A! Aalto-yliopisto
Aalto-universitetet
Aalto University

Master’s Programme in Automation and Electrical Engineering

Comparative analysis of Life Cycle Assessment data sharing models for supply chains

Ville Strengell

Master’s thesis
2024

Aalto-yliopisto
Aalto-universitetet
Aalto University

Author Ville Strengell

Title of thesis Comparative analysis of Life Cycle Assessment data sharing models for supply chains

Programme Automation and Electrical Engineering

Major Control, Robotics and Autonomous Systems

Thesis supervisor Prof. Valeriy Vyatkin

Thesis advisor(s) Dr. Udayanto Atmojo

Date 21.11.2024 **Number of pages** 41+5 **Language** English

Abstract

The reliability of Life Cycle Assessment (LCA) is highly dependent on the quality of the data used. This thesis examines the current landscape of LCA data sharing within supply chains and collaborative consortia, with a particular focus on three critical aspects: interoperability, data provenance and scalability. The study evaluates three key data sharing architectures, Blockchain, International Data Spaces (IDS) and a Trusted Operator system, as well as three of the most prominent data formats, ILCD, EcoSpold and JSON-LD. Using a qualitative comparison, the analysis addresses criteria such as efficiency, data provenance, immutability, scalability, and security. The results reveal that IDS provides the best balance of scalability and flexibility, making it suitable for complex supply chain applications. While the choice of data format was found to be less critical, compatibility with regional databases and commonly used software remains essential.

Keywords Life Cycle Assessment, LCA, Data Sharing, Blockchain, International Data Spaces, IDS

Aalto-yliopisto
Aalto-universitetet
Aalto University

Tekijä Ville Strengell

Työn nimi Comparative analysis of LCA data sharing models for supply chains

Koulutusohjelma Automation and Electrical Engineering

Pääaine Control, Robotics and Autonomous Systems

Vastuuopettaja/valvoja Prof. Valeriy Vyatkin

Työn ohjaaja(t) Dr. Udayanto Atmojo

Päivämäärä 21.11.2024 **Sivumäärä** 41+5. **Kieli** Englanti.

Tiivistelmä

Elinkaariarvioinnin (Life Cycle Assessment, LCA) luotettavuus riippuu merkittävästi käytetyn datan laadusta. Tämä tutkielma tarkastelee LCA-datan jakamisen nykytilaa toimitusketjuissa ja yhteistyökonsortioissa keskittyen erityisesti kolmeen osa-alueeseen: yhteentoimivuuteen, datan jäljitettävyyteen ja skaalaavuuteen. Tutkimuksessa arvioidaan kolmea keskeistä datan jakamisen arkkitehtuuria: lohkoketjuja (Blockchain), data-avaruuksia (IDS) sekä luotettuun operaattoriin perustuvaa mallia. Lisäksi tarkastelun kohteena ovat kolme yleisesti käytettyä tiedostoformaattia: ILCD, EcoSpold ja JSON-LD. Kvalitatiivisen vertailun avulla analyysi käsittelee kriteerejä, kuten tehokkuus, datan jäljitettävyys, muuttumattomuus, skaalaavuus ja tietoturva. Tulokset osoittavat, että IDS tarjoaa parhaan tasapainon skaalaavuuden ja joustavuuden välillä, mikä tekee siitä sopivan vaihtoehdon monimutkaisiin toimitusketjujen sovelluksiin. Tiedostoformaatin valinnan todettiin olevan vähemmän kriittinen, mutta yhteensopivuus alueellisten tietokantojen ja käytettävien ohjelmistojen kanssa on edelleen olennaista.

Avainsanat Life Cycle Assessment, LCA, Data Sharing, Blockchain, International Data Spaces, IDS

Preface and acknowledgements

I want to thank Professor Valeriy Vyatkin and my thesis advisor Dr Udayanto Atmojo for their good advice and guidance.

I also want to thank project partners Miro Eklund, Mateo Saavedra Del Oso, Max Ek, Hansani Perera and Tommi Karhela for insightful discussions and guidance throughout the process.

Otaniemi, 21 November 2024
Ville Strengell

Table of contents

Symbols and abbreviations

Abbreviations

CA	Certificate Authority
DAPS	Dynamic Attribute Provisioning Service
GDPR	General Data Protection Regulation
GHG	Greenhouse Gas
IDS	International Data Spaces
IDSA	International Data Spaces Association
IDS-RAM	IDS Reference Architecture Model
LCA	Life Cycle Assessment
LCI	Life Cycle Inventory Analysis
LCIA	Life Cycle Impact Assessment
ParIS	Participant Information Service
PCR	Product Category Rule
ILCD	International Reference Life Cycle Data System
ELCD	European Reference Life Cycle Database
PEF	Product Environmental Footprints
OEF	Organizational Environmental Footprints
XML	Extensible Markup Language
JSON-LD	JavaScript Object Notation for Linked Data
W3C	World Wide Web Consortium
SLA	Service Level Agreement
PPI	Pulp and Paper Industry
TPS	Transactions Per Second

1 Introduction

Sustainability has been important across various industrial ecosystems for decades, especially the pulp and paper industry (PPI). Globally, PPI produces over 400 million tonnes of paper and paperboard per year, of which European industry covers a quarter (FAO, 2018). This makes it one of the largest industrial contributors to energy and water usage, deforestation, greenhouse gas (GHG) emissions and landfill waste (Furszyfer Del Rio et al., 2022).

The importance of sustainable practices is ever increasing, as the production of paper is expected to rise to over 900 million tonnes per year by 2050 (Furszyfer Del Rio et al., 2022), and environmental issues are becoming a larger concern within the society (Schandl et al., 2017). This has caused increasing pressure on companies and governments to decrease waste and emissions, and to operate in a more ecologically responsible manner (Velte, 2021).

The PPI produces emissions and waste to air, water and land, during all its stages, from gathering of materials to papermaking processes and waste disposal. Therefore, accurately measuring the environmental impact and identifying the specific causes for a finished product is complicated.

One of the most common ways for businesses to measure their sustainability is through Life Cycle Assessment (LCA), a well renowned and standardized method of quantifying the environmental footprint of a product, service or material (British Plastics Federation, 2024). LCA examines the entire life cycle of a product, from the extraction of resources to production, use, recycling, and the disposal of remaining waste (*European Commission - Joint Research Centre - Institute for Environment and Sustainability,* 2010). It looks beyond the direct effects, and investigates, for example, the energy used in production, fuel used in transport, and end-of-life ecological costs.

LCA consists of four main stages, Definition of Goal and Scope, Inventory Analysis, Impact Assessment and Interpretation (Ecochain, 2024). The second stage, Life Cycle Inventory Analysis (LCI), involves the collection of data on environmental inputs and outputs related to a product. The inputs and outputs represent all flows within the product's supply chain, including raw materials, water consumption, emissions to air, water and land, and energy use.

The collection of data for LCI is often complex, due to the nature of supply chains. Especially if the inputs and outputs are sourced from a large array of companies, accurate LCA reporting relies heavily on the effortless flow of data between different entities. However, the process is complicated by differing standards and reporting methods across companies, as well as concerns over data security. Because environmental data is often sensitive, companies are hesitant to share their data without complete trust in the data protection and access control (Lin et al., 2021), which is difficult to establish.

In addition, essentially every environmental input has its own chain of processes behind it. Consequently, simplifications are often necessary, and industry averages are used instead of primary data. These industry averages are called secondary data. Using secondary data is necessary, because obtaining primary data for every input is unfeasible. However, excessive secondary data usage causes uncertainty because of several factors, such as geographical, temporal and technological differences (Cellura, Longo and Mistretta, 2011).

Stora Enso, a Finnish forest industry company, is facing similar challenges within its pulp supply chain. Stora Enso collects primary data from its suppliers with yearly questionnaires, which is labour intensive and prone to errors. There is a clear gap to improve these practices with modern data sharing tools.

The aim of this thesis is to evaluate the current landscape and state-of-the-art methods for LCA data sharing within supply chains and other consortia, with a particular focus on aspects such as interoperability, data provenance and scalability. This will assist Stora Enso in making informed decisions about their future data sharing strategies. The thesis will explore various data sharing architectures, such as International Data Spaces (IDS), Blockchain, and a trusted operator system, comparing their suitability to Stora Enso's needs.

To better evaluate the alternative LCA data sharing platform and assess clearer the challenges during development and maintenance of a data portal, a proof-of-concept portal was developed as a part of this thesis. This portal serves as a hands-on demonstration of how a data portal can be employed and how it benefits the process of data collection.

This thesis is focused specifically on the primary data collection phase within the pulp supply chain of Stora Enso. Broader topics, such as Life Cycle Impact Assessment (LCIA) and the interpretation of results, remain outside the scope of this research.

The remainder of this thesis is divided into four chapters. Chapter 2 reviews the literature on LCA data sharing formats and architectures. Chapter 3 compares the data formats and architectures, evaluates them, and presents the proof-of-concept data portal. Chapter 4 provides a summary of the thesis.

2 Background

2.1 Data formats and Standards

One of the major criticisms LCA receives is the inherent subjectivity and variability in its outcomes (Barahmand, Z. and Eikeland, 2022). Multiple practitioners can arrive at different results for the same product or process due to the decisions they make at various stages of the assessment. Factors such as quality of data, choice of LCIA method, and assumptions about system boundaries all influence the results. These factors are debatable, and the decisions are often shaped by the practitioner's perspectives or external motivations, including economic and political incentives. For instance, companies may seek to influence the assessment to present their products in a more favourable light by selecting parameters that minimize perceived environmental impact.

Standardization plays a critical role in addressing this variability and ensuring consistency in LCA results. General standards, such as ISO 14040 and 14044, lay the foundational framework for conducting LCAs. ISO 14040, primarily written for a managerial audience, outlines the principles and framework of LCA, while ISO 14044 provides detailed requirements and guidelines for practitioners (Matthews, Hendrickson, and Matthews, 2014). These standards define how LCAs should be structured, ensuring that assessments follow a standardized process; goal and scope definition, inventory analysis, life cycle impact assessment, and interpretation (ISO, 2006a).

In addition to these general standards, more specific guidelines exist, such as National Standards and Product Category Rules (PCRs). These build on the broader principles set by ISO but provide sector-specific guidance, helping to standardize LCAs for product categories. PCRs, for example, offer detailed instructions for assessing specific product types, ensuring that LCAs within the same category are comparable.

Consistent and transparent documentation of data is also crucial for ensuring the reliability of LCAs. The technical ISO standard 14048 provides the requirements and the structure for a data documentation format to be used for transparent and unambiguous exchange of LCA data.

It states that LCA data documentation formats should consist of three parts: process, modelling and validation, and administrative information. The data format should also accommodate for metadata, such as the source of the information, time period for its validity and geographic region of validity.

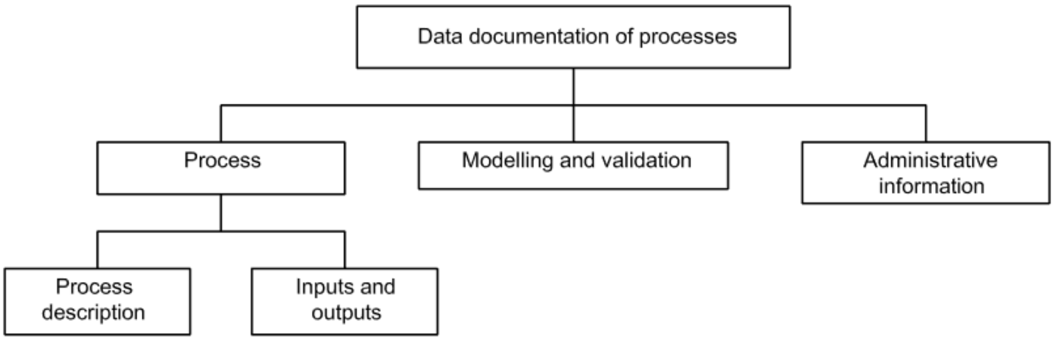


Figure 1. The Data documentation format of ISO 14048.

Processes are at the core of ISO 14048, where a process can be a unit process or a combination of unit processes. Unit processes refer to individual steps in the life cycle of a product. For each unit process, specific data flows must be documented, including inputs from the environment and resources, e.g. consumption of wood or freshwater usage, and outputs to disposal and to the environment, e.g. waste and emissions (ISO, 2002). These flows are standardized through a reference flow, which quantifies the process output. For instance, the emissions may be quantified as “x kg of CO2 per 1 ton of produced pulp”.

Given these ISO requirements, different data formats have been developed to ensure that LCA practitioners can efficiently document, manage, and share life cycle data. Formats such as ILCD and EcoSpold2 have become widely adopted in the industry. The following sections will explore these formats in more detail.

2.1.1 International Reference Life Cycle Data system (ILCD)

The International Reference Life Cycle Data System (ILCD) was developed by the European Joint Research Centre (European Commission - Joint Research Centre, 2010). The development was driven by several key requirements. First, it aimed to create a data format for the European Reference Life Cycle Database (ELCD). Second, it sought to establish a common format to support data exchange, both importing and exporting ELCD reference datasets between various databases and software tools. Additionally, ILCD was developed to facilitate the transfer of LCA datasets along supply chains and between different data networks. More recently, it has also been used to support data under the Environmental Footprint scheme, including Product Environmental Footprints (PEF) and Organizational Environmental Footprints (OEF). (Life Cycle Initiative, 2022)

The ILCD format is widely supported by most major LCA software applications and has been adopted by numerous national LCA databases. Notable examples include the European ELCD, the SICV database in Brazil, MYLCID in Malaysia, and the Thai National LCI database (Life Cycle Initiative, 2023).

An ILCD file is referred to as a package, and it consists of multiple folders, each representing a key component of LCA data. These components, shown in Figure 2, are structured to enhance organization and data accessibility. These components are made up of smaller elements, such as individual processes or flows, which represent specific data points within the LCA structure. They are illustrated in Figure 3.

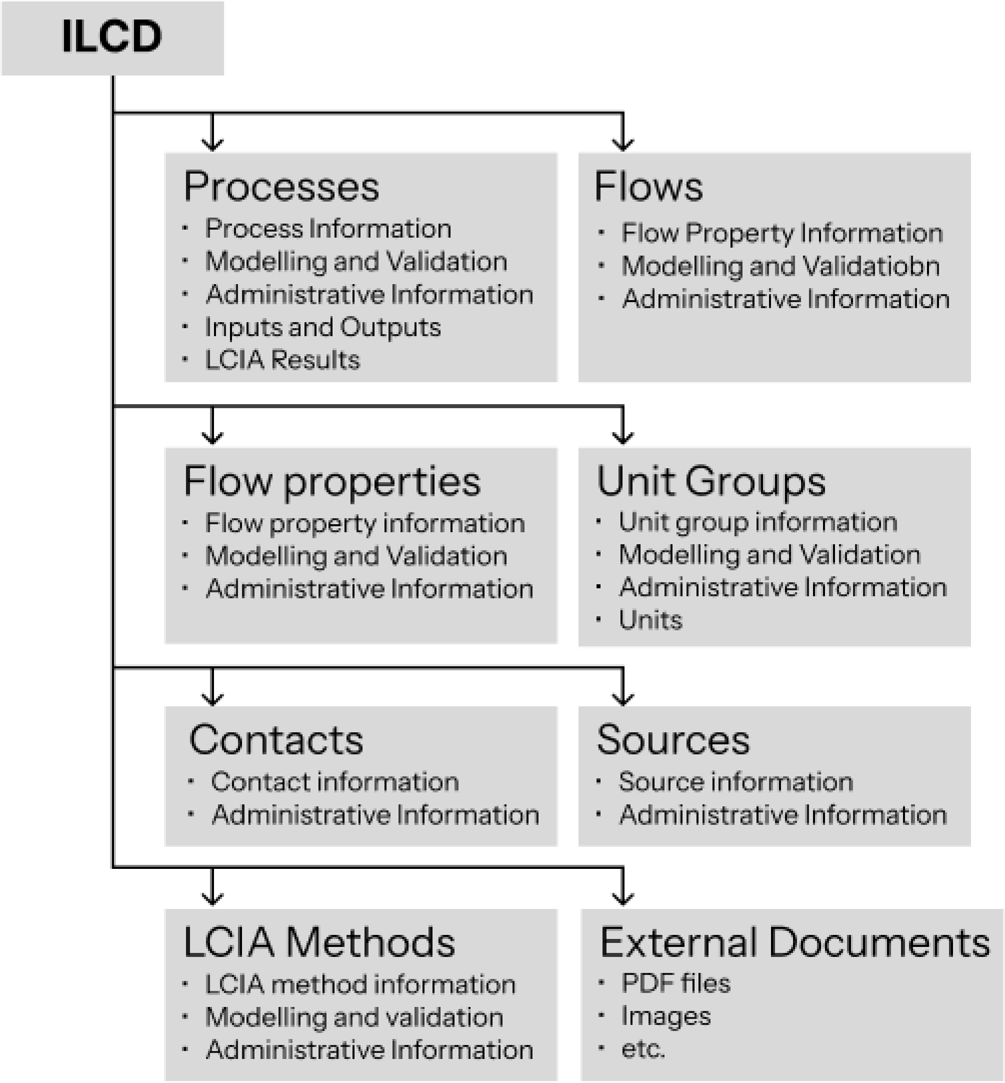


Figure 2. ILCD's structure and key components

The individual elements hold the actual data. They are Extensible Markup Language (XML) files that are structured according to the data documentation format of ISO 14048. They include sections for general information, as well as modelling and validation, and administrative information, although not all elements contain all three. Figure 3 illustrates the structure of process and flow data sets in greater detail.

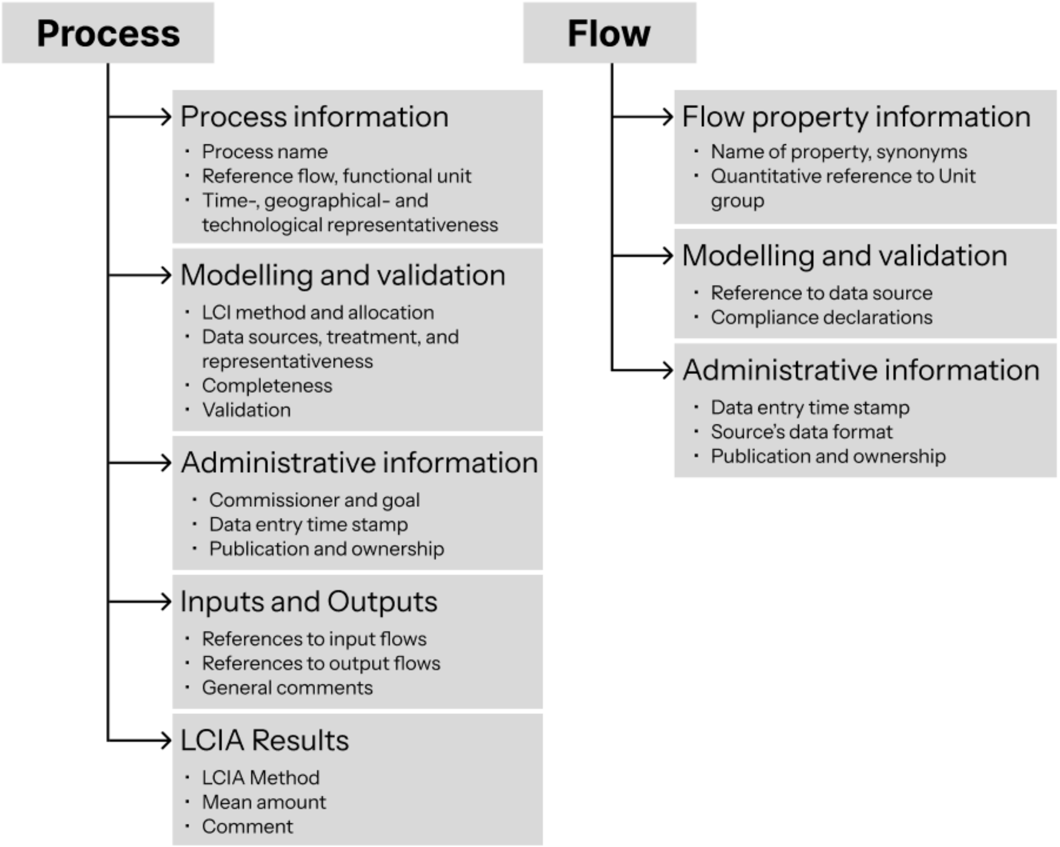
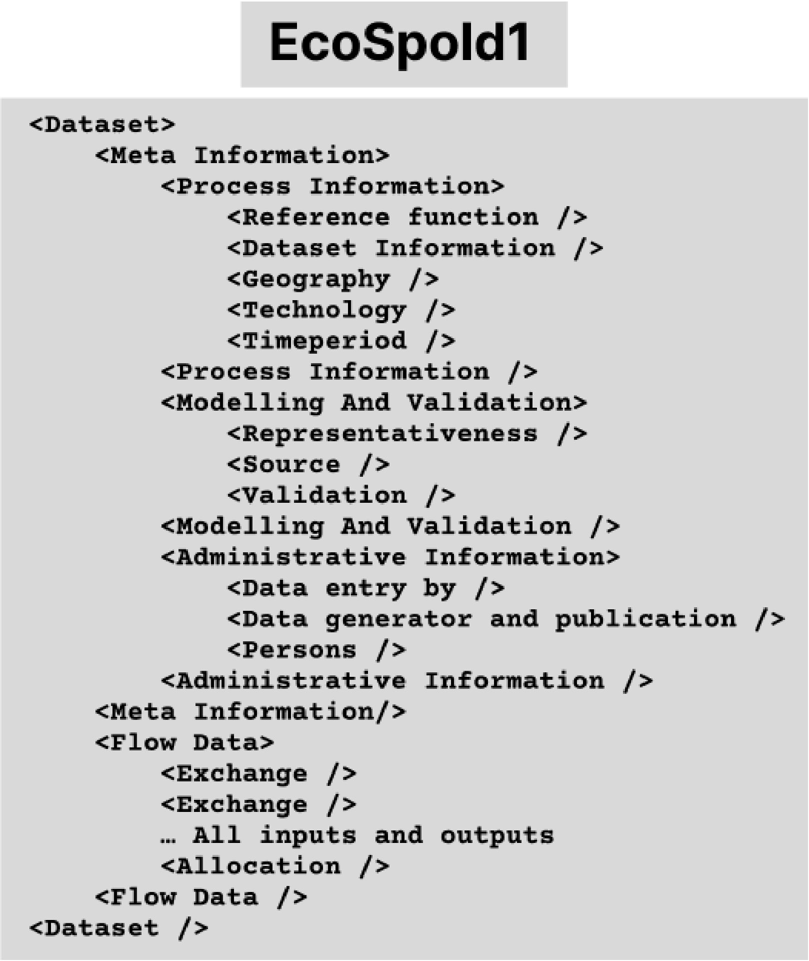

Process
Process information
Process name
Reference flow, functional unit
Time-, geographical- and technological representativeness
Modelling and validation
LCI method and allocation
Data sources, treatment, and representativeness
Completeness
Validation
Administrative information
Commissioner and goal
Data entry time stamp
Publication and ownership
Inputs and Outputs
References to input flows
References to output flows
General comments
LCIA Results
LCIA Method
Mean amount
Comment
Flow
Flow property information
Name of property, synonyms
Quantitative reference to Unit group
Modelling and validation
Reference to data source
Compliance declarations
Administrative information
Data entry time stamp
Source's data format
Publication and ownership

Figure 3. Structure of process and flow datasets

The XML files are named by their Universally Unique Identifiers (UUID), making the raw format difficult for humans to interpret. ILCD is designed for machine readability and LCA practitioners can easily navigate and manage the data using appropriate software. This design ensures that the data is not only secure and unambiguous but also interoperable across various LCA tools, reducing the likelihood of errors or inconsistencies during data exchanges. (PRé Sustainability, 2019)

2.1.2 EcoSpold

The EcoSpold data format is an XML-based standard developed for the exchange and storage of Life Cycle Inventory (LCI) data. It has been developed iteratively since the 1990s to address the growing needs of LCA practitioners for a structured and reliable format. The earlier version, EcoSpold1, introduced in the early 2000s, is simpler and lacks some advanced features. Although compliant with ISO 14048 and supported by major LCA software systems, EcoSpold1 is no longer widely used today (GreenDelta, 2024). Most software that support it uses some sort of dialect, making interoperability between databases and software difficult (GreenDelta, 2015). Its simplicity and standardization helped it become one of the first widely adopted LCA data formats, but the absence of key features, such as parameters and formulas, limited its long-term applicability. It is primarily used by the Ecoinvent v2 database.



EcoSpold1

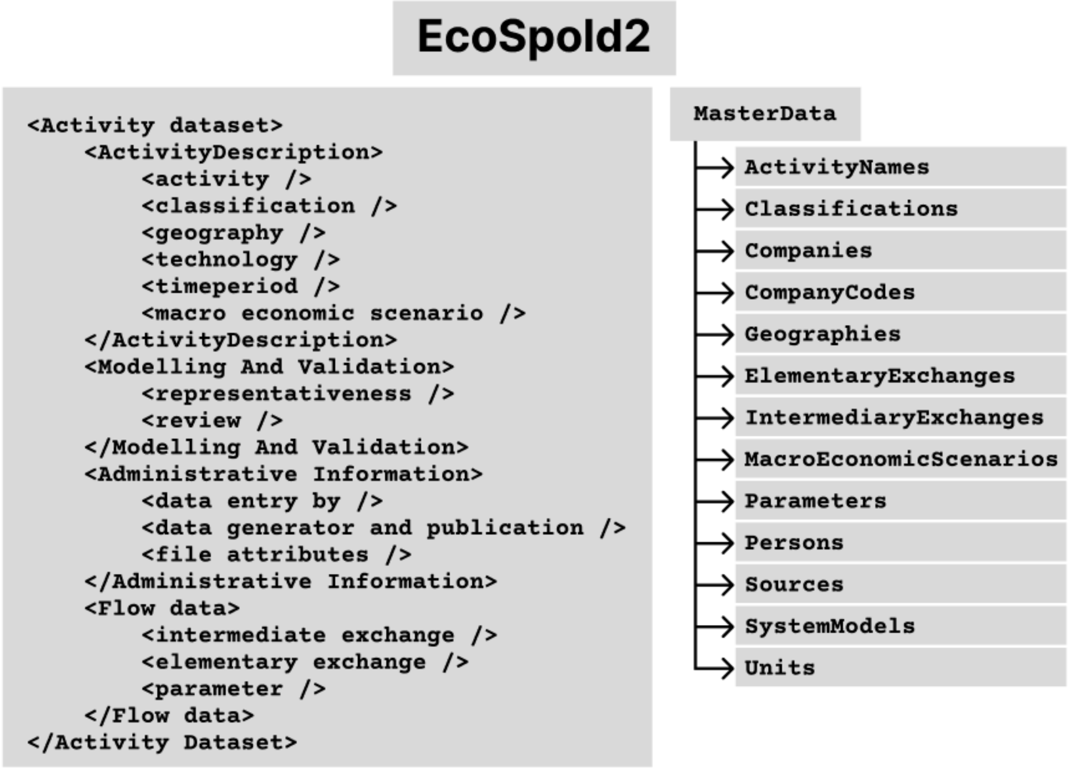
```
<Dataset>
    <Meta Information>
        <Process Information>
            <Reference function />
            <Dataset Information />
            <Geography />
            <Technology />
            <Timeperiod />
        <Process Information />
        <Modelling And Validation>
            <Representativeness />
            <Source />
            <Validation />
        <Modelling And Validation />
        <Administrative Information>
            <Data entry by />
            <Data generator and publication />
            <Persons />
        <Administrative Information />
    <Meta Information/>
    <Flow Data>
        <Exchange />
        <Exchange />
        … All inputs and outputs
        <Allocation />
    <Flow Data />
<Dataset />
```

Figure 4. Structure of EcoSpold1 Dataset

EcoSpold2 builds upon EcoSpold1 and incorporates various updates to cater to the evolving requirements of LCA. It was mainly developed for the Ecoinvent v3 database, but it is open source, allowing for broader community input and usage beyond the Ecoinvent Centre. This open approach ensures that EcoSpold2 is flexible enough to be adapted to the needs of various LCA software tools and datasets, as noted by Meinshausen, Müller-Beilschmidt, and Viere (2016).

The introduction of EcoSpold2 resolved both the structural limitations of EcoSpold1 and addressed new requirements from LCA software. One of the key structural improvements in EcoSpold2 was the enhanced use of master lists, which are referenced by universally unique identifiers (UUIDs) (Ecoinvent, 2023). In EcoSpold1, information was stored directly within each dataset, whereas EcoSpold2 uses UUIDs in the dataset to link to records within the MasterData folder, which contains all master lists. This approach prevents redundancy, so if, for example, freshwater is referenced multiple times, only a single entry is needed. Additionally, terminology was updated: processes are now called activities, and elementary flows have been renamed Exchanges with the environment (Ecoinvent, 2023).

Overall, EcoSpold2 is designed to handle all three key stages of LCA, Unit Process Records (UPR), Life Cycle Inventory (LCI), and Life Cycle Impact Assessment (LCIA). The inclusion of new features like exchange properties, mathematical formulas, variable names for exchanges and parameters, along with enhanced documentation and data referencing capabilities using UUIDs, allowed EcoSpold2 to meet the more complex requirements of modern LCA software, while facilitating data exchange and database maintenance on a larger scale. (Ecoinvent, 2023)



```
<Activity dataset>
    <ActivityDescription>
        <activity />
        <classification />
        <geography />
        <technology />
        <timeperiod />
        <macro economic scenario />
    </ActivityDescription>
    <Modelling And Validation>
        <representativeness />
        <review />
    </Modelling And Validation>
    <Administrative Information>
        <data entry by />
        <data generator and publication />
        <file attributes />
    </Administrative Information>
    <Flow data>
        <intermediate exchange />
        <elementary exchange />
        <parameter />
    </Flow data>
</Activity Dataset>
```

Figure 5. Structure of EcoSpold2 Activity Dataset and an example of the MasterData folder's contents

2.1.3 JSON-LD

JavaScript Object Notation for Linked Data (JSON-LD), specified by the World Wide Web Consortium (W3C), is a lightweight format designed to represent linked data using the popular JSON format (W3C, no date). One of its significant advantages is its compatibility. A JSON-LD document is always also a valid JSON document. Therefore, all of the standard JSON libraries work seamlessly with JSON-LD documents (W3C, 2020). This also makes

JSON-LD highly adoptable, as most developers are already familiar with JSON.

JSON-LD in itself is not compliant with the ISO 14048 standard, as it lacks any inherent structure. However, because LCI (Life Cycle Inventory) data contains numerous links, JSON-LD, a lightweight format for linked data, presents itself as a fitting option for structuring LCI data. Additionally, the existence of a JSON-LD representation for provenance, PROV-JSONLD, could be advantageous (Huynh, Michaelides and Moreau, 2016).

GreenDelta has used JSON-LD to develop openLCA schema, an LCI compliant alternative to XML, by combining elements from EcoSpold, ILCD and SimaPro CSV data formats (GreenDelta, 2024). It has been in use by OpenLCA since 2015.

2.2 Architectural Approaches to Data Sharing

When selecting a data sharing architecture, one must consider several relevant aspects, including efficiency, security, interoperability, data provenance, data integrity, scalability and trust. As companies try to seamlessly collaborate and exchange data, the underlying choice of architectural model becomes key. An effective architecture ensures that data sharing is not only secure and controlled but also efficient and scalable.

One fundamental requirement for any data-sharing network is efficiency. In supply chains, this means facilitating data access and transfer in a way that minimizes manual effort. In the context of this thesis, data transfer occurs once a year across multiple suppliers, making it crucial that the system allows users, primarily those retrieving and processing the data, to quickly and intuitively access the necessary information. While it is less critical for the data providers to have an exceptionally streamlined experience, due to the infrequent nature of these exchanges, efficiency also means that as the network grows, the system should still allow smooth interactions without adding too much extra work for users or straining the system. (SCSN, 2024)

Other concepts which are as important are data security and sovereignty. Data sovereignty refers to a company's ability to retain control over its data, including the right to delete it or regulate its use even after it has been shared. This becomes a concern any time a company shares confidential data, which LCI data often is. Without complete trust in the receiving party and its ability to keep the data secure, companies are hesitant to share their data. Protection of their own data has a higher priority than sharing it to benefit the entire supply chain (Lin et al., 2021). The architecture must incorporate mechanisms that ensure that the data is only accessed and modified by properly authorized entities.

Data immutability and integrity must also be ensured. Once data is entered into the system, no one should be able to alter it, unless a legitimate reason arises. To maintain a complete and transparent record, any changes must be logged with details about what was changed, by whom, and when.

This ties into the concept of data provenance. When the data is processed, there should be a clear indication of its origin, providing the historical context and authenticity of data.

Another critical consideration is the balance between data transparency and confidentiality. In many supply chains, certain transactions or partnerships involve sensitive information or confidential business relationships. For example, some companies may want to maintain anonymity or restrict access to specific data points to prevent the exposure of sensitive information. There may be instances where companies prefer to limit the visibility of their data to safeguard proprietary information or maintain confidential partnerships. Therefore, the architecture should incorporate mechanisms that allow selective transparency, ensuring that while data sharing is functional, sensitive relationships and information are protected. (Hyperledger, 2024)

When considering data-sharing architectures, three models are commonly identified: the two-corner (Point to point), three-corner (Hub and spoke), and four-corner models. The two-corner model is the simplest form of data exchange, where every entity within a network forms a unique connection to each other. An example of this is a network where all data is exchanged via direct emails. While this model ensures a high level of security, immutability and sovereignty with minimal initial setup, it scales incompetently. For a network of N companies, every new entity that enters the network, N new connections must be made. (SCSN, 2024)

The three-corner, or Hub and Spoke model, addresses the scaling issue by introducing a cloud platform or broker model where an intermediary IT party connects multiple companies. Thus, every company only needs to establish a connection to them. When a new company joins the network, they simply form one connection to this central broker, who takes care of the particularities. While this improves scalability, it introduces a central point of control. The reliance on a single intermediary can raise concerns about data sovereignty and trust, because the central intermediary gains access to the data and its frequency. (SCSN, 2024) (Derilinx, 2024)

The four-corner model is typically a more advanced architecture that further optimizes digital transactions along multiple parties. Instead of a single central provider, a four-corner model introduces a network of service providers, who communicate and facilitate data exchange between each other. Each company is free to choose their own provider. This fixes the issue of single point of control while maintaining scalability. However, this model is complex to implement and maintain, and it often involves a managing body, such as the SCSN Foundation or International Data Spaces Association (IDSA, 2024).

2.2.1 Data portal with a trusted operator

When considering how to share data among a network of participants who may not fully trust each other, a common solution is to hire a trustworthy third party. All parties would form a connection to them and the third party would manage the data exchanges while retaining sensitive information from the rest of the network.

There are numerous software providers offering platforms for secure business to business data sharing, such as Databricks Delta Sharing, Adeptia Connect or Vendia Share. To find the most suitable provider for a given situation, one would have to contact these providers separately to discuss the details and the estimated cost of the system. This approach is straightforward to implement, because only the functions matter, and as long as the data is stored centrally and not on the premises of every participant, the technology behind the chosen platform can be treated as a black box. If trust with the provider is established legally or contractually, and all necessary data management practices are assured through service level agreements (SLAs) or certifications, the underlying technology should not matter as much. For instance, the provider could guarantee compliance with the ISO 27001 standard for information security. Also, a provider with a strong reputation and a solid customer base has a shared interest in maintaining credibility, which serves as an additional assurance of reliability.

However, with this method, the data, which may be sensitive, is accessible by the third party and might raise concerns over data ownership and sovereignty. Even if the data was encrypted, gaining insights such as data frequency and business relations within the network could be valuable and might cause problems if they were to end up in the wrong hands. As stated by Lin X. et al (2021), some companies could be unwilling to participate in the data sharing if this is the case.

Another approach, which circumvents this distrust of third parties, would be for Stora Enso to act as the trusted operator itself, because they are the ones utilizing the data in the end. By doing so they would also have more control over the data management process and more flexibility to customize the platform to their exact situation, in the present and in the future. The obvious downside is the cost, both timely and monetary, of developing such a platform.

Outside the ones discussed in later sections, several architectures could be considered. A basic file-based transfer system, e.g., Secure File Transfer Protocol (SFTP) server with a tool like WinSCP, could support secure, encrypted transfers with minimal setup, though this limits scalability and is closer to a two-corner model than three-corner model.

A custom platform with Role-Based Access Control (RBAC) and an authentication standard like OAuth 2.0 would enable fine-grained permissions for different users within the supply chain. In addition, these solutions could be layered with logging and auditing mechanisms to ensure data traceability and meet compliance standards. This option is further explored by introducing the proof-of-concept data portal in section 3.3.

2.2.2 Blockchain

Blockchain is a data structure, that can be described as a decentralized digital ledger of transactions. It records these transactions securely across a peer-to-peer network, without a middleman. So, unlike traditional centralized systems, it does not require a governing body. Instead, every participant holds their own copy of the entire ledger locally. (Hackius, Petersen, 2017)

When a transaction is made between two participants, it is verified using public-private-key cryptography and added to a new block. A block can represent a single transaction or a group of transactions. All members of the network can verify the transactions in a block, and if a consensus is reached, the block is added to the ledger. If not, the block is rejected. For each block, a cryptographic hash is generated. In addition to the transaction records, each block also contains the hash of the previous block. Thus, every block is linked to all previous blocks, and a Blockchain is formed. (Hackius, Petersen, 2017)

If a malicious entity wanted to alter a transaction within the chain, they would not only have to alter it in most of the member's local devices, but also create a new cryptographic hash for every block after it. These are both extremely difficult tasks computationally. Thus, it can be said that immutability within a blockchain is granted. (Pilkington, 2016).

Reaching consensus is a critical step in a blockchain network to ensure the integrity of data. In traditional public, or permissionless, blockchains, consensus is often achieved through energy-intensive protocols like Proof of Work (PoW) or alternatives such as Proof of Stake (PoS). By contrast, private or permissioned blockchains use more efficient consensus protocols, such as Proof of Authority (PoA), Federal Byzantine Agreement (FBA), Practical Byzantine Fault Tolerance (PBFT) and Raft (Lashkari, B. & Musilek, P., 2021). These methods support data consistency without placing excessive demands on computational resources.

The best proof of the reliability of the blockchain as a technology is the cryptocurrency Bitcoin. It has been running since 2009 with no errors in its functionality or safety (Hackius, Petersen, 2017), even though it has a market capitalization of over 100 billion US dollars (Forbes, 2024).

In the context of LCA and supply chain management, blockchain is a promising technology due to its innate properties of data provenance security, immutability and transparency. This does not mean that it comes without limitations and challenges.

In 2020 a blockchain-based LCA (BC-LCA) framework was proposed by Zhang A. et al., and later refined in 2021 by Lin X. et al. to improve the availability, privacy, accuracy and the timeliness of LCI data, with less manual work. The benefits of using blockchain relies on its decentralized nature and built-in security and data provenance. The paper proposed a system architecture, where together with internet of things (IoT) and big data analytics, the entire LCA process could be automated with minimal human interaction (Zhang A., et al., 2020).

As Stora Enso requires only infrequent data updates, the proposed integration of IoT devices for real-time updating capability is excessive. However, the potential of blockchain technology to ensure data provenance remains highly relevant. One challenge arises with the nature of permissioned blockchains, where every participant is identified, and all transactions are transparent within the network. Some confidential relationships may require anonymity, and access to certain transactions may have to be restricted. Therefore, private sub-networks can offer a valid solution. In Hyperledger Fabric, these sub-networks are called channels.

Hyperledger is an open-source umbrella project initiated by the Linux Foundation in 2015 (The Hyperledger White Paper Working Group, 2018). It offers a wide range of tools and resources for companies to develop their own blockchain applications, such as the channels for more private data sharing. There are many Hyperledger frameworks, from which Fabric is the most widely used. As is mentioned in the Hyperledger Fabric's documentation, earlier blockchain platforms are being adapted for enterprise use, while Fabric has been designed for it from the beginning.

Hyperledger Fabric has a modular architecture, which provides organizations with the flexibility to configure the blockchain network according to their specific requirements. Key configurable components include cryptographic libraries and consensus protocols, identity and key management, making Fabric especially suited to environments where data privacy and selective transparency are essential.

In supply chains and LCA, where data privacy and selective transparency are essential, Fabric offers the concept mentioned earlier: channels. Channels are essentially private sub-networks within the larger Fabric network, allowing certain participants to conduct transactions that are visible only to those within a given channel. For instance, two companies can conduct data exchange privately without exposing anything to the rest of the network, preserving the confidentiality of sensitive information while still benefiting from the integrity and security of the blockchain.

2.2.3 International Data Spaces

International Data Spaces (IDS) is a framework designed to enable secure and standardized data exchange between organizations, with emphasis on data sovereignty and full control over shared information (IDSA, 2024). IDS aims to balance conflicting aspects of data management, ensuring data ownership, security, and value protection while promoting interoperability and sharing.

IDS is grounded in several foundational principles, that guide its structure and implementation: data sovereignty, security, decentralization, governance, openness, and trust (Nagel & Lycklama, 2021). Based on these principles, are the building blocks of IDS.

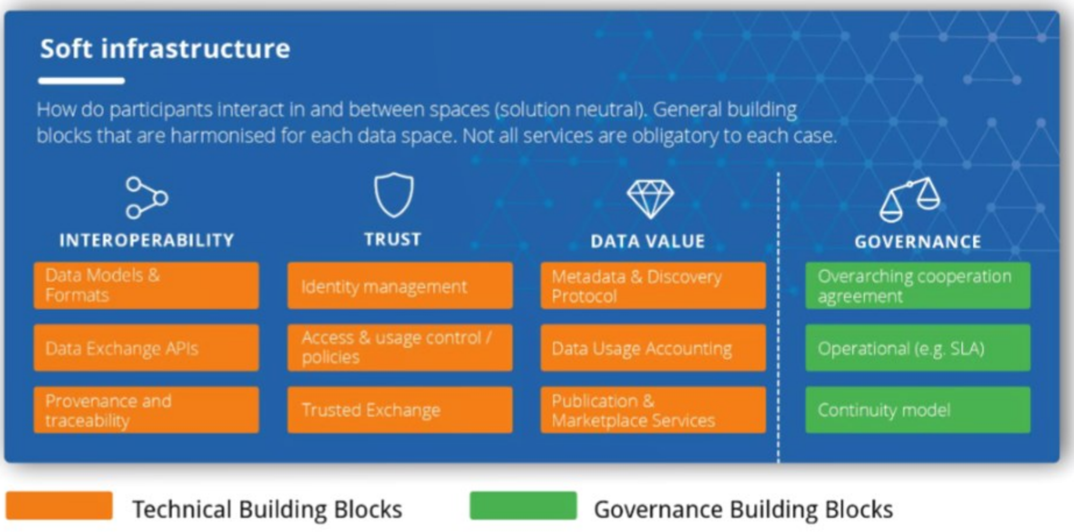
Technical Building Blocks (Nagel & Lycklama, 2021):

- Data Interoperability: Ensures compatibility of APIs, data formats, and traceability.
- Data Sovereignty: Manages identity, access, and trustworthiness.
- Data Value Creation: Facilitates data offerings, discovery, and accounting.

Governance Building Blocks (Nagel & Lycklama, 2021):

- Business Agreements: SLAs, data policies, and pricing models, often supported by smart contracts.
- Operational Agreements: Policies for compliance with GDPR and other directives.
- Organizational Agreements: Governance bodies and procedures defining data space operations.

These elements create sector-agnostic, customizable building blocks, aligning each data space with IDS Reference Architecture Model (IDS-RAM) while allowing flexibility to meet specific industry needs.



Soft infrastructure

How do participants interact in and between spaces (solution neutral). General building blocks that are harmonised for each data space. Not all services are obligatory to each case.

INTEROPERABILITY	TRUST	DATA VALUE	GOVERNANCE
Data Models & Formats	Identity management	Metadata & Discovery Protocol	Overarching cooperation agreement
Data Exchange APIs	Access & usage control / policies	Data Usage Accounting	Operational (e.g. SLA)
Provenance and traceability	Trusted Exchange	Publication & Marketplace Services	Continuity model

🟧 Technical Building Blocks 🟩 Governance Building Blocks

Figure 6. Data spaces building blocks (Nagel L., Lycklama D. 2021)

Data spaces represent a higher architecture model, and there are almost as many options for them as there are companies, domains of use and individual needs (Nagel L., Lycklama D. 2021). The IDS-RAM forms the structural foundation of IDS, offering guidelines and components for configuring such data spaces (IDSA, 2022).

IDS-RAM defines key roles for data spaces, which include data providers, consumers, and intermediaries. Providers supply data, consumers utilize it, and intermediaries enable data flow between parties. Intermediary roles like the broker, app store, and clearing house enhance data discoverability, usability, and accountability within IDS frameworks. (IDSA, 2022)

IDS-RAM includes several key components, which are the technical implementations to fulfill roles. These components include:

- IDS Connector: Each data provider and consumer in the IDS network implements its own IDS connector, which is used to interact with the other IDS connectors within a data space. In a data space, all data exchange happens directly between two connectors, ensuring that the intermediary parties do not gain access to the data. Data connectors also enforce data usage policies, which are set by the data provider. They support additional functionality through IDS Apps, allowing for custom operations such as data transformations, analytics, and enhanced usage controls. Connectors can be operated on-premises or in a cloud environment. (IDSA, 2022)
- Metadata Broker: Serving as the search engine of a data space, Metadata Broker stores the metadata and endpoints, known as self-descriptions, of every connector. It offers an interface for data consumers, and facilitates data discoverability by indexing and managing searchable data descriptions, linking connectors to relevant data sources. Technically, the Metadata Broker is an IDS Connector with specialized endpoints for registration, publication, and querying. (IDSA, 2022)
- The Identity Provider ensures secure identity and access management through three main components: Certificate Authority (CAs), which issues and manages identity certificates, the Dynamic Attribute Provisioning Service (DAPS), which provides real-time identity updates via dynamic tokens, and the Participant Information Service (ParIS), which maintains verified business-related information on IDS participants. Together they enable reliable identification and authorization within the data spaces. (IDSA, 2022)
- Clearing House records each data transaction, enabling auditing and enforcing data provenance. (IDSA, 2022)
- Vocabulary Hub standardizes terminology across data spaces, improving interoperability by ensuring a shared understanding of data semantics across systems. (IDSA, 2022)

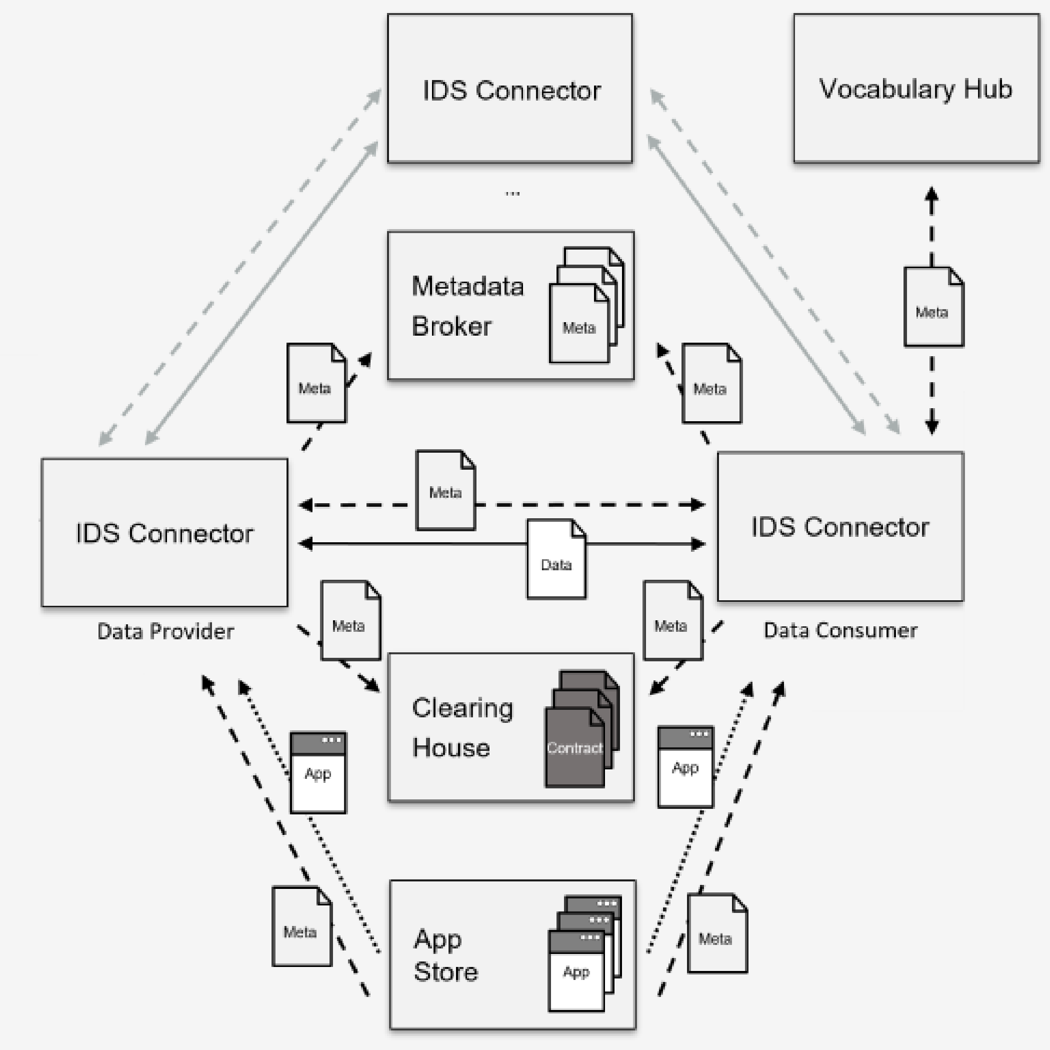


Figure 7. Interaction of IDS components within a data space.

IDS-RAM follows the FAIR data principles, which have been defined by an international group of researchers and organizations, and formally presented in a 2016 paper by Mark D. Wilkinson et al. (Steinbuss, S., 2020).

The FAIR principles map directly to IDS-RAM components (Steinbuss, S., 2020):

- Findable: The IDS Broker enables data discoverability by indexing the metadata of connectors, making datasets easier to locate.
- Accessible: The IDS Connector ensures standardized data access by managing digital identities, authentication, and authorization.
- Interoperable: The IDS Information Layer's information model fosters semantic interoperability, ensuring connectors can understand both the syntax and usage policies of shared data.
- Reusable: Usage policies and their enforcement is at the core of IDS-RAM, supported by legally binding contracts and a certification

approach for both organizations and technical components. This ensures data reusability while retaining data sovereignty.

The International Data Spaces offers a distinct approach to traditional three-corner models by limiting the scope of intermediary access. Only metadata is shared with brokers, providing them with access to transaction frequency without exposing the actual data content. Anonymity from the broker is also supported by IDS connectors by enforcing identity concealment policies, ensuring that only the necessary nodes involved in the data exchange are informed of the relevant identities.

Legal contracts remain fundamental in IDS, as in any business transaction. IDS does not replace these contracts but complements them by adding a layer of technologically enforced agreements. The connection between legally binding contracts and Usage Contracts is part of the IDSA Rulebook. (Steinbuss S. Et al., 2021)

3 Results and discussion

This section presents the findings of a qualitative comparison of data formats and data sharing architectures, focusing on their applicability, strengths, and limitations within the context of Life Cycle Inventory (LCI) data sharing. By evaluating these methods based on security, scalability, data integrity, the analysis highlights key factors influencing their suitability for LCA applications

3.1 Comparison of Data Formats

ILCD and EcoSpold are both XML-based data formats designed for structuring LCA data and processing it throughout all stages of LCA. Both formats are compliant with the ISO 14048 standard, and both are process centric at their core. JSON-LD is mostly supported by OpenLCA and could be an option for small web-based applications.

In ILCD, each process is represented as an individual XML-file, containing references to other elements of LCI, such as input and output flows, through UUID's. These referenced flows are stored in separate XML-files in a designated folder within the ILCD package. Similarly, EcoSpold2 represents processes, referred to as activities, as individual XML-files. These files use UUID's to reference other elements of LCI, which are stored in MasterData lists within the same EcoSpold2 package.

Both formats also support hierarchical relationships between processes, so that a parent-process can contain multiple child-processes. These parent-child relationships are also established by UUID references.

The main difference between ILCD and EcoSpold is that they are used by different LCA databases, ILCD by the European databases and EcoSpold2 by Ecoinvent. They both also contain some number of unique fields, which are not mappable to any other field in the other format. Thus, converting from ILCD to EcoSpold or vice versa is possible, but some data loss may happen. (Turner, I. et al. 2020)

Therefore, the choice of data format in LCA data sharing is less about the inherent advantages and disadvantages of the formats. Instead, it's more about the compatibility with regional databases and software in use.

Format	**Typical use case**	**Software support**	**ISO 14048 Compliant**
ILCD	European databases	Supported by major software	Yes
EcoSpold	Ecoinvent databases	Supported by major software	Yes
JSON-LD	Web-based and lightweight applications	Supported and used under the hood by OpenLCA	Not inherently

Table 1. Comparison of the Data Formats

The results indicate that the choice of data format does not have a substantial impact on the data-sharing process, however, compatibility with local or commonly used databases is still essential. Practitioners should prioritize formats that align with the databases and software they primarily utilize.

Also, ISO-compliant formats are more critical in the later stages of LCA than in the LCI data transfer phase. Given that data is collected from numerous sources, using a custom format designed for the transfer phase may be advantageous. This approach would allow for streamlined data collection without extra overhead in LCI-phase, after which the data can be converted into an ISO-compliant format of choice without data losses to be used for the LCIA-phase.

Another issue that arose was the lack of up-to-date documentation for EcoSpold2. For example, In Ecoinvent's support Knowledge base, it is stated that in EcoSpold2, separate dataset definitions are applied for elementary flow, process, and impact category. Though, in the documentation files only one overarching dataset is found, with a bunch of separate schemas. The shortcomings of available resources and reliable information make the developing and maintaining of a platform challenging, as seen with OpenLCA's discontinuation of support for EcoSpold2 (OpenLCA, 2020).

3.2 Comparison of Data Architectures

In this section, we compare the three prominent architectures which were identified and researched in section 2.2, Data portal with trusted operator, Blockchain, and International Data Spaces (IDS), according to relevant attributes and challenges to LCI data sharing, identified in section 2.2. A table containing summarized results can be found in Section 4, Table 2.

3.2.1 Efficiency

Data portal with trusted operator: With a centralized data access through a single operator, this approach should be highly efficient due to low communication overhead. The efficiency, however, relies heavily on the ability of the middleman to adapt and scale with the changes in the network.

Blockchain: Throughput of a blockchain network can be limited by many factors. Because multiple nodes must validate every transaction, choosing a wrong consensus protocol can decrease throughput as the network scales. According to Prinz, W. et al. (2022), blockchain applications are technically inferior, in terms of performance and scalability, to centralized systems with established technologies.

In 2019, Kuzlu, M. et al. conducted performance testing on Hyperledger Fabric using an AWS EC2 instance. They found that Fabric could sustain a network between two organizations with 100,000 participants at a transaction rate of 200 transactions per second (TPS) with a latency below one second. Around the same time, Christopher Ferris at IBM (2019) tested a network of 32 organizations and 325 individual channels and managed to reach around 13,000 TPS. This indicates that with a robust network infrastructure and hardware, fabric is capable of enterprise-grade efficiency.

International Data Spaces has many possible configurations and the performance-related overhead depends on the specific data space implementation. Therefore, there are no straightforward metrics to assess IDS efficiency. However, since data is exchanged directly between participants, the impact of central bottlenecks is minimal. The primary sources of friction arise from the complexities of data governance, including legal contracts, usage policies, and the negotiation processes that govern data sharing.

3.2.2 Data Provenance

Only storing provenance data is insufficient, as the data must also be available to answer questions, such as 'who created a specific data product and when?' and 'who modified it and when?'. For these needs, querying capabilities are essential for parsing the logs and making provenance data accessible and actionable. Implementing a robust provenance system is time consuming and can be costly.

Blockchain: Blockchain inherently embeds data provenance by recording every transaction on its ledger, which is immutable and transparent. Due to its robust provenance capabilities, blockchain has been explored and proposed as an enabler of provenance for frameworks like International Data Spaces (Prinz, W. et al. 2022), as well as independent cloud platforms (Liang, X., et al., 2017) (Tosh, D., et al., 2019).

In Hyperledger Fabric, the ledger consists of two parts, a world state and a blockchain (Hyperledger Fabric, 2024). The world state represents the current state of the database, making it easier for programs to directly access the current values. The blockchain is the transaction log that records every change that resulted in the current world state. Fabric offers an API for provenance queries, so programs can access the historic information of any given data record.

International Data Spaces: The component that is responsible for data provenance within IDS is the Clearing House. IDS Connectors with integrated local tracking components send the provenance data of each transaction to the Clearing House, which stores it centrally. The role of a Clearing House can be assumed by the same organization as other intermediary roles, because they all require a trusted entity between the data supplier and data consumer. (IDSA, 2022)

3.2.3 Immutability and Integrity

Data Portal with Trusted Operator: Immutability and integrity are reliant on the operator's security practices. With a self-hosted platform, all data can be validated before it is committed to the database, ensuring data quality. While the data is not technically immutable, robust logging can ensure integrity by providing change history for verification.

Blockchain: Immutable by design, blockchain ensures that once data is recorded, it can not be altered without consensus. Therefore, with proper data validation, data integrity is granted within a blockchain. With Hyperledger Fabric, data validation can be performed at the time of commit, ensuring that the transaction satisfies the custom endorsement policies (Hyperledger Fabric, 2024).

International Data Spaces: Instead of offering strict immutability, IDS allows controlled updates and enforces data integrity through comprehensive contract negotiation and usage control processes. Every data exchange starts with a contract negotiation, which is an automated or semi-automated process of specifying conditions for the data usage. If either side rejects the contract, the data exchange process is eliminated. The contracts are validated and recorded by their syntax and content, as well as the participants' digital signatures by the Clearing House. Once they are validated, the contract agreements are stored in both the data providers and the data consumers connectors, leaving an accessible audit trail. (IDSA, 2022)

3.2.4 Scalability

Data Portal with Trusted Operator: Scaling within a three-corner system is not a major issue. Adding new organizations to the network only requires one connection from the new participant to the central operator. As the amount of data increases, the operator may have to invest more hardware resources, which is easy and not too costly. All of the software providers mentioned in section 2.2.1 offer consumption-based pricing models. Any overhead that may result from actions such as logging or data validation, technically should not be as major as the computational overhead within a blockchain.

Blockchain: Scalability is one of the main weaknesses of blockchain (Swathi, P. & Venkatesan, M., 2021). To maintain data integrity and transparency, every node in the blockchain network must hold a complete or partial copy of the entire ledger. As the number of transactions increases, this results in considerable data storage demands. Consequently, participants must continually invest more into their hardware resources. To prevent the size of the ledger from reaching unmaintainable levels, a common solution is to use off-chain storage for larger datasets (Victor O., 2024). Off-chain solutions are, however, a trade-off between scalability and data transparency and governance.

In 2017, Dinh et al. conducted that Hyperledger Fabric is not well suited for large scale data processing workloads. However, since then several techniques have been introduced to address these issues, such as pipelining and parallelism in 2018. Thakkar and Natarajan (2021) implemented an approach, SmartFabric, that utilizes pipelined execution and so-called sparse peers and were able to achieve two to three times higher TPS than vanilla Fabric.

Within a permissioned blockchain, there are several techniques that enterprises can employ. Horizontal scaling means dividing the computational load evenly across the network and can be done by adding more peers to the network, by load balancing, or by partitioning and sharding. Vertical scaling focuses on optimizing the capabilities of individual nodes, such as increasing CPU or memory capacity of participants. Selecting a better suited consensus mechanism is also considered vertical scaling, for example Practical Byzantine Fault Tolerance (PBFT), Kafka, or Raft, which are more suitable for large networks. All of these techniques can make for faster transaction processing times and decreased latency. (Thakran, D., 2023)

International Data Spaces: Within IDS the data remains on the premises of the data owner rather than being stored centrally. This decentralized architecture allows for a more flexible scaling of storage compared to a network with a central operator, because every participant is responsible for storing their own data. Adding new organizations to the network is also similarly straightforward. The growth of the data space should not introduce any extra complexity to the data exchange, as it happens directly between two connectors.

3.2.5 Security

Data Portal with a Trusted Operator: When a single entity controls all data exchange and storage, it is both a weakness and a strength for cyber security. The central data portal makes for a single point of access and failure for cyber criminals, which can be heavily protected with established security measures. The operator can implement standardized encryption, firewall protections and role-based access controls across the portal. The centralized system can help mitigate the risk of inconsistent security practises across the participants by assigning roles and permissions to each individual user based on their responsibilities, and enforcing frequent authentication through a standardized method, such as OAuth 2.0 or multi-factor authentication (MFA).

However, in the case of a successful cyber attack or data breach, most, if not all of the data is compromised. This risk makes it imperative for the operator to implement robust cybersecurity measures and keep them up to date.

Blockchain: While blockchain provides full transparency and traceability of all actions, the decentralized nature means that the entire ledger is accessible for each participant. If an attacker compromises a single node, they will gain access to everything that is stored on the ledger. The channels of Hyperledger Fabric are somewhat of a solution to this vulnerability, since a channel has its own separate ledger which is only visible to a certain subset of nodes. This partition of data reduces the effect of a single compromised node.

Another solution would be to utilize off-chain storage, and only store the metadata and references of the most sensitive data on the ledger. Therefore, an attacker would only gain access to the pointers and not the sensitive data itself. This does however complicate the entire network and take away some of the inherent advantages of blockchain.

International Data Spaces: IDS offers multi-layered security framework, which starts with the onboarding process. To join a data space, every participant is verified and authenticated and must acquire a certified IDS Connector. This ensures that only authorized entities gain access to the network.

After the onboarding, all data exchanges happen in two phases: the Control Phase and the Transfer Phase (IDSA, 2022). Before any data is exchanged, contract negotiation takes place during the Control Phase, and strict usage policies for the data are established. These policies may specify security levels for authorized users, the intended purpose of data usage, limits on the number of times the data can be accessed, or the allowed duration of data use (IDSA, 2022). These policies are enforced contractually and technically through smart contracts. Only after these negotiations are through and policies are accepted, does the data exchange happen directly between the two parties, using an IDS-specific communication protocol. This decentralized and direct data exchange between two nodes circumvents the single point of vulnerability, reducing risks of network-wide data breaches.

3.3 Description of the data portal

The portal was developed using Typescript for the frontend, Spring Boot for the backend and MongoDB as the database. Keycloak is used to handle user authentication and role-based access control. Everything is deployed as Docker containers, ensuring that they can be deployed across different environments without compatibility issues.

The primary function of the portal is to facilitate requesting and sending data between companies. Any company within the network can send a data request to any other company within the network.

The data request is answered by the target company by filling a form. Multiple users within the company can participate in filling out the form, dividing sections to those with proper knowledge.

Keycloak offers a straightforward solution for implementing authentication and role-based access control. Keycloak groups are used to represent companies, and possible roles include modifyIn, modifyOut, readIn, readOut and groupAdmin. Users can belong to multiple companies and have multiple roles at once.

- ModifyIn allows users to modify incoming data requests, e.g. fill out the forms.
- ModifyOut allows users to modify outgoing data requests, e.g. create the form templates, send the requests or delete data requests that have been sent.
- ReadIn allows users to view their company’s incoming data requests, but not answer them.

- ReadOut allows users to view their company's outgoing data requests, but not send them or define the templates. They can view the received data and download it.

Every time a request is made to the backend, the Keycloak token associated with the user is verified on the server side to ensure that the user is authorized to perform the requested action.

Among other descriptive and metadata fields, a data request contains a target company, a source company and a form template. Source company is the source of the request, therefore they are receiving data from the target company. The template is made by the source company, and it defines the data that is being requested.

3.3.1 Custom data format

The data format used within the data portal is a custom format called FlexibleLCAFormData. It was developed out of need for top-to-bottom customizable form templates.

The initial iterations of the portal used a basic format, where each field represented a single fixed value. This was hard-coded to replicate the questionnaire that was used for collecting data from pulp suppliers. While this hard coded approach would have been sufficient if the data requirements remained static and identical for all suppliers, it quickly became limiting as the need for flexibility and customization arose.

To match the categorized nature of the original questionnaire, the following hierarchical model, FlexibleLCAFormData, was developed. At the top level, FlexibleLCAFormData contains metadata, a label and a list of FlexFormSheets. The metadata is for data provenance to keep track of who has made changes and when.

Sheets represent pages of the form, and act as a top layer of the hierarchy, covering a major section or a topic within the data. They have a label and a list of FlexFormCategories.

Categories are subsections within sheets that group related fields, providing structure and organization. They can contain lists of fields and subcategories, allowing for recursive nesting and flexible use. A category can represent a group of records, such as all fossil fuels consumed, or a single record which requires multiple fields for complete definition, such as a transportation that needs details on distance, mode of transportation and load rate of the vehicle.

At the lowest level, there is the FormField class, which represents individual data points. A FormField can store various types of information, such as text, numbers, boolean values or selections from predefined options using dropdowns. It is designed to be highly configurable for most use-cases.

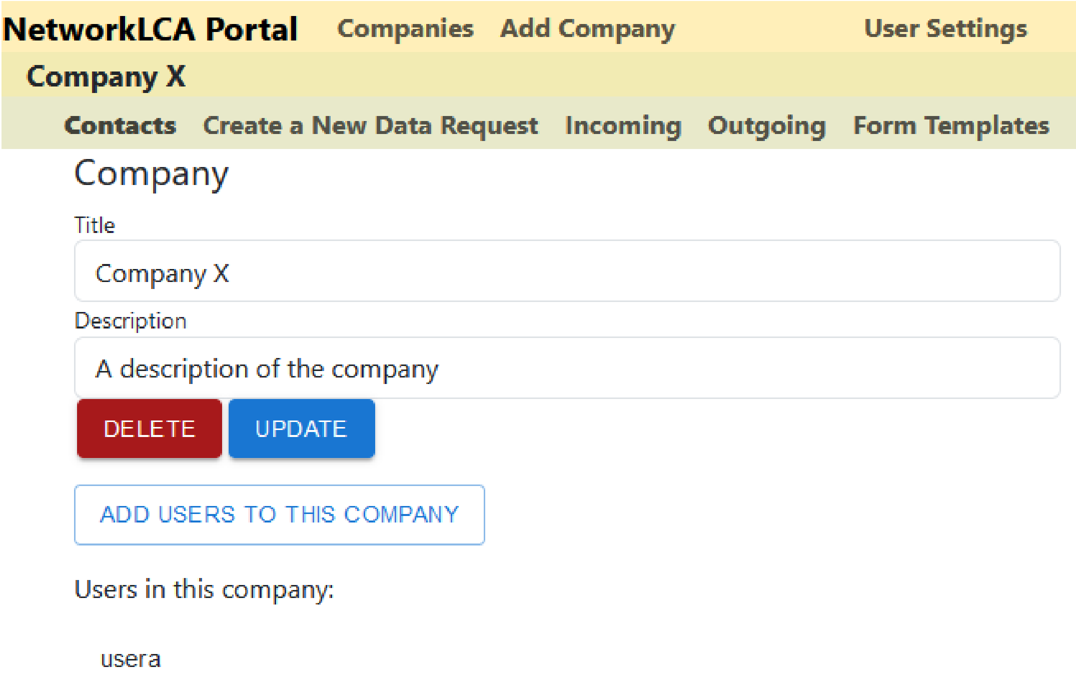
Each field can include optional validation rules, to ensure that the input value remains within a plausible range to maintain data integrity and reduce input errors. A threshold can be set to define an acceptable percentage change from the previous year's value. If it is exceeded, an explanation is required from the user.

```
public class FlexibleLCAFormData {
  private String id;
  private String author;
  private Date createdAt;
  private List<FlexFormSheet> sheets;
}
public class FlexFormSheet {
  private String label;
  private String databaseId;
  private List<FlexFormCategory> categories;
}
public class FlexFormCategory {
  private String label;
  private String databaseId;
  private List<FormField<?>> fields;
  private List<FlexFormCategory> subCategories;
}
public class FormField<T> {
    private T value;
    private String databaseId;
    private String type;
    private String label;
    private String min;
    private String max;
    private String threshold;
    private String unit;
    private String explanation;
    private List<DropdownOption> options;
}
```

Definition for FlexibleLCAFormData and its subclasses

3.3.2 User interface and workflow

Upon successful authentication through Keycloak, users are presented with a list of the companies they are associated with. By selecting a company, the user is directed to the dashboard, which serves as the central hub for navigating the portal's main functions. The dashboard provides access to key features, such as managing contacts, creating new data requests, and viewing form templates and incoming or outgoing requests.



Company

Title: Company X

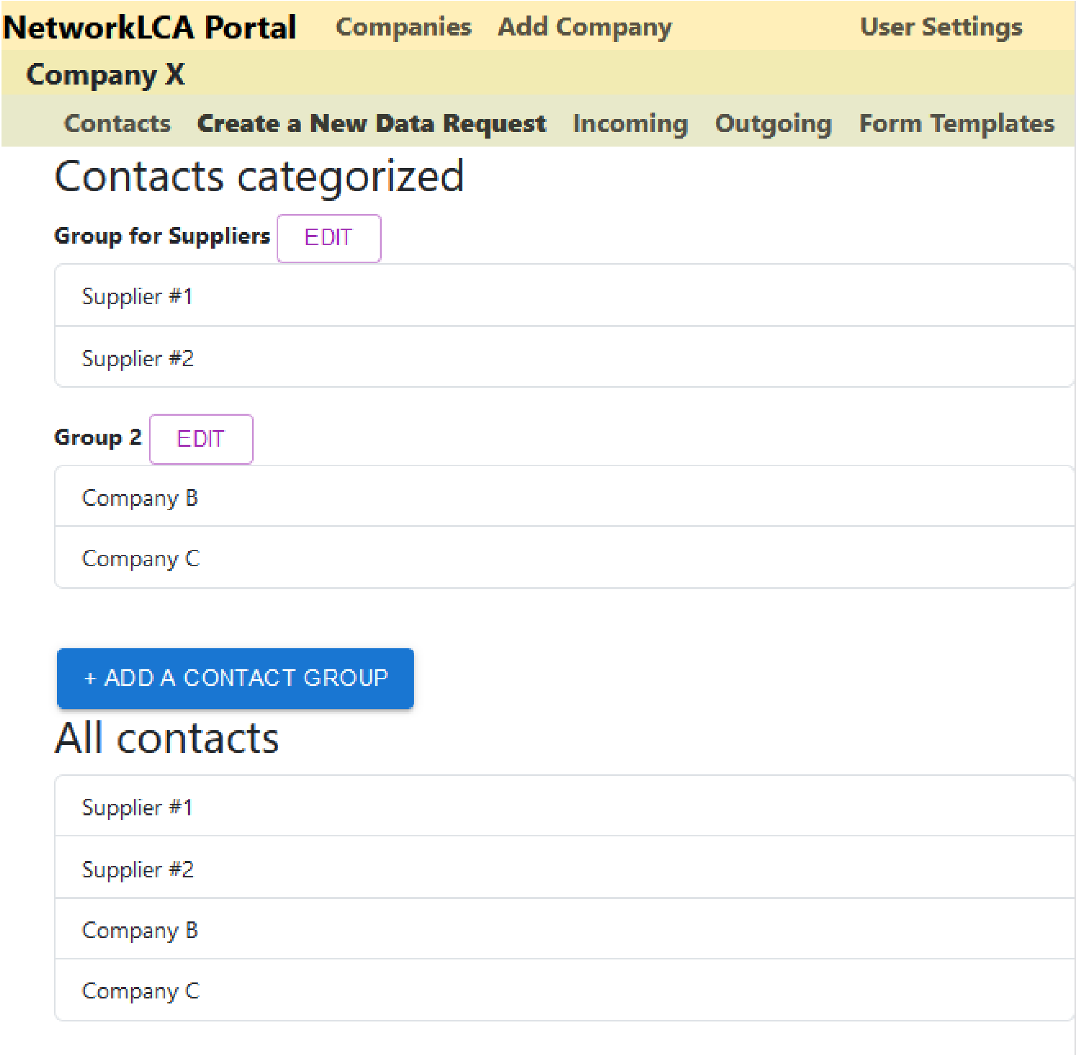
Description: A description of the company

Users in this company:

usera

Figure 8. Dashboard and general outlook of the portal

The Contacts page allows users to view all companies within the network with which their company has had correspondence with at some point in the past. This page also allows users to create contact groups which are used to organize companies and group them in labeled groups for better manageability. When dealing with large and complex supply chains, it is crucial to be able to manage and categorize the individual entities into single structures. Data requests are sent to these groups instead of individual entities, to enhance the efficiency of requesting data.



Contacts categorized

Group for Suppliers

Supplier #1

Supplier #2

Group 2

Company B

Company C

All contacts

Supplier #1

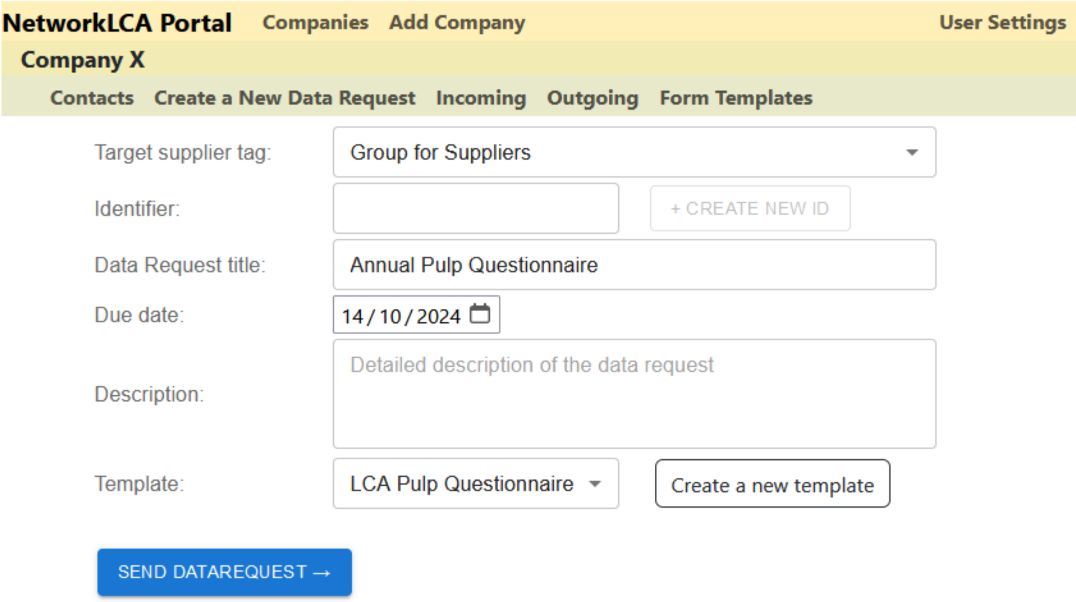
Supplier #2

Company B

Company C

Figure 9. Contacts page

The Create a New Data Request page enables users to send data requests by defining the necessary parameters. Users can select the target contact group to send the request to, define descriptive metadata for the request, such as the title, identifier, due date, and description. They can also choose or create a form template, which defines the structure of the requested data. This page is designed to be simple and to have only the necessary metadata parameters to choose from. This is to enhance both the efficiency and ease of use.



NetworkLCA Portal Companies Add Company User Settings

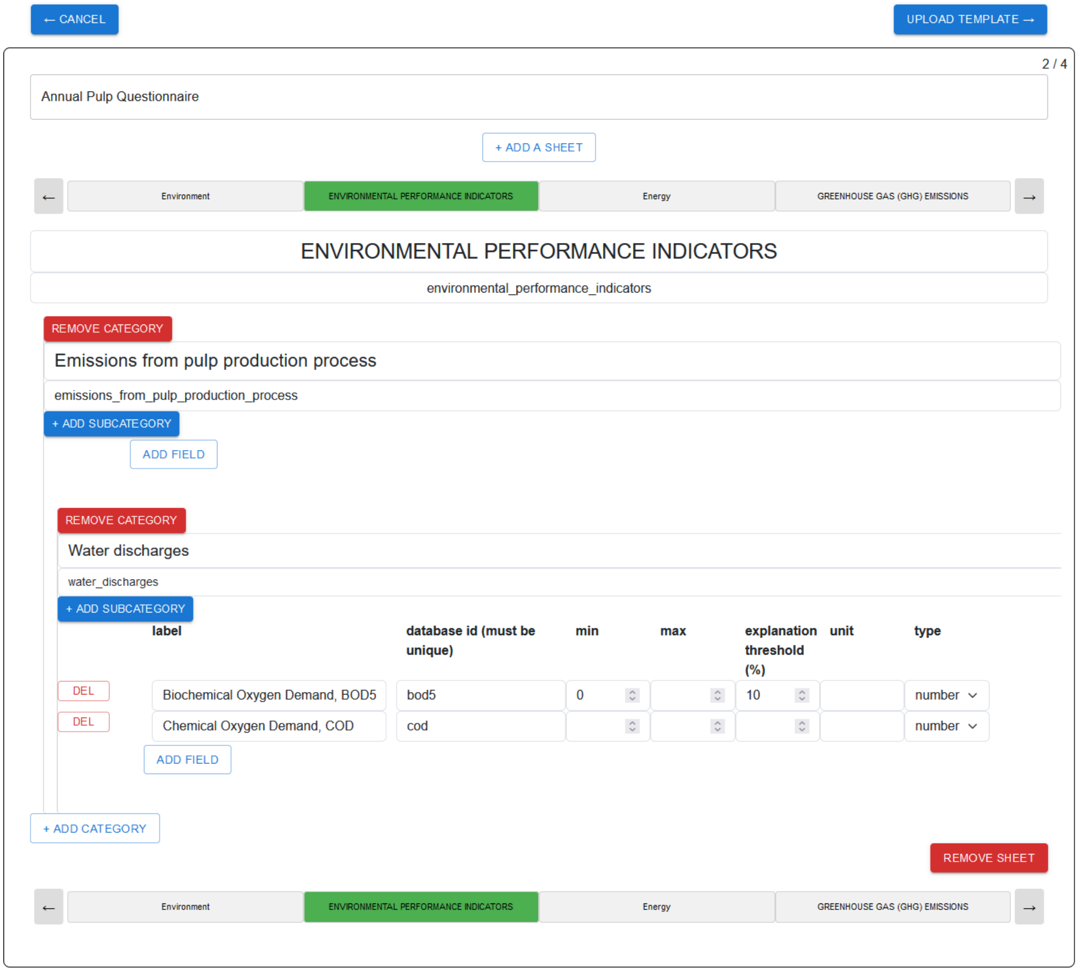
Company X

Contacts Create a New Data Request Incoming Outgoing Form Templates

Target supplier tag:	Group for Suppliers
Identifier:	+ CREATE NEW ID
Data Request title:	Annual Pulp Questionnaire
Due date:	14 / 10 / 2024
Description:	Detailed description of the data request
Template:	LCA Pulp Questionnaire Create a new template

SEND DATAREQUEST →

Figure 10. Creation of a Data Request



← CANCEL UPLOAD TEMPLATE →

2 / 4

Annual Pulp Questionnaire

+ ADD A SHEET

← Environment | ENVIRONMENTAL PERFORMANCE INDICATORS | Energy | GREENHOUSE GAS (GHG) EMISSIONS →

ENVIRONMENTAL PERFORMANCE INDICATORS

environmental_performance_indicators

REMOVE CATEGORY

Emissions from pulp production process

emissions_from_pulp_production_process

+ ADD SUBCATEGORY

ADD FIELD

REMOVE CATEGORY

Water discharges

water_discharges

+ ADD SUBCATEGORY

	label	database id (must be unique)	min	max	explanation threshold (%)	unit	type
DEL	Biochemical Oxygen Demand, BOD5	bod5	0		10		number
DEL	Chemical Oxygen Demand, COD	cod					number

ADD FIELD

+ ADD CATEGORY

REMOVE SHEET

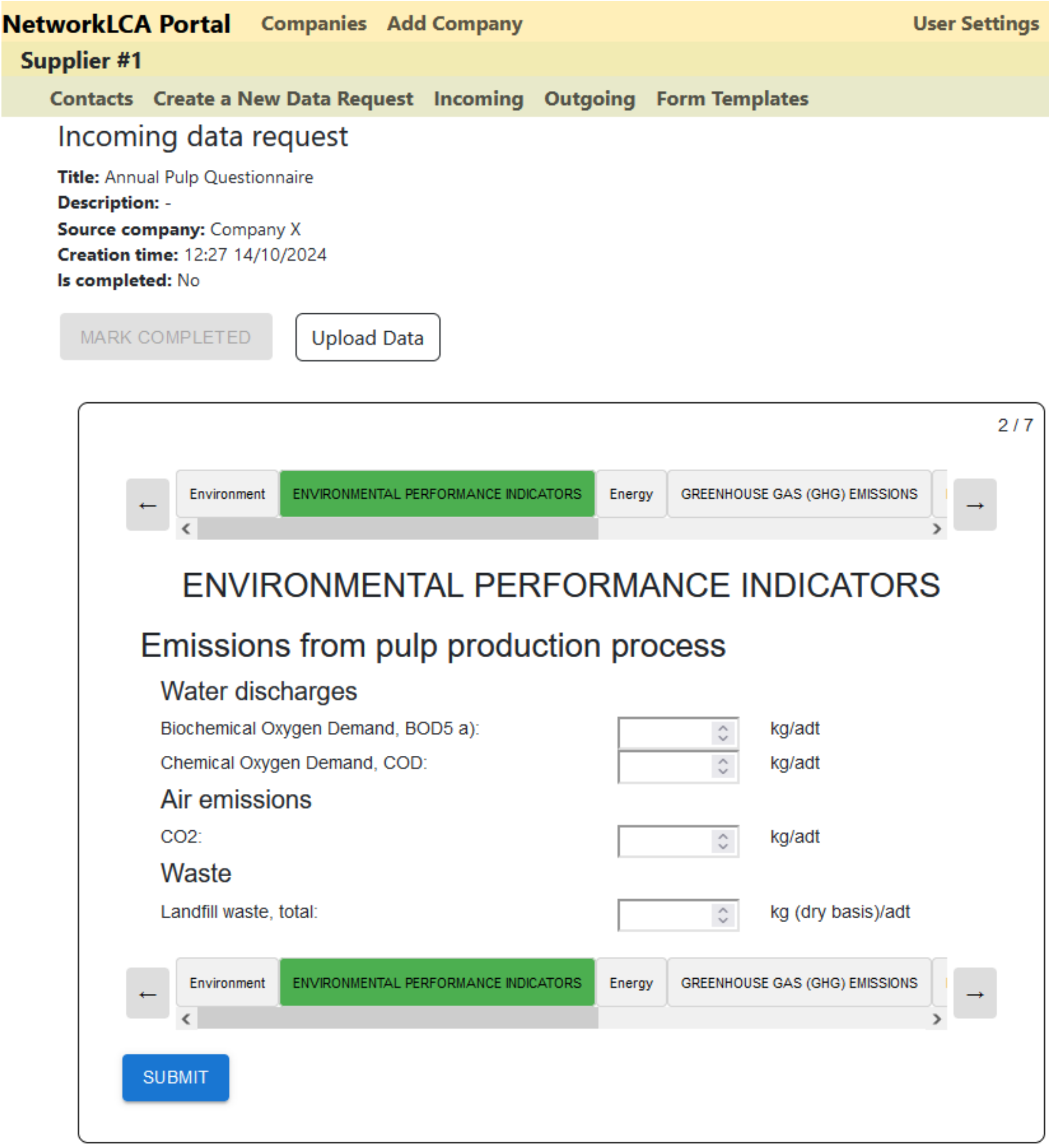
← Environment | ENVIRONMENTAL PERFORMANCE INDICATORS | Energy | GREENHOUSE GAS (GHG) EMISSIONS →

Figure 11. Creation of a Form Template

The created form templates can be viewed as a list on the Form Templates page. From here they can also be edited. Editing a template does not update a template on the backend directly, instead it flags the template as deprecated and creates a new one containing the updated data. This is because the old template might already be used on a data request and updating it could cause some unwanted problems. The deprecated templates are hidden from the list on Form Templates page but are used for the rendering of previously sent data requests' forms.

Incoming and outgoing pages provide lists of respective data requests, enabling users to monitor data flow in both directions. When accessing an outgoing data request users are presented with the descriptive metadata, such as the title, description, target company, creation date, and completion status. If the target company has completed their data request, the data is presented here with an option to export it as a JSON file. JSON was chosen for the ease of use and limited timeframe, in the future iterations of the portal there should be options to export to different data formats, such as ILCD and ecoSpold2. For now, the engineers at Stora Enso will have to map the data from the JSON to their desired data format.

When accessing an incoming data request, users are presented with the same metadata, but the source company instead of the target. Here users can submit the requested data by filling out a web form. After a submission is done, the data request can be marked as complete, so that the data is show to the source company.



Incoming data request

Title: Annual Pulp Questionnaire
Description: -
Source company: Company X
Creation time: 12:27 14/10/2024
Is completed: No

MARK COMPLETED
Upload Data
2 / 7
←
Environment
ENVIRONMENTAL PERFORMANCE INDICATORS
Energy
GREENHOUSE GAS (GHG) EMISSIONS
→

ENVIRONMENTAL PERFORMANCE INDICATORS

Emissions from pulp production process

Water discharges

Biochemical Oxygen Demand, BOD5 a):		kg/adt
Chemical Oxygen Demand, COD:		kg/adt

Air emissions

CO2:		kg/adt

Waste

Landfill waste, total:		kg (dry basis)/adt

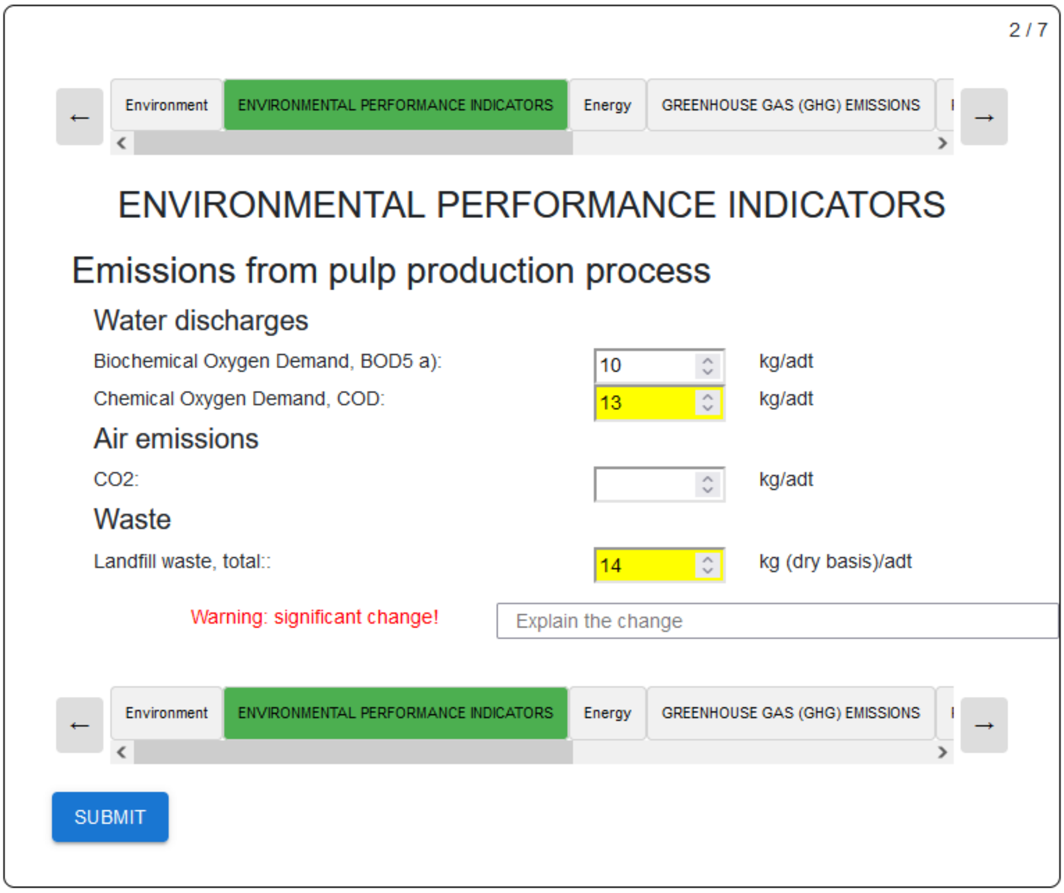
←
Environment
ENVIRONMENTAL PERFORMANCE INDICATORS
Energy
GREENHOUSE GAS (GHG) EMISSIONS
→
SUBMIT

Figure 12. Incoming Data Request

A single user does not have to fill in all the information, they can just submit a few fields without marking the request complete. This saves the incomplete data to the portal but does not reveal it to the source company. As many users as is needed can collectively fill in the form until it is deemed ready to be marked complete. Two users can not override each others' inputs without knowledge, because the server checks for overlapping changes and returns a conflict error if another user has updated the data between opening and submission of the form.

The data form itself is structured as a multi-step process, with categorized pages containing various input fields, such as text, numbers, checkboxes, or dropdowns. Each input field is subject to unique validation rules, defined during the creation of the template, which ensures data integrity. For example, number fields may have minimum and maximum values that are enforced during data entry, with users receiving prompts when their inputs fall outside acceptable ranges.

The data questionnaires are, in Stora Enso‘s case, sent yearly. When an incoming data request is opened, the portal automatically populates it with data from the previous year's questionnaire to save time from the operator, because often most of the values stay the same. When a user modifies prepopulated values, the altered input fields are highlighted yellow, to clearly indicate all changes. This visual cue alleviates error-checking by allowing users to quickly identify and review modifications. The system, however, does not track specific dates for prepopulating. Instead, it uses the most recent data request between two companies. Future iterations of the portal may incorporate features that allow users to manage multiple concurrent data request chains, with options for prepopulating based on specific criteria.



2 / 7
←
Environment
ENVIRONMENTAL PERFORMANCE INDICATORS
Energy
GREENHOUSE GAS (GHG) EMISSIONS
→

ENVIRONMENTAL PERFORMANCE INDICATORS

Emissions from pulp production process

Water discharges

Biochemical Oxygen Demand, BOD5 a): 10 kg/adt

Chemical Oxygen Demand, COD: 13 kg/adt

Air emissions

CO2: kg/adt

Waste

Landfill waste, total:: 14 kg (dry basis)/adt

Warning: significant change! Explain the change

←
Environment
ENVIRONMENTAL PERFORMANCE INDICATORS
Energy
GREENHOUSE GAS (GHG) EMISSIONS
→
SUBMIT

Figure 13. Examples of the highlighted changed values and the prompt to explain a significant change

To ensure accountability for significant data changes, the portal supports the use of thresholds for specific fields. A threshold is a percentage value that defines a significant change from the previous year's data. If a field's value changes by more than the predefined threshold, an explanation is required from the user. For instance, if the previous year's value was 10 and the threshold is set at 20%, any value outside the range of 8 to 12 would require an explanation. These explanations are linked directly to the data field and are visible to the requesting company when reviewing the submitted data.

Moreover, the system handles changes to templates between requests intelligently. Only fields that retain the same hierarchical structure—defined by the sheet → category → subcategory → field path from the previous year's template—are prepopulated. Each hierarchical element is assigned a unique database ID, which ensures consistency in prepopulating while allowing for changes in the labels. For example, if a field remains the same but is moved to a different category with a new database ID, it will not be prepopulated, ensuring that only valid data is reused. Fields that did not exist in the previous template are left empty to avoid the potential for incorrect data associations.

3.4 Validation of the data portal

To minimize the workload for suppliers, who may not have access to complex formats such as ILCD or EcoSpold2, a form-based data collection approach was chosen. This simplifies the process for suppliers by avoiding the need to learn and implement complicated data formats. Additionally, when data is only available in a specific format like EcoSpold2 or ILCD, conversion between formats can lead to data loss (GreenDelta, 2024).

By collecting data in a lightweight, custom format, lossless exporting to any desired data format is possible in the future. Plain JSON was chosen instead of JSON-LD, because the added complexity of JSON-LD's syntax is not necessary for the data transfer phase. LCA data contains many links and references, such as from processes to flows, flows to units, and processes to child-processes. None of these links are present in the data collected from suppliers.

Several challenges were identified earlier in the study and addressed throughout the design of the data portal:

- Trust: The portal follows a hub-and-spoke model, where the data is directly transferred between suppliers and the company using the data, without involving a third party. Since the portal is managed by the company that utilizes the data, suppliers can be confident that their data will not end up in the wrong hands, thereby reinforcing the trust already established between the parties.

- Data Sovereignty: One limitation of the portal is that it does not inherently provide strong data sovereignty for suppliers, as they lose control over their data once it is submitted and cannot remove it from the system. However, all data sharing between companies is request-based, ensuring that all data transactions are intentional and consensual.

- Efficiency. The largest work effort is at the beginning, when the manufacturer must construct the form template for data collection. Once this is set up, the same template can be reused infinitely. Because data requests are sent to individual suppliers once a year, this is ideal for our case. For the supplier, the data entry is straightforward, requiring only that they fill out the form and submit it. For subsequent data entries, the previous values are presented by default, so only the values that changed must be entered again. No stress testing has been done for the portal, so the performance with numerous suppliers remains to be seen.

- Security. The portal's authentication and authorization processes are managed by Keycloak, an established software solution developed by Red Hat, which is based on the OAuth2 protocol. Every request to the backend is validated individually by checking the user's permissions.

- Data Immutability and Integrity. Once data is submitted to the portal, it cannot be removed by the supplier. If updates are made, the original dataset remains stored in the database, although the user interface only displays the most recent version. The metadata associated with each dataset includes the creation date and the user's Keycloak ID, ensuring a clear record of data submissions and updates.

The technologies used, Typescript, Spring Boot and MongoDB, are widely adopted, free, and open source. These are well established within the community, ensuring that future maintenance is as easy as possible.

Several aspects of the portal are still lacking due to the limited time spent on development. For example, a proper logging system should be implemented to enhance data provenance, integrity and transparency. Currently, every data submission is saved as a separate record in the database. By having dedicated table for transaction logs, a clear separation between the primary and auxiliary data could be established. This separation would make it easier to manage and query provenance information.

Exporting to a format that can be natively imported to LCA software, such as ILCD or EcoSpold, would make the utilization of received data more straightforward. Another option would be to directly integrate the portal with the software that is being used at Stora Enso, so the two would work seamlessly together.

4 Summary

This thesis examined the challenges and solutions of LCA data sharing, focusing on aspects such as interoperability, data provenance and security. Qualitative comparison was conducted on three of the most prominent data formats, ILCD, EcoSpold and JSON-LD, as well as three key data-sharing architectures, Blockchain, International Data Spaces (IDS), and a trusted operator system. While the choice of data format is less impactful, compatibility with regional databases and software used is essential. The data sharing architectures were analyzed based on their efficiency, data provenance, immutability, scalability and security. The results showed that IDS offers the best balance of scalability and flexibility for complex and large consortiums.

A proof-of-concept data portal was developed as a part of this study to assess clearer the challenges during development and maintenance of a data portal. The portal offers a practical implementation of the trusted party system, with flexible request-based data transferring.

	Trusted operator	Blockchain	IDS
Efficiency	Good. Efficiency depends completely on the central operator. Minimal processing overhead.	Moderate. Least effective and resource intensive but can be improved with robust implementation and hardware.	Excellent. Direct data exchange between participants without central bottleneck.
Data Provenance	Moderate. Relies on operator's logging and querying capabilities.	Excellent. Inherent provenance via immutable ledger.	Strong. Relies on the implementation of the Clearing House, additional logging by the IDS Connectors.
Immutability and Integrity	Moderate. Relies on the operator's practices, immutability not guaranteed.	Excellent. Inherent immutability due to cryptographic and decentralized design.	Moderate. Ensures integrity through contracts and usage policies rather than immutability.
Scalability	Moderate. Bottlenecked by the central operator's capacity.	Limited. All nodes store the entire ledger, thus requires off-chain data storage and/or complex vertical and horizontal scaling solutions to maintain data integrity.	Excellent. Decentralized data storage and direct node-to-node data transactions allow for flexible scaling.
Security	Moderate. Single point of access and failure, which can be heavily protected but compromises all data if breached.	Moderate. Cryptographical security but risks arise if a single node is compromised.	Strong. Multi-layered security, no single point of failure. Every participant implements data security individually.

Table 2. Summary of the Comparison of Data Architectures

References

Zhang, A., Zhong, R.Y., Farooque, M., Kang, K., Venkatesh, V. G., (2020) Blockchain-based life cycle assessment: An implementation framework and system architecture, Resources, Conservation and Recycling, Volume 152, https://doi.org/10.1016/j.resconrec.2019.104512

Barahmand, Z. and Eikeland, M.S., (2022). Life Cycle Assessment under Uncertainty: A Scoping Review. World, 3(3), pp.692-717.
https://doi.org/10.3390/world3030039

British Plastics Federation (2024). Life Cycle Analysis (LCA) - A Complete Guide to LCAs [Accessed 18 September 2024]. Available: https://www.bpf.co.uk/sustainable_manufacturing/life-cycle-analysis-lca.aspx

Cellura, M., Longo, S. and Mistretta, M. (2011) 'Sensitivity analysis to quantify uncertainty in Life Cycle Assessment: The case study of an Italian tile', Renewable and Sustainable Energy Reviews, 15(9), pp. 4697-4705. Available: https://doi.org/10.1016/j.rser.2011.07.082.

Dinh, T. T. A., Wang, J., Chen, G., Liu, R., Ooi, B. C., & Tan, K.-L. (2017). BLOCKBENCH. A Framework for Analyzing Private Blockchains.
http://dx.doi.org/10.1145/3035918.3064033

Ecoinvent, (2023). Changes ecoSpold1 to ecoSpold2 [Accessed 23 October 2024]. Available: https://support.ecoinvent.org/changes-ecospold1-to-ecospold2

Ecochain (2024). Life Cycle Assessment (LCA) – Everything you need to know [Accessed 18 September 2024]. Available: https://ecochain.com/blog/life-cycle-assessment-lca-guide/#product-life-cycle-lca

European Commission - Joint Research Centre - Institute for Environment and Sustainability (2010) International Reference Life Cycle Data System (ILCD) Handbook - General guide for Life Cycle Assessment - Detailed guidance. 1st ed. EUR 24708 EN. Luxembourg: Publications Office of the European Union

Food and Agriculture Organization of the United Nations (FAO) (2018) Global Forest Products: Facts and Figures 2018. Rome: Food and Agriculture Organization of the United Nations. Available:
https://openknowledge.fao.org/server/api/core/bitstreams/80da7381-fc20-44bb-b0a0-7e8c35334a80/content

Furszyfer Del Rio, D.D., Sovacool, B.K., Griffiths, S., Bazilian, M., Kim, J., Foley, A.M., and Rooney, D. (2022) 'Decarbonizing the pulp and paper industry: A critical and systematic review of sociotechnical developments and policy options', Renewable and Sustainable Energy Reviews, 167, p. 112706. Available:
https://doi.org/10.1016/j.rser.2022.112706

GreenDelta GmbH. (2015). JSON-LD: A smarter format for LCA data interchange. [Accessed 22 October 2024] Available: https://www.greendelta.com/wp-content/uploads/2017/03/LCA_XV_JSON-LD_final.pdf

GreenDelta. (2024), openLCA schema. [Accessed 21 October 2024]. Available: https://greendelta.github.io/olca-schema/intro.html

Hackius, N. and Petersen, M. (2017). Blockchain in logistics and supply chain: Trick or treat? Proceedings of the Hamburg International Conference of Logistics (HICL), Vol. 23. Berlin: epubli GmbH, pp. 3-18.
https://doi.org/10.15480/882.1444

Hedemann, J. And Meinshausen, I., (2008). Ecoinvent 2000 - Documentation EcoSpold. Available: https://support.ecoinvent.org/data-formats-overview-2

Helo, P., Hao, Y., (2019). Blockchains in operations and supply chains: a model and reference implementation, Computers & Industrial Engineering, https://doi.org/10.1016/j.cie.2019.07.023

Hermundsdottir, F. and Aspelund, A. (2021) 'Sustainability innovations and firm competitiveness: A review', Journal of Cleaner Production, 280, p. 124715. Available: https://doi.org/10.1016/j.jclepro.2020.124715

Hyperledger Fabric, (2024). Hyperledger Fabric Docs - Introduction. Available: https://hyperledger-fabric.readthedocs.io/en/latest/whatis.html

IBM, (2024). Does Hyperledger Fabric Perform at Scale? IBM Insights. Available at: https://www.ibm.com/think/insights/does-hyperledger-fabric-perform-at-scale [Accessed 11 November 2024].

International Data Spaces Association (IDSA), (2024). Our mission and vision. [Accessed 29 October 2024] Available: https://internationaldataspaces.org/why/

International Data Spaces Association (IDSA), (2022). IDS Reference Architecture Model Version 4.1. [Accessed 31 October 2024] Available: https://github.com/International-Data-Spaces-Association/IDS-RAM_4_0/tree/main/documentation

ISO (2006a). ISO 14040: Environmental management – Life cycle assessment – Principles and framework. International Organization for Standardization.

ISO (2006b). ISO 14044: Environmental management – Life cycle assessment – Requirements and guidelines. International Organization for Standardization.

Joint Research Centre, Institute for Environment and Sustainability, (2010) General guide for Life Cycle Assessment: provisions and action steps. Publications Office. https://data.europa.eu/doi/10.2788/94987

Jurmu, M.; Niskanen, I.; Kinnula, A.; Kääriäinen, J.; Ylikerälä, M.; Räsänen, P.; Tuikka, T. (2023). Exploring the Role of Federated Data Spaces in Implementing

Twin Transition within Manufacturing Ecosystems. Sensors 2023, 23, 4315. https://doi.org/10.3390/s23094315

Kuzlu, M., Pipattanasomporn, M., Gurses, L. & Rahman, S., (2019). Performance Analysis of a Hyperledger Fabric Blockchain Framework: Throughput, Latency and Scalability. [online] DOI: 10.1109/Blockchain.2019.00003

Lashkari, B. & Musilek, P., (2021). A Comprehensive Review of Blockchain Consensus Mechanisms. Available: https://doi.org/10.1109/ACCESS.2021.3065880 [Accessed 1 September 2024].

Liang, X., Shetty, S., Tosh, D., Kamhoua, C., Kwiat, K. and Njilla, L. (2017) 'ProvChain: A Blockchain-Based Data Provenance Architecture in Cloud Environment with Enhanced Privacy and Availability', 2017 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID), Madrid, Spain, pp. 468-477. doi: 10.1109/CCGRID.2017.8.

Life Cycle Initiative. (2022). LCA Database Helpdesk, Data formats. [Accessed: 14 October 2024]. Available: https://helpdesk.lifecycleinitiative.org/distribution/data-formats/

Lin, X., Li, X., Kulkarni, S. & Zhao, F. (2021) The Application of Blockchain-Based Life Cycle Assessment on an Industrial Supply Chain. Sustainability, 13(23), p. 13332. https://doi.org/10.3390/su132313332

Matthews, H. S., Hendrickson C. T., and Matthews D. H. (2014). Life Cycle Assessment: Quantitative Approaches for Decisions That Matter. p. 83.

Meinshausen, I., Müller-Beilschmidt, P. & Viere, T. (2016). The EcoSpold 2 format—why a new format?. Int J Life Cycle Assess 21, 1231–1235. https://doi.org/10.1007/s11367-014-0789-z

Nagel L., Lycklama D. (2021): Design Principles for Data Spaces. Position Paper. Version 1.0. Berlin DOI: http://doi.org/10.5281/zenodo.5105744

OpenLCA, (2019). 'Importing ecoinvent 3.6 cutoff database to openLCA', *ask.openLCA*. [Accessed: 18 November 2024] Available: https://ask.openlca.org/2561/importing-ecoinvent-3-6-cutoff-database-to-openlca?show=2561#q2561.

Pilkington, Marc, Blockchain Technology: Principles and Applications (2015). Research Handbook on Digital Transformations, edited by F. Xavier Olleros and Majlinda Zhegu. Edward Elgar, 2016, Available: https://ssrn.com/abstract=2662660

Prinz, W., Rose, T. and Urbach, N. (2022) 'Blockchain Technology and International Data Spaces', in Otto, B., ten Hompel, M. and Wrobel, S. (eds.) *Designing Data Spaces*. Cham: Springer. https://doi.org/10.1007/978-3-030-93975-5

PRé Sustainability. (2019). The ILCD format – solving LCA data exchange problems. [Accessed 14 October 2024]. Available: https://pre-sustainability.com/articles/the-ilcd-format-solving-lca-data-exchange-problems/

Schandl, H., Fischer-Kowalski, M., West, J., Giljum, S., Dittrich, M., Eisenmenger, N., Geschke, A., Lieber, M., Wieland, H., Schaffartzik, A., Krausmann, F., Gierlinger, S., Hosking, K., Lenzen, M., Tanikawa, H., Miatto, A. and Fishman, T. (2018), Global Material Flows and Resource Productivity: Forty Years of Evidence. Journal of Industrial Ecology, 22: 827-838. https://doi.org/10.1111/jiec.12626

Steinbuss Sebastian, (2020). IDS and the FAIR DATA PRINCIPLES. [Accessed 29 October 2024] Available: https://internationaldataspaces.org/ids-and-the-fair-data-principles/

Steinbuss S., Ottradovetz K., Langkau J., Punter M. et al. (2021) IDSA Rule Book. International Data Spaces Association. https://doi.org/10.5281/zenodo.5658294

Smart Connected Supplier Network. (2024). How it works [Accessed 17 September 2024]. Available: https://smart-connected.nl/en/about-scsn/how-it-works

Tobias Koch, Derilinx (2024) The Power of Data Sharing – Part 2 [Accessed 18 September 2024] Available: https://derilinx.com/blog-the-power-of-data-sharing-part-2-models/

Tosh, D., Shetty, S., Liang, X., Kamhoua, C. and Njilla, L.L. (2019) 'Data Provenance in the Cloud: A Blockchain-Based Approach', IEEE Consumer Electronics Magazine, 8(4), pp. 38-44. doi: 10.1109/MCE.2019.2892222

Huynh, T.D., Michaelides, D.T. and Moreau, L. (2016). PROV-JSONLD: A JSON and Linked Data Representation for Provenance. In: M. Mattoso and B. Glavic, eds. Provenance and Annotation of Data and Processes, IPAW 2016. Lecture Notes in Computer Science, vol. 9672. Cham: Springer. Available: https://doi.org/10.1007/978-3-319-40593-3_15

Thakkar, P. & Natarajan, S., (2021). Scaling blockchains using pipelined execution and sparse peers. *Proceedings of the ACM Symposium on Cloud Computing (SoCC '21)*. Seattle, WA, USA: Association for Computing Machinery, pp.489–502. Available: https://doi.org/10.1145/3472883.3486975

Thakran, D. (2023) 'Scaling Hyperledger Network: Key Considerations for Effective Scaling', *AST Consulting Blog*, 2 August. [Accessed: 8 November 2024] Available: https://astconsulting.in/blog/2023/08/02/scaling-hyperledger-network/).

The Hyperledger White Paper Working Group, (2018). An Introduction to Hyperledger, Available: https://8112310.fs1.hubspotusercontent-na1.net/hubfs/8112310/Hyperledger/Offers/HL_Whitepaper_IntroductiontoHyperledger.pdf

Turner, I., Smart, A., Adams, E. et al. (2020). Building an ILCD/EcoSPOLD2–compliant data-reporting template with application to Canadian agri-food LCI data. Int J Life Cycle Assess 25, 1402–1417. https://doi.org/10.1007/s11367-020-01748-2

Velte, P. (2022) 'Meta-analyses on Corporate Social Responsibility (CSR): a literature review', Management Review Quarterly, 72(3), pp. 627–675. Available: https://doi.org/10.1007/s11301-021-00211-2

Victor Oshimua, (2024). Off-Chain Data Storage Significance in Blockchain. [Accessed 12 November 2024] Available: https://dev.to/victor_isaac_king/off-chain-data-storage-significance-in-blockchain-cd0

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. and Mons, B., (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3:160018. doi:10.1038/sdata.2016.18.

World Wide Web Consortium (W3C). (No date). JSON for Linking Data. [Accessed 22 October 2024] Available: https://json-ld.org/

World Wide Web Consortium (W3C). (2020). W3C Recommendations, JSON-LD 1.1. [Accessed 21 October 2024]. Available: https://www.w3.org/TR/json-ld11/

www.forbes.com. (No date). Cryptocurrency Prices, Market Cap and Charts. [Accessed 24 September 2024] Available: https://www.forbes.com/digital-assets/crypto-prices/