acknowledge data providers and users, and a provision for community feedback to ensure data quality. Many of these steps could be adopted by the GLAD platform.

Evaluate FAIR data sharing. While there are no specific guidelines or metrics to validate the FAIR data sharing related to LCA data, FAIR metric tools developed by EOSC include *F-UJI* (a web service to assess FAIRness of research data objects based on metrics developed by the Horizon EU FAIRsFAIR project) (Devaraju and Huber 2020) and *FOOPs* (a scanner for FAIR principles of semantic ontologies) (Garijo and Poveda-Villalon 2020). These tools provide the possibility to benchmark and strategize data sharing for a specific domain based on the FAIR principles.

Strategic funding for data sharing. The use of LCA has grown exponentially with several research incentives to use the tool for decision support or methodology development. The domain can further benefit from strategic, sustainable, and coordinated investment in the development of research data infrastructures, data sharing tools, and services. While some efforts by funders (e.g., Horizon Europe) are increasingly mandating data management and sharing, researchers and data curators need to understand and implement data management plan to strategize and support FAIR management.

Develop technology for FAIR compliance. The tools and services required to fulfill the needs of data producers and users must be easy to adopt. Facilitating automated processing and interoperability frameworks is vital. This can be done with the support of semantic technologies (Ghose et al. 2022). Database search engine works by matching terms or understanding the interconnections between terms. Semantics gives meaning or explains the interconnections of disparate terms or data. Use of semantic technologies can support challenges related to findability and interoperability of data. However, there is a steep learning curve to understand and use semantic web technologies for researchers not accustomed to these models. Development of web applications that automate the process of annotating, integrating, querying, and using semantically linked data is required (Ghose et al. 2022).

Leverage training and skill development. Knowledge and use of data sharing infrastructure services, formats, tools, and workflows can be promoted with access to training. It is vital to develop and implement skills in data stewardship for researchers.

Recognize and reward the best practice in data sharing. Finally, the key barrier to data sharing is simply the lack of willingness to share data. Recognition and

rewarding FAIR data stewardship (Collins et al. 2018) might incentivize the willingness to share data. Citation of data and other research outputs needs to be encouraged. In addition, developing a core set of metrics at community level to recognize the contributions beyond publications and citations can be considered.

6 Conclusion

This study provides a broad overview of data sharing in practice and the available infrastructure for data curation for the LCA community. While there is awareness and motivation about data sharing in the LCA community, in practice data sharing seldom follows the FAIR guidelines. With respect to adopting the FAIR guidelines, LCA data curation is limited to sharing data in generic repositories that are free and provide multiple tools to enable the adoption of FAIR data sharing. Currently, LCA-specific data repositories do not provide the tools and services to enable FAIR data sharing. There have been some efforts in the development of LCA-specific vocabularies and metadata templates, but their use has been sporadic. There is a lack of guidelines and workflows to ensure FAIR data sharing. Data management plans of recent research projects performing LCA do not outline how data specifically collected and analyzed to perform LCA will be shared. This study advocates the development of a FAIR data ecosystem for the community and proposes a workflow using current tools as well as recommendations to incentivize LCA practitioners to share data based on FAIR principles. Based on the current available tools, this study also presents a workflow that can guide researchers to share FAIR-compliant data. While the workflow recommends the use of currently available tools for data curation, it also addresses the need data sharing incentives and platform specific for the LCA community. In conclusion, the study recommends further efforts on skill and technology development, strategic funding, and recognition of the best practices in data sharing.

Acknowledgements The paper was developed in relation to the data management plan developed for the ALIGNED project (Horizon Europe grant agreement no. 101059430). This paper was also inspired by several discussions with Ugo Javourez, Jorge Senan Salinas, Pierre Jouannais, Maddalen Ayala Cerezo, Massimo Pizzol, and Flora Champetier.

Funding Open access funding provided by Aalborg University The author would like to acknowledge the financial support of the European Union's Horizon Europe research and innovation program under grant agreement no. 101059430.

Data availability All supplementary information linked to this study are available in a Zenodo repository and can be found in <https://doi.org/https://doi.org/10.5281/zenodo.10136803>.

Declarations

Competing interests The author declares no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- ARGOS OpenAIRE (2022) ARGOS- what is it? <https://argos.openaire.eu/splash/about/how-it-works.html>. Accessed 1 Oct 2023
- Asem-Hiablie S, Battagliese T, Stackhouse-Lawson K, Rotz A (2018) A life cycle assessment of the environmental impacts of a beef system in the USA | SpringerLink. <https://link.springer.com/article/10.1007/s11367-018-1464-6>. Accessed 25 Jul 2023
- Australian Research Data Commons (2021) Research activity identifier (RAiD). In: RAiD. <https://www.raid.org.au>. Accessed 16 Sep 2022
- Bandrowski AE, Martone ME (2016) RRIDs: a simple step toward improving reproducibility through rigor and transparency of experimental methods. *Neuron* 90:434–436. <https://doi.org/10.1016/j.neuron.2016.04.030>
- Bast R (2019) A FAIRer Future *Nat Phys* 15:728–730. <https://doi.org/10.1038/s41567-019-0624-3>
- Brewster C, Nouwt B, Raaijmakers S, Verhoosel J (2020) Ontology-based access control for FAIR data. *Data Intell* 2:66–77. https://doi.org/10.1162/dint_a_00029
- Brose D (2011) Data needs for life-cycle assessment. https://sites.nationalacademies.org/cs/groups/pgasite/documents/webpage/pga_068711.pdf. Accessed 14 Sept 2022
- Collins, Sandra, Genova F, Harrower N, et al (2018) Final report and action plan from the European commission expert group on FAIR data. European Commission. <https://op.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF/source-80611283>. Accessed 14 Sept 2022
- Creative Commons (2016). <https://creativecommons.org/share-your-work/cclicenses/>. Accessed 1 Oct 2022
- Devaraju A, Huber R (2020) F-UJI - an automated FAIR data assessment tool (v1.0.0). Zenodo. <https://doi.org/10.5281/zenodo.3934401>
- DublinCore (2022) DCMI metadata terms. <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>. Accessed 17 Nov 2022
- European Commission (2018a) Life cycle data network. In: Eur Platf LCA EPLCA. <https://eplca.jrc.ec.europa.eu/LCDN/>. Accessed 11 Oct 2023
- European Commission (2018b) European platform on LCA | EPLCA. In: Nodes Approv. Wait. Approval. <https://eplca.jrc.ec.europa.eu/LCDN/contactListILCD.xhtml>. Accessed 11 Oct 2023
- European Commission (2020) Data governance and data policies - at the European Commission
- EOSC (2020) The EOSC portal. In: EOSC Portal. <https://eosc-portal.eu/about-eosc-portal>. Accessed 14 Apr 2023
- Figshare Team (2022) About Figshare. <https://knowledge.figshare.com/about>. Accessed 14 Apr 2023
- Fritter M, Lawrence R, Marcolin B, Pelletier N (2020) A survey of Life Cycle Inventory database implementations and architectures, and recommendations for new database initiatives. *Int J Life Cycle Assess* 25:1522–1531. <https://doi.org/10.1007/s11367-020-01745-5>
- Frontiers Author guidelines. <https://www.frontiersin.org/guidelines/author-guidelines>. Accessed 25 Jul 2023
- Garijo D, Poveda-Villalon M (2020) Best practices for implementing FAIR vocabularies and ontologies on the web. In: Applications and practices in ontology design, extraction, and reasoning. https://dgarijo.com/papers/best_practices2020.pdf. Accessed 11 Oct 2023
- Ghose A, Lissandrini M, Hansen ER, Weidema BP (2022) A core ontology for modeling life cycle sustainability assessment on the Semantic Web. *J Ind Ecol* 26:731–747. <https://doi.org/10.1111/jiec.13220>
- Giacovelli C (2020) About | GLAD. <https://www.globalcadataaccess.org/about>. Accessed 14 Apr 2023
- Go FAIR (2022) FAIRification process. In: GO FAIR. <https://www.go-fair.org/fair-principles/fairification-process/>. Accessed 15 Jun 2023
- Ingwersen WW (2015) Test of US federal life cycle inventory data interoperability. *J Clean Prod* 101:118–121. <https://doi.org/10.1016/j.jclepro.2015.03.090>
- ISO 14048 (2002) Environmental management - life cycle assessment - data documentation format
- Kahn E, Antognoli E, Arbuckle P (2022) The LCA commons—how an open-source repository for US federal life cycle assessment (LCA) data products advances inter-agency coordination. *Appl Sci* 12:865. <https://doi.org/10.3390/app12020865>
- Keller F, Voss RL, Lee RP, Meyer B (2022) Life cycle assessment of global warming potential of feedstock recycling technologies: case study of waste gasification and pyrolysis in an integrated inventory model for waste treatment and chemical production in Germany. *Resour Conserv Recycl* 179:106106. <https://doi.org/10.1016/j.resconrec.2021.106106>
- Köhler S, Pizzol M (2019) Life cycle assessment of bitcoin mining. *Environ Sci Technol* 53:13598–13606. <https://doi.org/10.1021/acs.est.9b05687>
- Kuczynski B, Marvuglia A, Astudillo MF et al (2018) LCA capability roadmap. *Int J Life Cycle Assess* 23:1685–1692. <https://doi.org/10.1007/s11367-018-1446-8>
- Kusche O (2020) Get involved! GLAD. <https://www.globalcadataaccess.org/get-involved>. Accessed 14 Apr 2023
- Kwon D (2020) The push to replace journal supplements with repositories. In: The scientist. <https://www.the-scientist.com/news-opinion/the-push-to-replace-journal-supplements-with-repositories-66296>. Accessed 24 Jul 2023
- Matthews T (2022) Community repositories: the best way to share the data underlying your research. In: Res. Data Community. <http://researchdata.springernature.com/posts/community-repositories-the-best-way-to-share-the-data-underlying-your-research>. Accessed 21 Jun 2023
- Michener WK (2015) Ten simple rules for creating a good data management plan. *PLOS Comput Biol* 11:e1004525. <https://doi.org/10.1371/JOURNAL.PCBI.1004525>
- Miranda Xicotencatl B, Kleijn R, van Nielen S et al (2023) Data implementation matters: effect of software choice and LCI database evolution on a comparative LCA study of permanent magnets. *J Ind Ecol* 27:1252–1265. <https://doi.org/10.1111/jiec.13410>

- Muñoz I, Rodríguez C, Gillet D, M. Moerschbacher B, (2018) Life cycle assessment of chitosan production in India and Europe. *Int J Life Cycle Assess* 23:1151–1160. <https://doi.org/10.1007/s11367-017-1290-2>
- Musen MA, O'Connor MJ, Schultes E et al (2022) Modeling community standards for metadata as templates makes data FAIR. *Sci Data* 9:696. <https://doi.org/10.1038/s41597-022-01815-3>
- Neylon C (2017) Compliance culture or culture change? The role of funders in improving data management and sharing practice amongst researchers. *Res Ideas Outcomes* 3:e14673. <https://doi.org/10.3897/rio.3.e14673>
- Nielsen LH, Smith T (2014). Zenodo Overview. <https://doi.org/10.5281/zenodo.8428>. Accessed 14 Apr 2023
- PANGAEA (2023) Data publisher for earth & environmental science. <https://www.pangaea.de/>
- Parvatker AG, Eckelman MJ (2019) Comparative evaluation of chemical life cycle inventory generation methods and implications for life cycle assessment results. *ACS Sustain Chem Eng* 7:350–367. https://doi.org/10.1021/ACSSUSCHEMENG.8B03656/ASSET/IMAGES/LARGE/SC-2018-03656C_0007.JPEG
- Pauliuk S, Heeren N, Hasan MM, Müller DB (2019) A general data model for socioeconomic metabolism and its implementation in an industrial ecology data commons prototype. *J Ind Ecol* 23:1016–1027. <https://doi.org/10.1111/jiec.12890>
- Pop M, Salzberg SL (2015) Use and mis-use of supplementary material in science publications. *BMC Bioinformatics* 16:237. <https://doi.org/10.1186/s12859-015-0668-z>
- Rotz A, Stout R, Leytem A et al (2021) Environmental assessment of United States dairy farms. *J Clean Prod* 315:128153. <https://doi.org/10.1016/j.jclepro.2021.128153>
- ScienceDirect (2023) Guide for authors. In: *J Environ Manag J Clean Prod*. <https://www.elsevier.com/journals/journal-of-cleaner-production/0959-6526/guide-for-authors>
- Shanahan H, Bezuidenhout L (2022) Rethinking the A in FAIR data: issues of data access and accessibility in research. *Front Res Metr Anal* 7. <https://doi.org/10.3389/frma.2022.912456>
- Springer (2023) Submission guidelines. In: *Submiss. Guidel. - Int. J. Life Cycle Assess.* <https://www.springer.com/journal/11367/submission-guidelines?IFA>
- Stenzel A, Waichman I (2023) Supply-chain data sharing for scope 3 emissions. *Npj Clim Action* 2:1–7. <https://doi.org/10.1038/s44168-023-00032-x>
- Thonemann N, Pizzol M (2019) Consequential life cycle assessment of carbon capture and utilization technologies within the chemical industry. *Energy Environ Sci* 12:2253–2263. <https://doi.org/10.1039/C9EE00914K>
- Venkataraman S (2022) Introduction to research data management and open research, CODATA/RDA Data Science Summer School, Trieste. <https://indico.ictp.it/event/9806/>. Accessed 14 Apr 2023
- Weber S, Peters JF, Baumann M, Weil M (2018) Life cycle assessment of a vanadium redox flow battery | *Environmental Science & Technology*. *Environ Sci Technol* 52:10864–10873. <https://doi.org/10.1021/acs.est.8b02073>
- Whyte A (2019) Towards FAIR data management plans, from principles to practice - Discussion Collaborative session notes: https://docs.google.com/document/d/1sMaf891ou4jb5s50hmtlhWGL5zTQSJaFPWxiF_APsv0/edit?usp=sharing. Accessed 14 Apr 2023
- Wilkinson MD, Dumontier M, Aalbersberg IJJ, et al (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2016 31 3:1–9. <https://doi.org/10.1038/sdata.2016.18>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.