2nd International Conference on Problems of Logistics, Management and Operation in The East-West Transport Corridor (PLMO 2023) Baku, Azerbaijan
DOI:10.54381/plmo2023.12

Feasibility study of the practical application of the data pipeline concept for improving the grain corridor efficiency using UN/CEFACT standards

Dmytro IAKYMENKOV
UNECE Consultant, DUX Consulting, Ukraine
Galyna ROIZINA
UNECE Consultant, DUX Consulting, Ukraine
E-mail: d.iakymenkov@dux.consulting, g.roizina@dux.consulting

Abstract. This article describes the results of the United Nations Economic Commission for Europe (UNECE) assessment of a feasibility study of the practical application of the Data pipeline concept. The scope of the assessment is the humanitarian transport corridor under the Black Sea Grain Initiative (BSGI). The efficiency of this transport corridor is of exceptional importance in connection with the political and economic situation in Ukraine and its impact on world food security. The main objective of the project was to discover a way to improve the efficiency of this transport corridor through trade facilitation and digital transformation of data exchange and business processes. The concept of the Data pipeline was introduced in 2014 and is actively developed by the UN/CEFACT till then. The approach develops the authors' deliverables, presented in theirs's previous work on the GUAM transport corridor [1] – such as the functional transformation of data requirements sets and distributed convertors concept. The authors also considered the experience from the participation in several single window projects in the maritime and port domain and the influence of the European eFTI Regulation. The key concept of the assessment is the use of UN/CEFACT standards and artifacts, as a holistic set of requirements that provide the possibility to harmonize data stemming from various trade operations, transport modalities, and jurisdictions. The scientific achievement of this assessment is the approach of a functional transformation over such a set of requirements. The authors developed further their own approach to apply the UN/CEFACT standards and recommendations in general and particularly the data pipeline concept for the practical projects focused on the facilitation of trade and transport procedures. This assessment shows the feasibility of the approach and can be used both as a roadmap for piloting real-world projects and as a basis for further assessments.

Keywords**:** *data pipeline concept; eFTI Regulation; UN/CEFACT standards; supply chains; Multimodal Transport*

1. BACKGROUND FOR THE ASSESSMENT

The military aggression of the Russian Federation against Ukraine led to a humanitarian catastrophe not only for Ukraine, but it also provoked a global food crisis. The blockage of Ukrainian seaports and the impossibility to export Ukrainian grain and other agricultural products to the world market led to a risk of famine in many countries and a significant increase in world food prices.

Out of the thirteen Ukrainian seaports, only three (Izmail, Reni, and Ust-Dunaysk) continued to operate. In accordance with the indicators of cargo handling for 2021, these three ports handled about

3% of the total amount of cargo in the seaports of Ukraine. A critical factor before February 2022 was the significant orientation of cargo flows to maritime transport - about 75% of the total cargo turnover and about 90% for grain and vegetable oil.

The objective limitations of the transport infrastructure of the land routes accessible for cargo transportation after February 2022 did not allow for the transshipment of goods in volumes necessary to compensate for the loss of maritime traffic.

The limited capacity of existing transport corridors and the impossibility of radically increasing it by traditional methods in feasible time has led to the need to find ways and approaches to improve the efficiency of the existing logistics infrastructure by trade facilitation and digital transformation of data exchange and business processes.

A significant factor for the re-establishment of exports of Ukrainian agricultural products was the Black Sea Grain Initiative (BSGI) - an agreement signed between the UN, Turkey, and Ukraine (and between the UN, Turkey, and the Russian Federation) on July 27, 2022, unblocking the three largest Ukrainian sea ports (Odesa, Chornomorsk, and Pivdenny) for grain exports. Already on August 1, 2022, the first ship with grain left the port of Odesa. By the end of December 2022, the BSGI allowed for additional exports of more than 14 million tons of agricultural products. This helped prevent a humanitarian catastrophe in many developing countries and stabilize world grain prices. The logistics corridor within the BSGI began to play an important role in the total volume of Ukrainian exports, and the efficiency of its functioning is critical not only for Ukraine but for the whole world.

Based on the results of the UN Development Account COVID-19 project on the digitalization of multimodal and cross-sectoral data and document exchange (https://unttc.org/stream/electronic-trade-and-transport-documents-and-data), following the recommendation of the 2022 UNECE Odesa workshop to support the digitalization of multimodal information exchange in the context of the UN Black Sea Grain Initiative(BSGI) [2], UNECE initiated an assessment to investigate the practical application of the data pipeline concept to data exchange in this emergency corridor to improve the efficiency of this logistics route using UN (UN/CEFACT) standards and recommendations. The UN/CEFACT trade facilitation recommendations [3] (currently over 40) are of utmost importance to government agencies and participants in supply chains, which investigate the development of such data pipeline projects.

2. DATA PIPELINE CONCEPT DEFINITION

The assessment included an analysis of the existing flows of documents required for the clearance of goods handled under the BSGI. This included collecting the documents used for clearance under BSGI, extracting the data set from each document, and mapping it to the UN/CEFACT Multimodal Transport Reference Data Model (MMT RDM).

As was already mentioned in the assessment [1], the MMT RDM is currently the only functionally complete data model that combines data used by all modes of transport. As part of the hierarchy of UN/CEFACT data models, the MMT RDM provides the possibility to present information exchanged in trade and transport operations along the whole supply chain at both the administrative and business levels. The model also allows for the transformation of data from trade and transport documents into the data model format of the World Customs Organization (WCO DM), which provides a full cycle of data use in the supply chain. The model is based on a harmonized data dictionary - the UN/CEFACT Core Components Library (UN/CEFACT CCL), which allows to represent accurately links to entities from a specific domain area and transforms such entities between domain areas (or jurisdictions).

The key concept of the assessment is the use of UN/CEFACT standards and artifacts, such as the MMT RDM, as a holistic set of requirements that provide the possibility to harmonize data stemming from various trade operations, transport modalities and jurisdictions. Based on this concept, the underlying data structure is directly the MMT RDM. Using a single data model for a holistic view of a logistics corridor as an element in an international supply chain allows for a practical approach to the implementation of the data pipeline concept, first introduced by David Hesketh [4] in 2014 and further developed by UN/CEFACT within its “transport and logistics domain” [5]. The assessment explores the

possibility of practical implementation of the data pipeline concept in compliance with the UN/CEFACT recommendations and standards.

Following the data pipeline concept, data is captured directly at its source and only once. It is reused throughout the supply chain, regardless of the mode of transport, the party or regulator that needs access to this data.

The basic principles of the concept of data pipelines are as follows:

- Getting data from the right source at the right place and at the right time,
- Data acquisition once and multiple use in the supply chain,
- Passing data to the pipeline at the point of origin,
- Data to be requested from the pipeline on demand and then sent to the recipient upon assessment of their responsibility.

The main goal of data pipelines is to improve the quality of data and ensure its seamless transmission within the information flow of the supply chain by shifting the paradigm from the "documentary" exchange of information to support international trade to the concept of the "datasets". The key difference between these two concepts is that the documentary model is based on a rigid (paper or electronic) document structure, while the dataset model involves the presentation of information in the form of flexible structures - datasets (business information entities and aggregated business information entities - BIE and ABIE). The latter can be formed based on universal data models on the fly upon request for the information by the recipient in a form that meets the specific requirements of the business process.

For the practical application of the data pipeline concept in a specific logistics corridor, the degree of readiness of the IT systems of its participants is of the essence. The study analyzed the degree of adoption of UN/CEFACT standards in the documents and documentary procedures required for the clearance of goods in the framework of the BSGI. A total of 17 documents were analyzed in 6 domain areas. Further, we will take a closer look at two examples.

3. THE APPROACH DEFINITION

In this study, the authors offer their views on a possible approach to the practical application of the concept of UN/CEFACT data pipelines for the BSGI corridor. This approach is based on the definition of the concept of data pipelines in the White Paper Data Pipelines, Business Requirements Specification (BRS) Data Pipeline Carrier Pipeline Data Exchange Structure (PDES) [6], as noted in previous works of the authors.

The key principle of this approach is to present the data pipeline not as “another” information system that unites the participants in the supply chain, but as a harmonized set of requirements that define interfaces for the interaction of existing IT solutions of the participants.

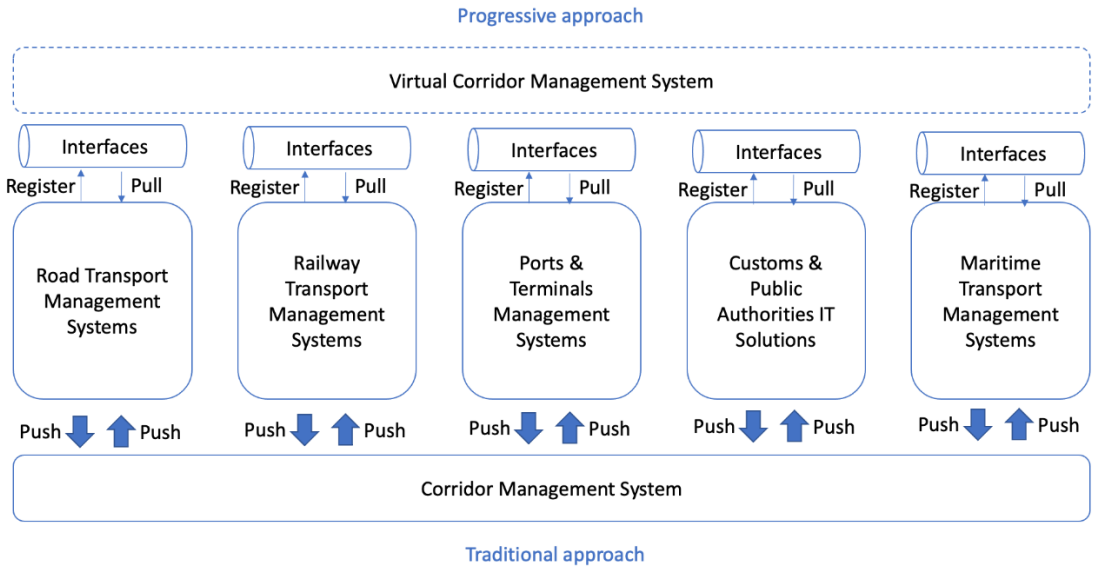

Fig. 1. Approach definition

Thus, instead of confrontation and competition with the numerous IT solutions on the market today, used and promoted by both private parties and governments, we propose a cooperation approach that considers the interests of all parties involved. The UN/CEFACT Reference Data Model Hierarchy and the Core Component Library (CCL) are considered as a common language for information exchange among participants from different modes of transport and jurisdictions. At the same time, the reference data models play the role of an information structure, and the Core Component Library the role of a dictionary. This makes it possible not to create another data model illustrating the concept of a pipeline, but to inherit it from the canonical UN/CEFACT model, recognized as an international standard. It is important to note that the mechanism used in UN/CEFACT for profiling models for a specific domain can be successfully applied in this case as well.

The technical implementation of the approach implies the use of a distributed ledger to register data sets in the pipeline, and such registration should include a unique and unambiguously identifiable link to the source of such a data set. Such a reference, no doubt, can serve as a decentralized identifier (DID) [7]

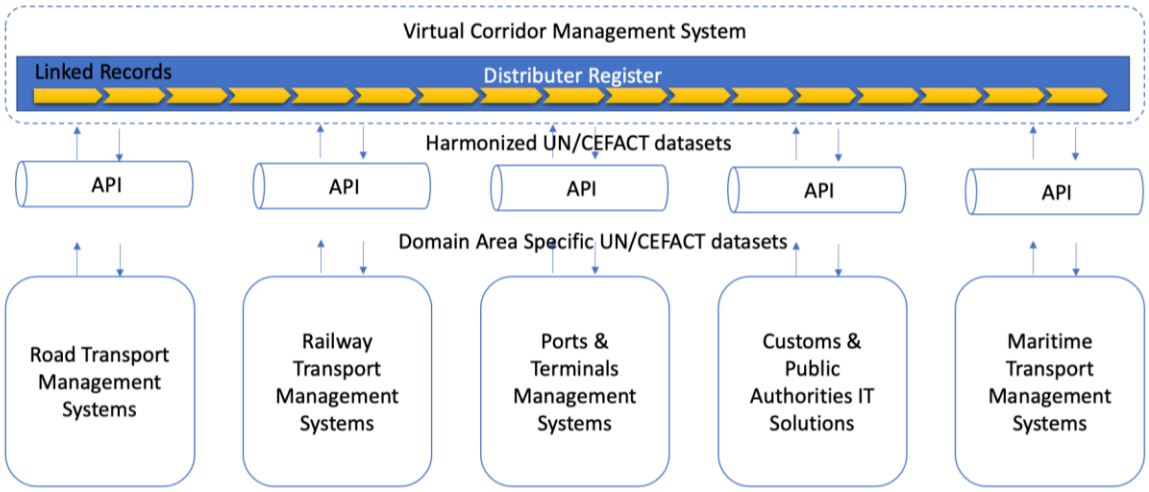

Fig. 2. Approach implementation vision

As seen from the illustration in Figure 2, interaction with existing IT solutions of supply chain participants is harmonized in the form of application programming interfaces (API). Such API can be domain-specific or even specific for a certain IT solution. The function of such API is to harmonize the specific domain-area presentation of the data sets with the canonical requirements of the Virtual Corridor Management System. Due to the assumption that domains are also utilizing the UN/CEFACT reference data models (with profiling), such harmonization can be implemented as a transformation function over the original dataset.

By proposing UN/CEFACT tools as standards for defining the formats and structures for datasets passing through the APIs in a data pipeline, the authors deliberately try to maintain technological neutrality in the implementation of a distributed ledger. To date, there are several technologies that can be successfully used to build such a ledger - including distributed file system, blockchain projects, and Internet domain name systems (DNS). Considering that the described approach correlates with the principles declared during the implementation of the EU eFTI Regulation [8], the idea of using the European blockchain infrastructure for such a registry looks promising. At the same time, for the purposes of this study, it is sufficient to understand the general architecture of the approach, while practical pilot projects will be determined by the requirements of their participants.

4. THE APPROACH APPLICATION ROADMAP

The approach described in the assessment is based on the harmonization of requirements for the integration of existing IT solutions on a particular corridor using UN/CEFACT global standards. Thus, the roadmap for the practical application of the approach can be presented in the form of three main stages:

1. Identification of the specifics of a particular corridor
2. Building a peer-to-peer network of access points (or using one of the existing ones)
3. Implementation of a corridor-specific data access layer (distributed converters)

Identification of the specifics of a particular corridor

This is the analytical part of the assessment. It uses business analysis, data analysis, and compliance analysis methods.

1. Determine the scope and boundaries of the project
2. Determine the composition of the information flow in this corridor:
    a. Identify documents or datasets relevant for the digitalization project,
    b. Identify the interaction processes,
    c. Identify possible IT solutions for the digitalization of the information flow (existing, developed, planned),
    d. Determine the degree of adoption of standards (international, UN/CEFACT, etc.) in IT solutions within the scope of the project.
3. Specify the details of the information flow of the corridor:
    a. Select/collect datasets from documents,
    b. Map datasets to Reference Data Models,
4. Create an information model for this corridor:
    a. Profile the Reference Data Model for project tasks,
    b. Describe the event model,
    c. Define the Business Requirements Specification (BRS) details.

In this assessment the scope of the project was limited to the BSGI transport corridor. The data analysis and business analysis performed during the assessment help the author to identify the main information flow, the list of documents, and the datasets from these documents. This allows to map the data with the reference data model that is the UN/CEFACT MMT and to propose a subset (profile) of

such model for the project scope. The detailed report of the assessment and project deliverables is accessible on the www.unttc.org site.

The important part of the concept and valuable part of the assessment is the compliance analysis of the UN/CEFACT standards adoption level of the IT solutions in the scope of the project. The importance of this work is caused by the hypothesis described earlier that domain-specific IT solutions can present the dataset in the UN/CEFACT profiled data models. Compliance analysis uses the UN/CEFACT Recommendation 36 on interoperability among (electronic) Single Window systems[9]. The recommendation identifies five degrees of acceptance of the standard:

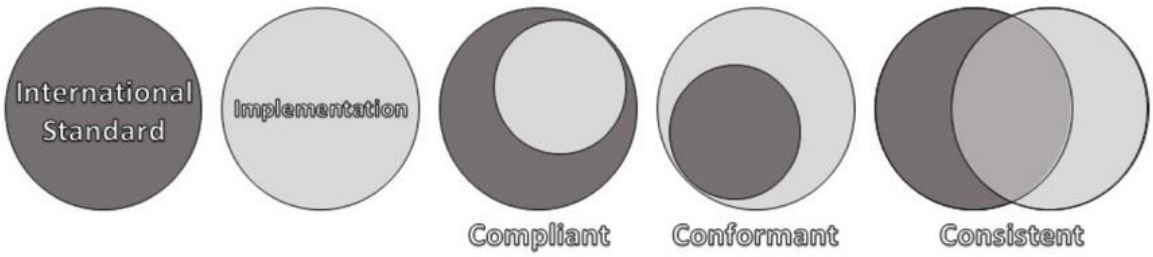

Fig. 3. Degrees of adoption of international standards in cases of implementations

We can talk about compliance (compliant) when the implementation of a particular solution is determined solely using the terms adopted for a particular standard, and within its scope. If, when implementing a particular solution, one or another standard is entirely used and additions are included on its basis, then we can talk about conformity (conformant). In this case, it may turn out that new additions are functionally incompatible with other solutions since they are not covered within the specified standard.

When implementing a particular solution, only separate parts of a standard are used, and additions are included on this basis, then we can talk about consistency (consistent). However, even in this case, it may turn out that additions are functionally incompatible with other solutions since they are not covered within the specified standard. Moreover, due to the use of only part of the specified standard, it may well turn out that the other party using the same standard will not be able to adapt to this solution, since certain parts of the standard will be missing in such an “agreed” solution.

Based on the methodology proposed above, the compliance analysis of the adoption of standards was carried out considering this gradation. For the convenience of analyzing the documents involved in the work of the grain corridor, the analysis is divided by mode of transport.

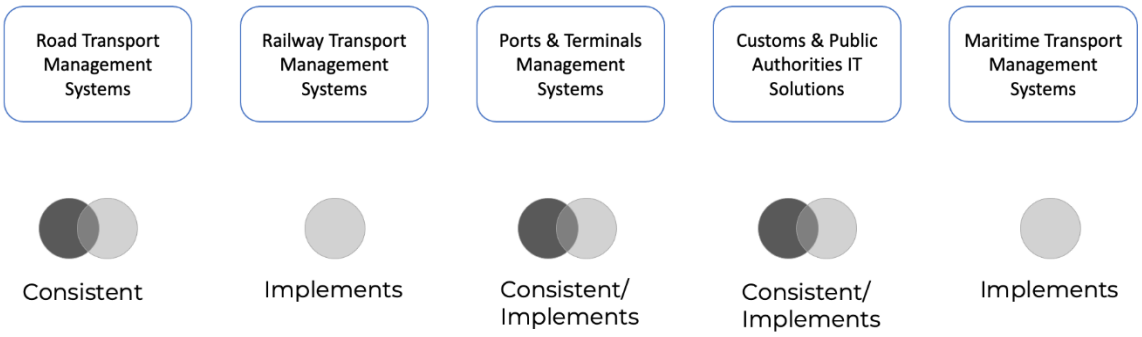

Fig. 4. Degree of the standard acceptance evaluation

Building an Access Point Peer-to-Peer Network

In the previous works [10], authors already described the high level of mistrust between parties involved in the transportation process in general and in transborder interaction particularly. The feasible solution of such an issue could be an equal-rights network (of IT solutions) – a pear-to-pear network. The key advantage of such an approach is the absence of a single central point of making decisions (or central authority point) and thus – the equal rights and requirements for all parties – both public and private independently from their residence.

The key advantage of the authors' approach to harmonize the requirements for the existing IT solutions is that the peer-to-peer network of access points can be either created specifically for the task of digitalizing certain transport corridor, or one of the existing similar networks can be used. As an example, we can cite the concept of a digital global network of logistics and transport providers in the framework of IATA's OneRecord project [11]. The concept of nodes in the implementation of the European eFTI Regulation can also be applied as such a peer-to-peer network.

The basic architecture of the peer-to-peer access point network includes four layers:

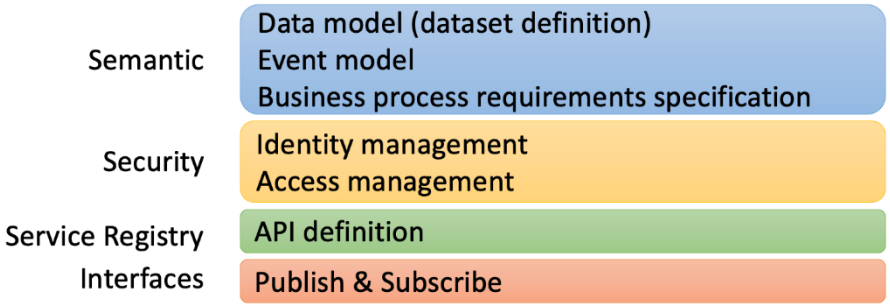

Fig. 5. The lavers of the access point architecture

The semantic layer is determined by the deliverables of the previous stage. The datasets formed on it become information entities for exchange within the framework of the network that is being created. The creation of a node (access point) in such a network can be described by the following sequence of actions (as an illustrative example):

1. Define a distributed identity schema for Information Entities (datasets),
2. Define the Security and Access Control mechanism for the Information Entities,
3. Create basic structures for storing information about the process of interaction in the network in the form of distributed registers (ledgers),
4. Implement the underlying API for the Information Entities defined in the previous stage.

Implementing a corridor-specific data access layer

This stage is necessary to link the universality of the peer-to-peer network, the functional completeness of reference data models, and the specifics of implementations of certain IT solutions in a particular corridor.

Depending on the results of the assessment of the level of adoption of standards, carried out at the first stage, the functionality and complexity of the implementation of distributed converters may differ significantly. In the simplest case - when a specific IT solution is implementing the standard - it is possible to organize direct access to data. In case of the BSGI transport corridor, the high level of the standards' adoption was detected.

The functionality of distributed converters can be described as follows:

1. (optional) Transforming the internal identifiers into decentralized identifiers in accordance with the identification scheme adopted in the peer-to-peer network,
2. Maintaining the canonical "interface" data model,
3. Transforming the data structure between the canonical structure of reference model and the internal representation of a particular IT solution,

4. Harmonization of non-standardized code lists,
5. Transforming the data format between an internal representation and one that is supported by a peer-to-peer network,
6. Transforming the API calls,
7. Transforming security layer requirements.

Overall data pipeline process mapping diagram for the BSGI

To illustrate the overall process mapping for a data pipeline implementation in the Black Sea Grain Initiative (BSGI) we use the documents listed in the Table 4 (Section 4.2) and put them in a data pipeline diagram, as was presented by the David Hesketh in his publication "Enhancing business and government interactions in global trade" [12]

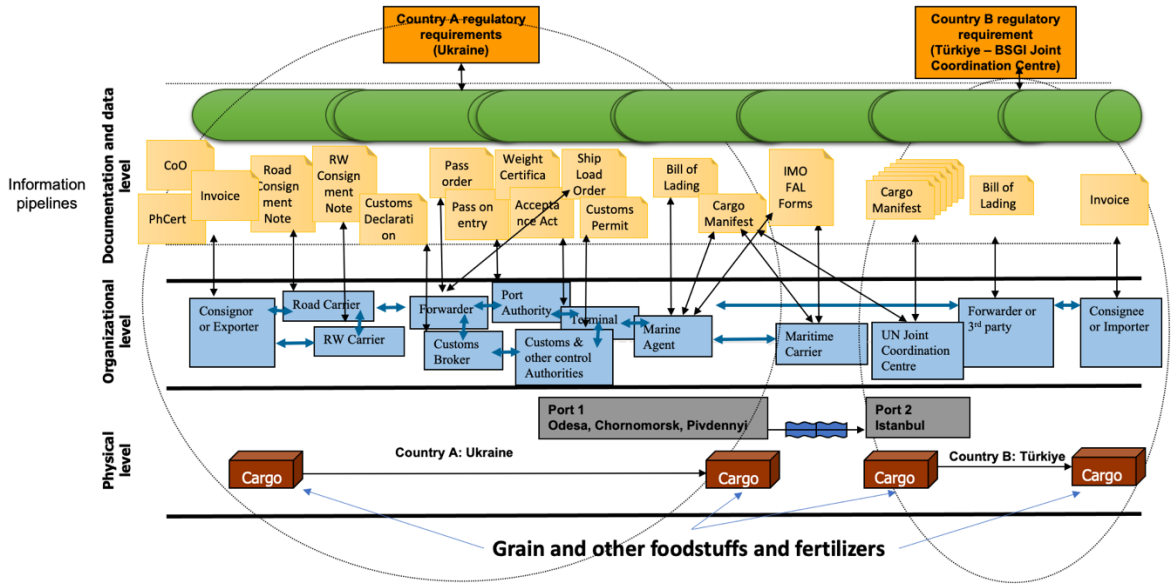

Fig. 6. Data pipeline process mapping in the BSGI transport corridor

As it was mentioned previously, the documents are interpreted as datasets. Moreover, the mapping of such datasets to the canonical reference data models of UN/CEFACT allows to operate with datasets as separate classes, that provide fine grained data access control. The data structure for exchange along the possible data pipeline will be defined by the Multimodal Transport and other UN/CEFACT reference data models.

5. THE FUNCTIONAL TRANSFORMATION

The key factor to ensure the interoperability of the parties' interaction with the data pipeline without the need to develop a specific solution for such interaction is the usage of the UN/CEFACT profiles of the reference data model for a specific domain area. In this case, the semantic requirements are fully consistent since the domain area is a subset of the canonical DM requirements. Thus, the transformation from the domain dataset to the pipeline dataset can be realized due to the functional completeness of the canonical model, the reverse transformation is also realized due to the functional completeness of the profile for a specific subject area (i.e. attributes that are not reflected in the profile of a specific domain are considered as non-functional in the scope of this domain). For transforming the representation formats of data sets - the transformation is implemented as a subset, which is the intersection of sets of formats supported by both the domain area and the data pipeline. In any case, there is at least one

representation format that is guaranteed to be supported by both scopes, namely the official schema publication format for the canonical data model and domain profiles.

The feasibility of this hypothesis to use UN/CEFACT standards for a specific transport domain is shown earlier in the compliance analysis description. To formalize this process, the distributed converters approach is proposed by the authors.

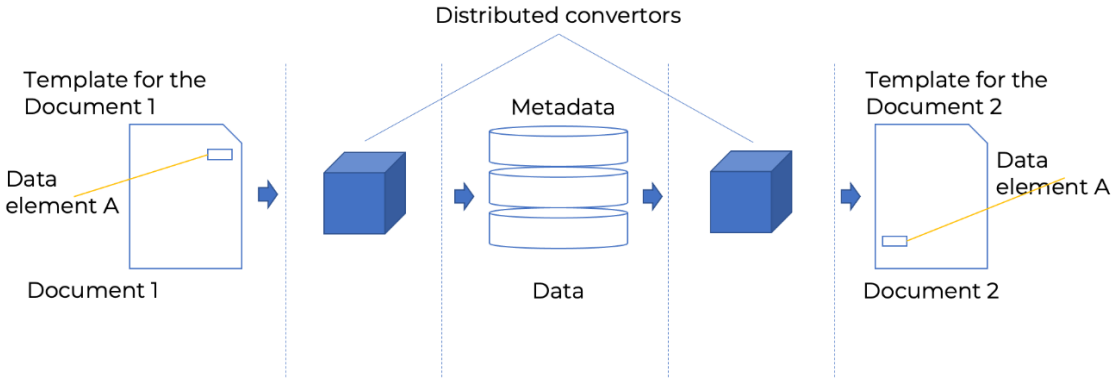

Fig. 7. Distributed conversion content

The documents involved into the process of informational interaction can be represented as a data sets from these documents, or a set of data attributes:

$$D = [A^1, A^2, \dots, A^N]$$

The presentation of the document in the particular jurisdiction (we assume the jurisdiction not only as a national, but also as a modality of transport) can be described also as a set of requirements:

$$R = [R^1, R^2, \dots, R^M]$$

The process of the in informational interaction in that case can be represented as function over such sets of attributes:

$$P = F(D, R)$$

The transformation of the data attributes always is caused by the transformation of the particular subset of requirements that influence on such attribute while changing the jurisdiction. Thus, the interaction process can be presented as a function over the set of requirements (that is in its turn a set of functions over a set of data attributes):

$$P = F'(R), where\ R = [F_1(D_1), F_2(D_2), \dots, F_M(D_M)]$$

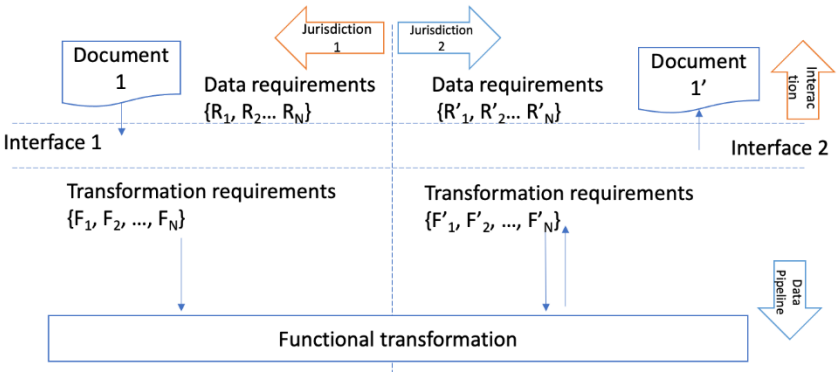

Fig. 8. Functional transformation of the requirements within the informational interaction process

Such functions F_1-F_M are the transformation requirements for the data requirements. Due to this, such requirements can be described formally. Harmonization of such requirements description lets us consider them as input parameters for the universal "transformation" function, that executes such converting. As a practical example for describing both sets of requirements, the Extended Markup Language (XML [13]) is proposed, that allows to use of an XSLT [14] transformation for describing the transformation algorithm. The execution of such transformation can be easily implemented in different IT solutions including open sources [15].

6. CONCLUSIONS

The authors developed further their own approach to apply the UN/CEFACT standards and recommendations in general and particularly the data pipeline concept for the practical projects focused on the facilitation of trade and transport procedures. The other deliverables are the dataset for the BSGI transport corridor, mapped with the canonical UN/CEFACT reference data model, compliance analysis results, bottlenecks elicitation, and recommendations. This assessment shows the feasibility of the approach and can be used both as a roadmap for piloting real-world projects and as a basis for further assessments.

References

1. Iakymenkov D., Roizina G., Alignment of documents used in multimodal transportation along the GUAM transport corridor with the UN/CEFACT reference data model. Available at: https://unttc.org/documents/alignment-documents-used-multimodal-transportation-along-guam-transport-corridor-uncefact
2. The United Nations, Initiative on the Safe Transportation of Grain and Foodstuffs from Ukrainian Ports, https://www.un.org/sites/un2.un.org/files/black_sea_grain_initiative_full_text.pdf
3. The United Nations Economic Commission for Europe (UNECE), Trade Facilitation Recommendations, https://unece.org/trade/uncefact/tf_recommendations
4. CLECAT, Data Pipelines: Innovation in Supply Chain Visibility, David Hesketh, https://www.clecat.org/media/David_Hesketh.pdf
5. The United Nations Economic Commission for Europe (UNECE), Data Pipeline Project, https://unece.org/fileadmin/DAM/cefact/cf_forums/2019_Geneva/T_L_DataPipeline2019.pdf
6. The United Nations Economic Commission for Europe (UNECE), Business requirements specification, data pipeline carrier pipeline data exchange structure (pdes), https://unece.org/fileadmin/dam/cefact/brs/t_l-brs_datapipeline_v1.pdf
7. The United Nations Economic Commission for Europe (UNECE), Regulatory & egovernment programme development area edata management domain, White paper verifiable credentials for cross border trade , https://unece.org/sites/default/files/2022-06/010_Verifiable-Credentials-CBT.pdf and https://www.w3.org/2022/07/pressrelease-did-rec.html.en
8. EUR-Lex, Regulation (EU) 2020/1056 of the European Parliament and of the Council of 15 July 2020 on electronic freight transport information https://eur-lex.europa.eu/eli/reg/2020/1056/oj
9. The United Nations Economic Commission for Europe (UNECE), Single Window Interoperability, Recommendation No. 36, https://unece.org/fileadmin/DAM/trade/Publications/ECE-TRADE-431E_Rec36.pdf
10. Iakymenkov D., Roizina G., Assessment of the use of electronic documents based on the UN/CEFACT Multimodal Transport Reference Data Model (MMT RDM) in the execution of formalities for the registration of the arrival and departure of ships in the seaports of Ukraine,

https://unttc.org/documents/assessment-use-electronic-documents-based-uncefact-multimodal-transport-reference-data
Iakymenkov D., Roizina G., Digitalization of data and document exchange for multimodal transport and trade in Azerbaijan, using the UN/CEFACT standards and reference data models., https://unttc.org/documents/digitalization-data-and-document-exchange-multimodal-transport-and-trade-azerbaijan-using

11. IATA, One record, https://www.iata.org/en/programs/cargo/e/one-record/
12. ResearchGate, Enhancing business and government interactions in global trade, https://www.researchgate.net/publication/233758106_Enhancing_business_and_government_interactions_in_global_trade
13. W3C, Extensible Markup Language (XML) 1.0 (Fifth Edition), https://www.w3.org/TR/xml/
14. W3C, XSL Transformations (XSLT), https://www.w3.org/TR/xslt-10/
15. Freeformatter.com, https://www.freeformatter.com/xsl-transformer.html