

Automatic data transformation from specific business information system to DPP

Marija Jankovic^{1*}, Alexandros Nizamis^{1,2}, Dimosthenis Ioannidis¹, Konstantinos Votis¹ and Dimitrios Tzovaras¹

¹ Centre for Research and Technology Hellas, 6th km Charilaou-Thermi Rd, GR 57001, Thessaloniki, Greece

² Department of Digital Industry Technologies, National and Kapodistrian University of Athens, Dirfies Messapies, GR 34400, Greece

Abstract

The lack of interconnection, interoperability, and transparency in data sharing among businesses has significantly reduced the potential for future collaboration across value chains. Recently, the concept of Digital Product Passport (DPP) has emerged as a promising approach to facilitate collaboration between potential partners, both in terms of policy and practical implementation across various industries. In this work, we provide a brief overview of the DPP initiatives specific to the battery sector, along with the current challenges faced. Additionally, we introduce a novel framework for an as-a-service data transformation that will extend DPP architecture. The proposed data transformation service not only supports data interoperability but also integrates the principles of data sovereignty and trust by leveraging the Data Spaces concept.

Keywords

DPP, Interoperability, Data Transformation, Data Spaces, ETL

1. Introduction

In an era of rapid technological development and escalating sustainability concerns, the digital transformation of supply chains has become crucial. The European Union, under the European Green Deal, has launched measures such as the Sustainable Products Initiative and the Ecodesign for Sustainable Products Regulation (ESPR) to strengthen sustainability across industries [1]. The introduction of Digital Product Passports (DPPs) to improve transparency and facilitate a circular economy, complementing the EU's broader strategies for digital transition and data sharing is integral to these measures [2].

The Battery Pass project embodies the ESPR's vision by showcasing the practical application and operationalization of DPPs in the battery sector. The project gained significant support from the European Commission's standardization request (SReq) in May 2023, which seeks to promote the integration and establishment of norms for DPPs across the continent [4]. The collaborative efforts of CEN-CENELEC JTC 24 'Digital Product Passport' and the Battery Pass Consortium are crucial to shaping a coherent System Architecture for the Battery Passport, addressing the technological infrastructure and the creation and adoption of shared standards for these systems [4].

Despite this, the absence of standardized protocols often hinders data sharing between business entities, thereby impeding cross-value-chain collaboration. There is a recognized need for an effective 'Data Transformation Service' that facilitates seamless data conversion from various proprietary business information systems to the DPP format [3]. By doing so, we will ensure that DPP data will

I-ESA 2024 12th International Conference on Interoperability for Enterprise Systems and Applications, April 10–12th, 2024, Crete, Greece
jankovicm@iti.gr (M. Jankovic); alnizami@dind.uoa.gr (A. Nizamis); djoannid@iti.gr (D. Ioannidis); kvotis@iti.gr (K. Votis); dimitrios.tzovaras@iti.gr (D. Tzovaras)
0000-0002-3600-3014 (M. Jankovic); 0000-0002-6954-4861 (A. Nizamis); 0000-0002-5747-2186 (D. Ioannidis); 0000-0001-6381-8326 (K. Votis); 0000-0001-6915-6722 (D. Tzovaras)



© 2025 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

be standardized, actionable, and interoperable across all the DPP ecosystem's components. This paper delves into the complexities of such transformation process, highlighting the key challenges that need to be overcome and proposes conceptual solution to establish a robust and efficient DPP framework by introducing as-a-service data transformation.

Following this introductory chapter, the rest of the paper is organized as follows: Chapter 2 presents a brief overview of DPP background and data transformation. Chapter 3 and 4 present the challenges and the proposed concept for sovereign data transformation. Finally, conclusions and future work are drawn in Chapter 5.

2. Background

The following section outlines the specific background related to our introduced work that consists of two main parts: (i) the DPP system that was initially introduced in battery sector and (ii) the data transformation services from EC funded project RE4DY aimed at delivering 'Data as a Product' [9] value ecosystems for Factory 4.0.

Digital Passport System. According to [3] the DPP is a decentralized system that consists of sectorspecific data and an interoperable system architecture. The DPP system plays a crucial role in coordinating various key stakeholders across different geographical locations and industrial sectors. It provides consolidated data for governance purposes and is managed by Economic Operators. The European Commission is responsible for critical infrastructure elements such as the web portal and registry that ensure unified system administration and streamlined access to vital data throughout the product's lifecycle. The *DPP Data* is specific to each sector, such as batteries, electronics, textiles, and construction materials, and is defined by different regulations. A comprehensive list of data attributes that need to be available in the battery passport is included in Article 77 and Annex XIII of the EU Battery Regulation [4]. The Battery Pass initiative proposes a comprehensive approach to data modeling by following the Model Driven Approach (MDA). This approach includes a semantic *Core Data Model* based on RDF, a *Platform Independent Model (PIM)*, and a *Platform Specific Model (PSM)*. The goal of this tiered approach is to establish a flexible data architecture that can promote efficient data exchange and support transparency, traceability, and sustainability within the battery industry. An RDF meta model based on the W3C Resource Description Language (RDF) is defined for the representation of the Battery Passport data points identified in the Battery Pass Content Guidance[4]. The SReq does not cover the domain data ecosystem, which makes it difficult to organize data. However, there are two RDF-based ontologies available, offering a structured approach, particularly in the battery sector [5] [6].

Automatic Data Transformation. The RE4DY project leverages advanced ETL (Extract, Transform, Load) processes to ensure robust and efficient data transformation. At the heart of these processes is the Sovereign Data Transformation Service (DTS), that extends the work that has been done in European Connected Factory Platform for Agile Manufacturing (EFPF) Data Spine [7]. DTS utilizes Apache NiFi, a leading-edge technology in data flow management. Apache NiFi is renowned for its flexibility and scalability, supporting a vast array of file formats including JSON, CSV, and plain text, thus facilitating seamless integration into various data ecosystems. The automatic data transformation in RE4DY is designed to address the three essential stages of ETL:

- **Extraction:** Data is sourced from diverse origins, encompassing a multitude of formats, which are effectively handled by the system.
- **Transformation:** The core transformation process utilizes Apache NiFi's processors to convert data into a standardized format, applying transformation rules, ontological mappings, and JOLT transformations to ensure the data's integrity and compliance with specified requirements.
- **Loading:** The transformed data is subsequently loaded into the designated target systems or databases, completing the ETL cycle.

This streamlined process is enhanced by the use of data space connectors that comply with the IDSA Reference Architecture Model (RAM). The connectors enable secure and sovereign data communication within the digital infrastructure of the RE4DY project, which adheres to the European

Industrial Data Space (EIDS) paradigm [8]. Additionally, the connectors ensure that data management confirms to strict regulations regarding access and sharing, thus reinforcing the system's reliability and trustworthiness.

3. DPP Data Transformation Challenges

Transforming data for efficient integration into DPP presents several challenges across the value chain [3]:

- **Data Collection and Aggregation:** The '*upstream value chain*' involves initial stages such as mining, refining, and producing components essential for batteries. At this stage, collecting static upstream data, which includes parameters like greenhouse gas emissions and certifications, is challenging due to the varying formats and standards across the sources. Ensuring the uniformity of this data is crucial for its integration into the DPP.
- **Data Processing:** The collected data should be transformed into structured DPP format that adheres to standardized DPP schema, ensuring consistency and regulatory compliance across the system.
- **Data Access via Battery Passport and Registry:** Access to DPP data varies among different user groups, from the public to regulatory authorities. The challenge lies in fulfilling the data access needs of these diverse groups while preserving data integrity and security. To achieve this, user interfaces for web portals and registries must be designed to make data retrieval both straightforward and accurate.

Addressing these challenges is crucial for enhancing the functionality of DPP system and promoting a transparent and sustainable data ecosystem.

4. Sovereign DPP Data Transformation

In this section, we describe the extended DPP System Architecture illustrated in Figure 1. The newly introduced Data Transformation Service (DTS) is systematically integrated within the existing DPP framework, providing interoperability and robustness to the data handling lifecycle. Besides, we provide a more detailed data information flow diagram illustrating a possible application of data space connectors for trusted and sovereign communication between the Economic Operator and specific business information systems.

4.1. Proposed extension of the DPP System Architecture

The DPP System Architecture proposed by the Battery Pass consortium is segmented into three principal service-oriented components as illustrated in Figure 1 [4]. Firstly, EC Central Services, which fall under the purview of the European Commission, provide a centralized framework for secure data exchange and access control within the battery passport ecosystem, encompassing APIs and rolespecific portals to facilitate efficient system oversight. The *Passport Data Service* offers a suite of functions for the validation, management, and logging of data, ensuring its integrity. A secure *Registry UID* maintains essential identifiers, enabling regulatory authorities to verify compliance. Additionally, *Support Services* deliver IT operations management to uphold service levels and ensure system performance and maintenance. Next, the **Third-Party Services** component involves external service providers who offer essential services that augment the DPP system's functionality, incorporating additional capabilities not directly covered by EC Central Services or the Economic Operators such as mandatory back up of the DPP data. Finally, the architecture features **Distributed DPP System Services**, that indicate a decentralized approach, where each Economic Operator or Service Provider may have its instance of the service, ensuring the DPP system's robustness and adaptability across various operational contexts. Interoperability within this varied system is achieved through a semantic data model that serves as a common denominator, independent of any specific platform.

In this context, the '**Data Transformation Services (DTSs)**' emerge as a critical component. They should provide a vital bridge, facilitating seamless integration and translation between the myriad of unique data formats of the proprietary systems used by different entities and the unified DPP data model, ensuring consistent and coherent data exchange throughout the network. To address this issue, we propose the introduction of the new DTS component to the Distributed DPP System Services Layer as illustrated in Figure 1:

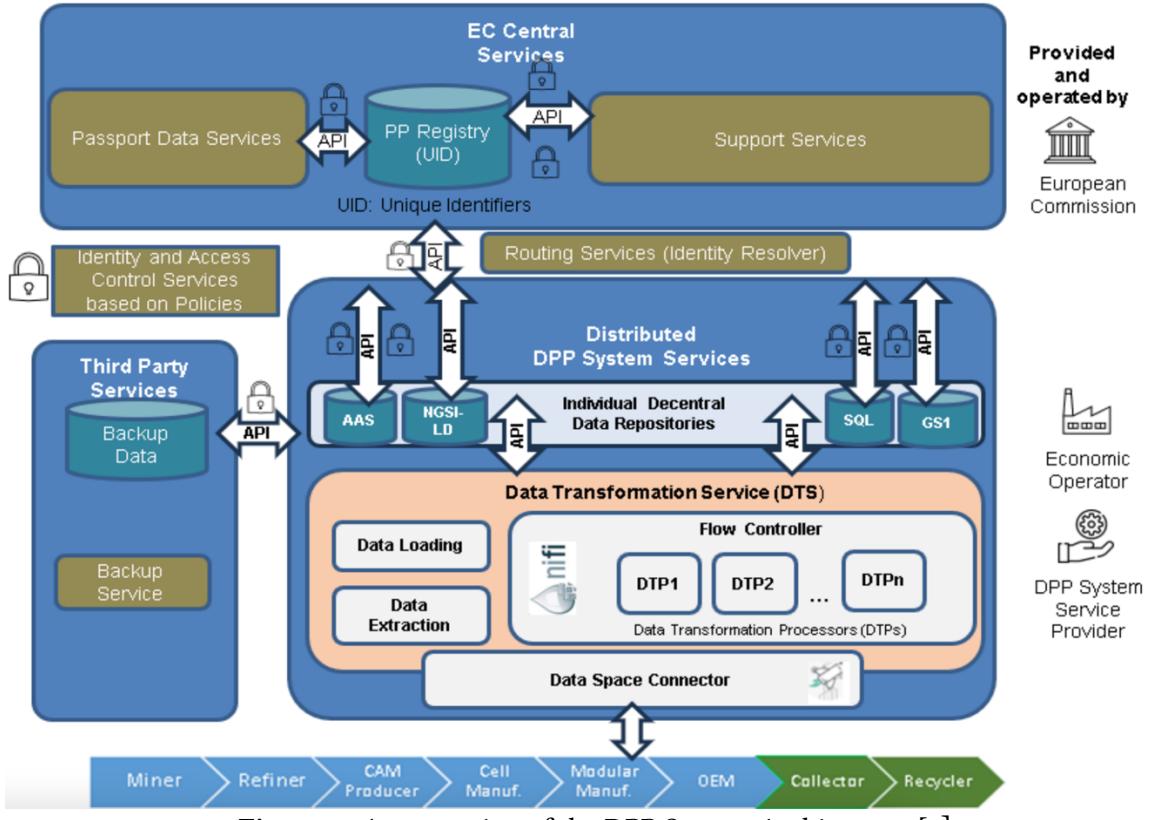


Figure 1: An extension of the DPP System Architecture [3]

The DTS architecture comprises several sub-components that work in tandem to facilitate the ETL process. At the core of the architecture is the **Flow Controller**, which coordinates the activities of the Apache NiFi and DPP-specific custom processors to ensure efficient and synchronized data transformation. The **Data Loading Sub-Component** is the first stage of the ETL process and tailored to meet the unique requirements of the DPP ecosystem. It supports source integration, adaptive data ingestion, and proficient error handling, which is crucial for accommodating diverse data streams and enabling a seamless data flow into the system. The **Data Extraction Sub-Component** is designed to meticulously clean and filter data, removing impurities and inconsistencies and preparing the data for the subsequent transformation phase. This sub-component is essential for ensuring the quality and reliability of the data entering the DPP system. The **Data Transformation Processors** are built upon the robust foundation of Apache NiFi and custom-designed to perform critical functions such as data conversion, data enrichment, data aggregation, business logic application, and data mapping. The use of JOLT processors within this sub-component underscores the capability to handle complex JSON transformations precisely and with agility. **Data Space Connectors** are vital in ensuring trusted and sovereign communication between data providers and consumers. They enable the transmission of transformed data between stakeholders, adhering to the sovereignty and security principles. The IDSA RAM-based Data Space Connectors establish guidelines for data access and sharing, making it easier for stakeholders to share data.

The integration of the IDSA Connectors, Apache NiFi, and JOLT forms a powerful technological nexus for the DTS. Apache NiFi provides the robust infrastructure needed to manage extensive data flows, while JOLT brings specialized flexibility for JSON data manipulation. This combination ensures

that the DTS can accommodate a wide spectrum of file formats, underscoring the adaptability of the services to integrate seamlessly into varied data ecosystems.

4.2. End-to-end Data Transformation Flow in DTS

Figure 2 illustrates a hypothetical scenario that demonstrates how the Economic Operator, Supplier, and DTS interact systematically based on the Apache NiFi platform and enable data interoperability through staged transformation. The data flow scenario progresses through various stages, each playing a critical role in facilitating data interoperability.

In the first stage, the Economic Operator, acting as a data consumer, initiates the process by requesting data that complies with the desired specification, in our example Asset Administration Shell (AAS). This stage is crucial in identifying the specific data requirement and triggers the initial data acquisition process from the last Supplier in the battery value chain, who acts as the Data Provider in this workflow.

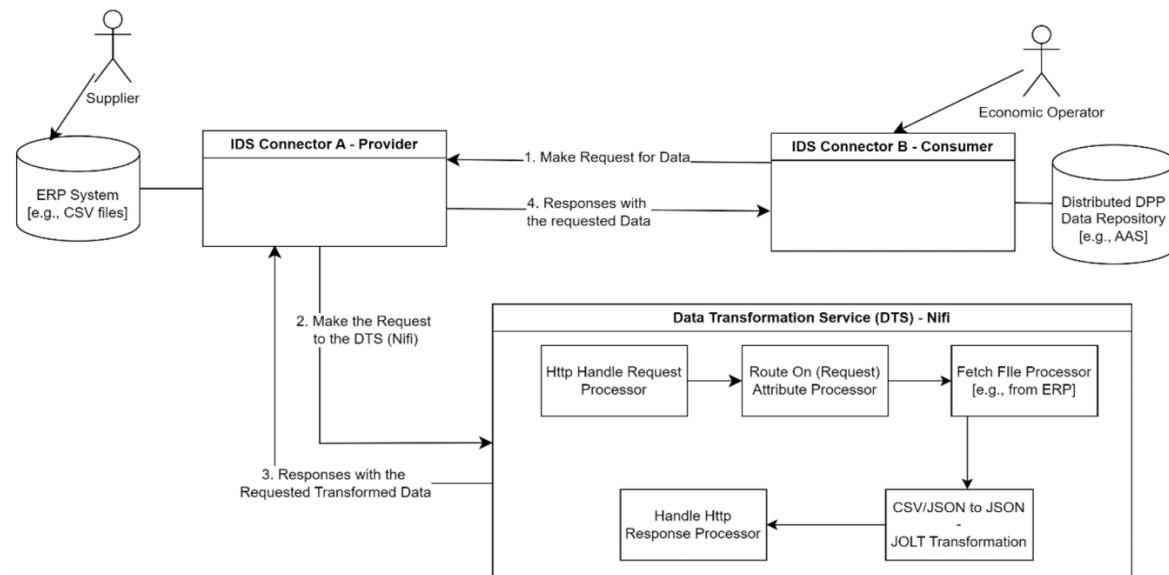


Figure 2: Data Flow between the Data Provider and Consumer

In stage two, when the Supplier receives the data request, he activates the DTS, which is built on the Apache NiFi framework - renowned for its robust data flow management and transformation capabilities. An HTTP handle request processor, integral to the NiFi ecosystem, serves as the flow controller. This processor orchestrates the workflow, mobilizing necessary subordinate processors to execute data transformations that align with the AAS standard, thus converting raw data into an interoperable structure. Raw data can be extracted from a Data Provider's resource, such as an ERP system, using an appropriate processor and then transformed into an intermediate JSON format, unless it is already in that format. Subsequently, JOLT library's capabilities are used to provide JSON to JSON transformation to achieve the target AAS-based format. An HTTP handle response processor is also utilized to return transformed data to Data Provider's Data Space Connector, thereby responding to the Data Consumer's request.

In the final stage, the DTS transforms the data into the requested format, which is crucial not only for responding to the initial request but also for ensuring data interoperability. The data, now compliant with AAS standard, is sent back to the Economic Operator. This stage emphasizes that the transformed data is ready for smooth integration and exchange across various data ecosystems, meeting both functional requirements and interoperability standards.

The presented data sequence outlines the various phases involved in setting up Data Provider and Data Consumer processors, with emphasis on the crucial role played by Apache NiFi in enabling data transformation. The workflow goes beyond a traditional ETL operation, depicting the journey from data sourcing (Stage I) to transformation (Stage II), ultimately leading to the attainment of data

interoperability (Stage III). The sequence showcases the Economic Operator's ability to request data in various standardized formats such as AAS and NGSI-LD that align with the DPP ontology. The DTS is versatile in transforming a wide range of data types, from CSV files to proprietary formats, but this transformation requires specialized Data Transformation Processors that must be engineered and configured precisely to handle the distinct needs of each data format. These processors enable the transformation of data into a structure compliant with the DPP framework. The system serves as a model for a comprehensive data transformation process within a data space ecosystem, where data is not just extracted and transformed but is also rendered in a format that is optimally suited for widespread application and interoperability.

5. Future work and conclusion

In conclusion, the extended DPP System Architecture, supported by the DTS, presents a progressive method for data interoperability. This comprehensive extension offers strong and dependable ETL services and can adapt to the specific needs of different enterprise actors. Future work will focus on further enhancing the interoperability and security of the DPP framework. Additional research is needed to explore new technologies and methodologies for data sharing and access. The integration of blockchain technology for enhanced data security could be a potential area of exploration. Acknowledging the progress in hierarchical security assessment models, there exists a compelling need to explore their applicability and potential integration within the DPP framework, indicating a valuable avenue for future research to improve data architecture and security assessment [10]. This could further strengthen the resilience and trustworthiness of the DPP framework. The adoption of industry standards will also be a focus for future work. This will ensure that the DPP framework remains compatible and interoperable with other systems and platforms. It is important to continue engaging with stakeholders to guide the evolution and refinement of the DPP framework. Moving forward, we are committed to integrating the DTS as an essential component of the DPP architecture, a step that promises to enhance data interoperability and security.

Acknowledgements

This work has received funding from the European Union's Horizon Europe innovation action programme under grant agreement No.101058384 – RE4DY

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

- [1] Ecodesign for Sustainable Products Regulation - European Commission. Accessed: Mar. 07, 2024. [Online]. Available: https://commission.europa.eu/energy-climate-change-environment/standardstools-and-labels/products-labelling-rules-and-requirements/sustainable-products/ecodesignsustainable-products-regulation_en
- [2] Common European Data Spaces | Shaping Europe's digital future. Accessed: Mar. 07, 2024. [Online]. Available: <https://digital-strategy.ec.europa.eu/en/policies/data-spaces>
- [3] Battery Pass Consortium, "Battery Passport Technical Guidance," vol. Version 1.0, Mar. 2024.
- [4] Battery Pass Consortium, "Battery Passport Content Guidance," vol. Version 1.1, Dec. 2023.
- [5] BattINFO - BIG-MAP, <https://www.big-map.eu>. Accessed: Mar. 09, 2024. [Online]. Available: <https://www.big-map.eu/dissemination/battinfo>
- [6] OntoCommons ontology catalogue. Accessed: Mar. 09, 2024. [Online]. Available: [https://data.ontocommons.linkeddata.es/vocabulary/BatteryValueChainOntology\(bvco\)](https://data.ontocommons.linkeddata.es/vocabulary/BatteryValueChainOntology(bvco))
- [7] Sofia, R. C., Coutinho, C., Scivoletto, G., Insolvibile, G., Deshmukh, R. A., Schneider, A., ... & Mastos, T. (2023). The EFPF approach to manufacturing applications across edge-cloud architectures. Shaping the Future of IoT with Edge Intelligence, 319.

- [8] Mertens, C., Alonso, J., Lázaro, O., Palansuriya, C., Böge, G., Nizamis, A., ... & Poulakidas, A. (2022). A framework for big data sovereignty: the European industrial data space (EIDS). In *Data Spaces: Design, Deployment and Future Directions* (pp. 201-226). Cham: Springer International Publishing.
- [9] Nizamis, A., Julian, M., Valero, C. I., Foti, M., Drigkopoulou, I., Esbrí, M. Á., ... & Palau, C. E. (2025). Data-as-a-Product to enable data-driven value networks in Industries 4.0 & 5.0: The Swiss Smart Factory experiment. *Procedia Computer Science*, 257, 793-800.
- [10] Siavvas, M., Kehagias, D., Tzovaras, D., & Gelenbe, E. (2021). A hierarchical model for quantifying software security based on static analysis alerts and software metrics. *Software Quality Journal*, 29(2), 431-507.