Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2022) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

Transport Research Arena (TRA) Conference

5G and Logistics Data Collection: the 5G-LOGINNOV Approach

Djibrilla Amadou Kountche[a]*, Mandimby Ranaivo Rakotondravelona[a], Pavlos Basaras[b], Ralf Willenbrock[c], Janez Sterle[d], Rudolf Susnik[d]

[a]*AKKA HIGH TECH, 7 Boulevrad Henri Ziegler, BLAGNAC 31700, France*
[b]*INSTITUTE OF COMMUNICATION AND COMPUTER SYSTEMS, Patission Str. 42, ATHINA 10682, Greece*
[c]*T-SYSTEMS, Hahnstrasse 43D, FRANKFURT AM MAIN 60528, Germany*
[d]*INTERNET INSTITUTE, CRNA VAS 128, LJUBLJANA 1000, Slovenia*

Abstract

Collaboration in logistics fosters the development and testing of new use cases especially those involving new technologies such as Internet of Things (IoT) or 5G. In this regard, 5G-LOGINNOV, a 3-year Horizon 2020 project aims at demonstrating the added values of 5G in port logistic operations. 5G-LOGINNOV will deploy and trial 11 use cases in 3 Living Labs (LLs): Athens, Hamburg, and Koper ports. This paper evaluates 5G-LOGINNOV approach to collect 5G and logistics data and its impact to data analysis and evaluation efficiency. Indeed, the data produced by the trials will serve in the evaluation of the impact of 5G technologies on the port operations as well as the socio-economic and environmental effects. Moreover, collaboration relies on an efficient data sharing between the stakeholders. Even though normalized by the United Nations Centre for Trade Facilitation and Electronic Business (UN/CEFACT), GS1 standards used in Global Data Synchronization Network (GDSN) and others, collecting logistics data in 5G related projects is still an important challenge. We have identified some key principles laying the basis for an efficient data sharing in the context of an evaluation: (i) Carefully designed metadata to document the shared data while maximizing comparability; (ii) Well defined interfaces between the stakeholders to smooth data sharing; (iii) Tools for data management accessible to all the stakeholders enhancing data usability. We have implemented these principles in 5G-LOGINNOV first by designing the metadata according to the requirements from the evaluation tasks and the trials description. Then we proposed harmonized interfaces for project-level data sharing between the LLs. Finally, we designed and developed a central data collection tool for data ingestion, indexing, visualization, and publication.

Peer-review under responsibility of the scientific committee of the Transport Research Arena (TRA) Conference

* Corresponding author. Tel.: +33772660650
E-mail address: djibrilla.amadou-kountche@akka.eu

2352-1465 © 2022 The Authors. Published by ELSEVIER B.V. This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the Transport Research Arena (TRA) Conference

Keywords: Collaboration; Data sharing; Logistic;, 5G; ORDP; FAIR

1. Introduction

Collaboration in logistics is increasingly important to overcome the challenges of a global economy where local incidents on the supply chain might have global consequences such as the Ever Given obstruction of the Suez Canal (BBC, 2022). Other events such as the CoVID-19 pandemic have implications which are hard to assess.

In the European Union (EU), important part of the global exchanges is done through ports which face several challenges including a fragmentation of views and interests of partners in the same port, operational problems, peak hour management to avoid bottlenecks, slot booking and capacity management. Digitalization including mobility data spaces, data exchanges between ports and replicable Information Technology (IT) System is considered as an answer to some of these challenges. In (BOOSTLOG project, 2021), the authors identify the main barriers hindering the coordination and the collaboration in logistics: scalability of governance, business and operational models, complexity of transition management, legal issues and soft/behavioral aspects as well as data sharing.

This paper focuses on the data exchange aspects which are at the heart of the collaboration between the stakeholders in the era of digitalization. The main challenges with respect to the data exchange include the proper and harmonized documentation of the data (metadata), the harmonization of the data models, the supporting technical solutions, data protection, privacy, and the preservation of business knowledge. Our approach to tackle the harmonization of the data sharing between ports stakeholders relies on a careful documentation of the data, a proper description of the data models and a central data collection tool supporting the data sharing. We have implemented our approach in 5G-LOGINNOV which aims to exploit 5G technology in 3 LLs and combine the hinterland aspects of Hamburg LL as well as in port operations of Athens and Koper. The set of main 5G technologies to be considered as part of the LL and use cases (UCs) evaluation are Enhanced Mobile Broadband (eMBB), Massive Machine Type Communication (mMTC), and Ultra-Reliable Low Latency Communication (URLLC) capabilities with 5G network slicing support, Physical Network Function (PNF) and Virtual Network Function (VNF) capabilities and corresponding management and orchestration (MANO), and finally Multi-access Edge Computing (MEC) and site Infrastructure as a Service (IaaS) capabilities as part of low-latency and operational data protection requirements (5G Observatory, 2022).

In recent years, several H2020 project were conducted on several challenges (especially data collections aspects) related to Logistics such as AEOLIX (European Commission, n.d.), SELIS (Selis project, 2022), 5G-GROWTH (5GGROWTH project, 2022), BOOSTLOG (BOOSTLOG project, 2021) and BE OPEN (The BE OPEN project, 2022) to name but few. SELIS produced a Big Data Analytics and Machine Learning System to address big data issues in logistics domain while BE OPEN studied various transport data model and provides e Big Data stores and the SELIS Knowledge Graphs and the SELIS Knowledge Observatory on transport data. Finally, for Field Operational Test (FOT) projects, The FESTA methodology (FOT-NET, 2022) devised a detailed description of the dataset which is important to take into consideration for any FOT projects. Also, several open-source projects around data collection, ingestion and storage have been undertaken for e.g. The Telegrah, InfluxDB, Koregraph and Kapacitor or TICK stack, the Elastic stack, the Elastic Common Schema (ECS), Fluentd and FluentBit, Apache Flink and Apache Samza to name but few. Our approach is based on Elastic Common Schema and adapt it to the logistic domain and to FOT.

2. The 5G-LOGINNOV's Approach

2.1. Overview

5G-LOGINNOV comprises a palette of port-driven technological and societal innovations, tailored to realise its objectives. In keeping with the EU Green Deal program (European Commission, 2022), 5G-LOGINNOV will implement and deploy: (i) a Green Truck Initiative using Connected, Cooperative and Automated Mobility (CCAM) & automatic truck platooning based on 5G technological blocks; (ii) a management and network orchestration platform, a device management platform ecosystem, and algorithms for optimal selection of yard trucks, optimal surveillance using video analytics and predictive maintenance; floating truck and emission data analysis; (iii) an automation for ports: port control, logistics and remote automation; (iv) experiment 5G in mission critical communications in ports, for Green Light Optimal Speed Advisory (GLOSA), Automated Truck Platooning and dynamic control loop for environment sensitive traffic management actions. The experimentation (also called storyboard) will be done in real operating conditions in 3 LLs environments, associated with Athens, Hamburg and Luka Koper ports. As part of 5G port LL, validation of a set of 5G technologies and its capabilities is going to be realised and evaluated for their relevance with regards to the port control, logistics, remote automation, surveillance and sensing, and security operation support. Thus, this validation requires data to be collected.

2.2. Data collection approach

Data collection in 5G-LOGINNOV is based on an adaptation of FESTA methodology to: (i) Devise the trial planning; (ii) Conceive the internal data production/collection tools; (iii) Design a common metadata model; and (iv) Conceive a centralized data collection tool. In H2020 projects, flexibility in defining data models is important to cope with the project's phases and challenges. Trial planning guarantees that scientific evaluators, port operators and LL leaders define beforehand the requirements on the trials for a given deployment of the project to ensure data quality and protection and prevent data loss.

Data production occurs in the LLs during the execution of the storyboards. However, each storyboard is executed in a different environment and setup with the implication of several partners. Therefore, to prevent data loss and ensure the coherence of the collected data, a coordination is mandatory between the LL partners. Each partner ensures the acquisition of certain data and applies internal data processing. This data production can be manual or rely on automatic processes and can involve many data producers. Therefore, each LL can be organized following two scenarios: (i) the first one corresponds to the case where a local data storage is used in a LL before data being sent to central tools used for scientific evaluation; (ii) the second scenario as shown on Fig. 1 covers independent partners in each LL where each partner has its own data aggregation and storage facilities and forward data to the central tools on its own.

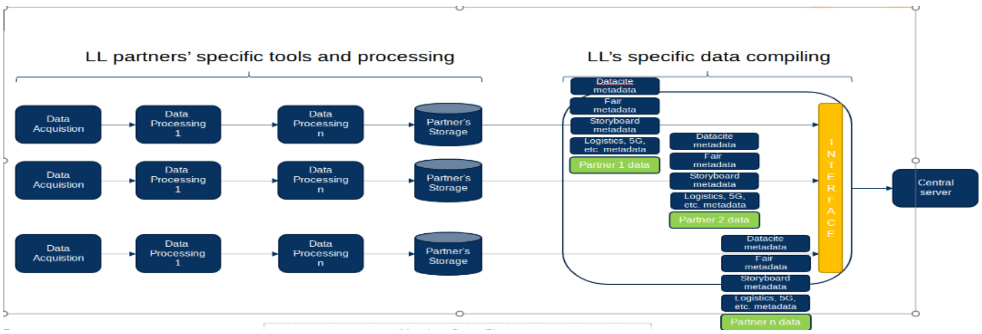

Fig. 1 Second scenario of LL's local data acquisition, processing, and storage

After the acquisition, the data are stored in the partner's own storage before being transferred to the LL's central data storage. From the LL's data storages, the evaluation data is transferred to the central data collections tool with the necessary metadata. Every data set sent through the data collection tools follows a common data model which is an important achievement of the methodology.

2.3. Metadata

The current metadata model, as show on Fig. 2, is a combination of several properties taken from the DataCite metadata schema (DataCite Metadata Working Group, 2021), Findable, Accessible, Interoperable, Reusable (FAIR) metadata and a metadata devised for 5G-LOGINNOV's storyboards. The properties of each metadata category are described in a table with the following columns:

- Property: the name and identifier of the metadata property

- Sub-property: the name and identifier of the metadata sub-property if any
- Required/Optional: indicates whether it is mandatory to provide the property or not
- Description: indicates what data are expected in the field

The metadata categories used in 5G-LOGINNOV are described in the following subsections.

2.3.1. DataCite and FAIR metadata

DataCite is a non-profit organization acting as a persistent identifiers (DOIs) provider. DataCite proposes an extensive metadata schema that has been adopted by other persistent identifiers providers. A subset of the DataCite metadata fields is used for 5G-LOGINNOV evaluation data.

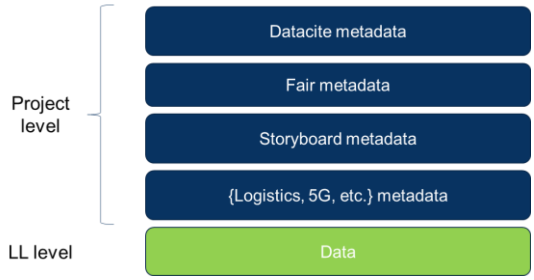

Fig. 2 5G-LOGINNOV's metadata model

FAIR metadata category can be divided into three sub-categories:

- Administrative metadata: they correspond to management data about the resource such as the owner, the collaborators, the funding, etc.
- Descriptive metadata: they consist in the data describing the resource allowing its discovery and identification such as the unique identifier, the title, the abstract, etc.
- Structural metadata: they provide information on how the resource has been constituted and how the resource is internally organized.

To sum up, FAIR metadata sub-categories are already covered by the DataCite and for each dataset, in addition to the DataCite metadata, the relevant data model must be provided as part of the FAIR metadata.

2.3.2. Storyboard, Logistics, 5G and other metadata

Storyboard metadata provide the relevant information to set up and perform the use cases during the trial. As such, they describe the context in which the evaluation data have been collected. Hence, the storyboard metadata covers identification and description of the storyboard, use cases, experience iteration, KPIs baseline and particular setup. Finally, our three steps approach allows other metadata models to be readily included in our model.

2.4. Data

During the trial, several data types can be collected and organized by project's needs. These types are described for each LL. For each dataset, the following information are provided: (i) Field: the name of the data field; (ii) Type: the field type for e.g., string, integer; (iii) Unit: the unit of the value of the field; (iv) Required/Optional: indicates whether it is mandatory to provide the field or not; (v) Description: short explanation of what is the data field about.

3. Implementation

The implementation of the data collection tool is based on an internal data production at each LL managed by an instance of data ingestion tool which can pre-process the data (data anonymization, data quality checks, etc.). The data are aggregated by a central data aggregator which can also apply additional data pre-processing before sending the data to a database. The central data collection tool offers a data visualization tool, an authentication and authorization using OAuth 2.0 and OpenID Connect. Finally, Open Research Data Pilot (ORDP) aspects are managed by the central data aggregator which sends data to ZENODO using a REST API. The next sub-sections describe the data collection process in the LLs and their interfacing with the central data collection tools.

3.1. Data sources from the Living Labs

3.1.1. Athens LL

Athens LL provide: (i) for 5G use cases: telemetry data from yard truck (CAN-Bus, localization, container presence, Video and other custom sensor data) transmitted to the PCT management platform via the integrated (on-truck) 5G telematics device; (ii) for the traffic monitoring system (TMS): aggregated telemetry data from the fleet of 5G trucks: logged/historical, operational and use case data of daily port operations for business and analytics; (iii) for the Enterprise Asset Management System (EAMS) at PCT datacenter: vehicles operational status (historical and recent data), e.g., hours of operation, breakdown events and duration, parts of the truck affected by a malfunction and the spare parts used for the repair, fuel consumption etc.; (iv) for external trucks inbound at the Athens Port: operational data are collected and processed at the Vodafone Innovus platform; v) Additionally, for machine learning and computer vision: key performance indicators (e.g., inference time, inference accuracy) to validate the efficiency of 5G&AI-enabled video analytics services targeting safety, security, operational efficiency, and logistics applications in Athens LL. Such software components are deployed as cloud-native services based on e.g., Kubernetes, establishing ease of transfer to heterogeneous infrastructures and monitoring of data collection activities. The transfer of data from LL Athens to the central data collection tools is done using a data ingestion tools named Fluentd.

3.1.2. Hamburg LL

Hamburg LL mostly provide processed data to the central data collection tool. If applicable, data available from partners Continental, tec4u, SWARCO and T-Systems are collected and aggregated in a data collection instance whose main purpose lies in acting as the main interface to transfer data to the central data collection instance via Fluentd. As the data foreseen for the evaluation tool is the result of complex algorithms it is not applicable for detailed processing by the evaluation tool and is only intended for limited aggregation and visualization.

3.1.3. Koper LL

Koper LL provide centralized monitoring point based on Fluentd. Data sources for multiple inputs to the Koper LL central collector are realized with Fluentd-based KPI collectors that are provisioned per use case partner (INTERNET INSTITUTE LTD, VicomTech, and Continental SRL, while Luka Koper and Telekom Slovenia are supporting use cases). The ININ KPI collector, a qMON Collector (responsible for gathering and parsing KPIs from qMON Agents) entity is extended to support the extraction of relevant 5G-LOGINNOV KPIs from the native qMON KPI datasets to Fluentd interface. For the other two partners, dedicated Fluentd agents process the files from the partner's use case scenarios. This way, the only requirements from use case partners are to provide JSON-based files containing relevant 5G-LOGINNOV KPIs and be able to put those files in selected folders as shown on Fig. 3.

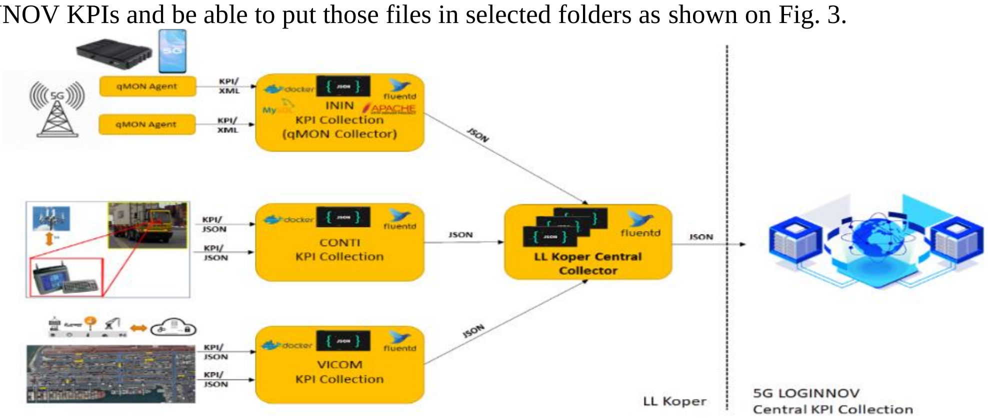

Fig. 3 Koper LL's data collections tools as illustration of typical LL data collection

3.2. Central Data Collection Tools

The central data collection tool relies on a set of open-source projects as shown on Fig. 4. The design of this architecture was based on the requirements from several project's partners which drove the choice of the components and modules with their roles: Fluentd for local data ingestion, central data aggregation and ORDP, Elasticsearch as database; Kibana for visualization and Keycloak for Authentication and authorization.

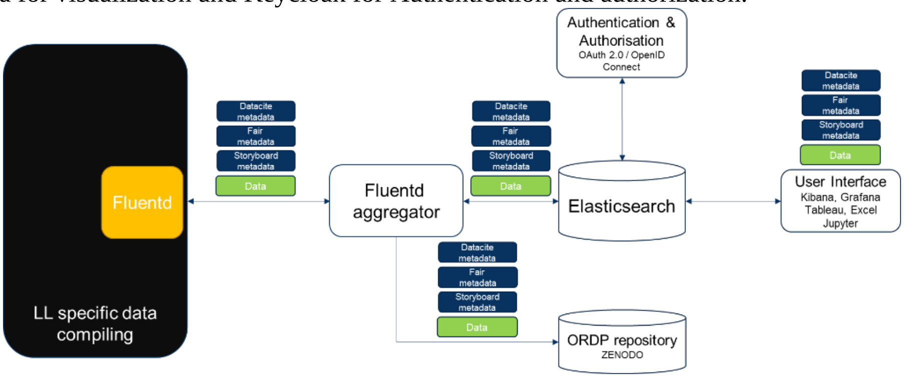

Fig. 4 Central data collection tools

Fluentd is widely used as data collector and aims to: (i) Unify the logging with a common JSON data structure; (ii) Offer a pluggable architecture based on data sources, data filters and data outputs; (iii) Use as minimum resources as possible and (vi) Offer a built-in reliability. In the central data collection tools, it aggregates received data from the LLs on an interface which uses the MessagePack protocol. It checks the data quality and ensures data anonymization using data filters (pre-defined or provided using for e.g., Python or Ruby) and send the data to the database.
Elasticsearch is the distributed search and analytics engine at the heart of the Elastic Stack. It provides near real-time search, analytics for all data types and data streams management capabilities. In 5G-LOGINNOV, Elasticsearch is used primarily as a database but other features such as searching, and the REST API are also exploited. Also, index templating is used to cope with data and metadata models evolution. To ensure consistency of the data, an index is created for each LL with static typing as follow: ll-athens, ll-hamburg, ll-koper and evaluation-kpis. Upon reception of data from Fluentd, Elasticsearch inserts them in the appropriate indexes when compliant with the data models. In case of error for e.g., non-compliance with the data models, a message is sent to Fluentd. Several user interfaces can be used for the evaluation purposes among which are Microsoft Excel, Jupyter, Grafana and Kibana. As part of the Elastic stack, Kibana is the default user interface provided by the central data collection tool. It uses the Elasticsearch API to manage Elasticsearch clusters and indices and serves as a data visualization and evaluation tool supporting Machine Learning libraries. Results from the evaluation are stored in a separate index on Elasticsearch.

Data protection is ensured by the usage of OAuth 2.0, OpenID Connect and TLS. Keycloak is the chosen implementation. The central data collection has been deployed using Ansible. Ansible roles have been created for each of the components (Fluentd, Elasticsearch, Keycloak and Kibana) and have been deployed on Microsoft Azure using a dedicated Ansible playbook.

3.3. Generation of Common Data Format

The implementation of the common metadata format and data models is shown on Fig. 5. The first step has been extensively described in section 2.2. The second step uses YAML to further describe and document the metadata and data models as shown on Table 1. The YAML files are then parsed and translated using Python scripts into JSON files consisting in the Elasticsearch indexes. The third and final phase as shown on Fig. 6 consists in the configuration of Elasticsearch with these indexes, publication of the data and scientific evaluation. The outcome of the third phase can serve to improve all the other steps.

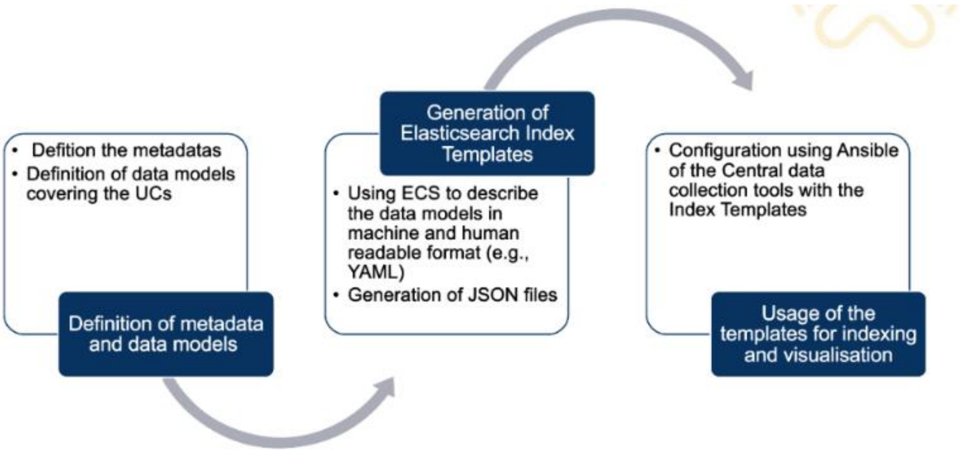

Fig. 5 5G-LOGINNOV Common data model generation procedure

Table 1 Excerpt of YAML description of 5G-LOGINNOV's data models

```yaml
---
- name: datacite
  title: Datacite
  group: 2
  short: metadata Fields
  description: >

  type: group
  fields:
    - name: identifier
      level: extended
      required: true
      type: keyword
      description: >
```

```yaml
---
- name: athens
  title: Athens
  group: 2
  short: Data fields of LL Athens
  description: >
  type: group
  fields:
    - name: inference_time
      level: extended
      type: float
      required: false
      short: The time..
      description: >
```

```json
{
  "_meta": {
    "description": "LL Athens composable template",
    "version": "8.0.0"
  },
  "composed_of": [
    "vglcs",
    "datacite",
    "storyboard",
    "athens"
  ],
  "index_patterns": [
    "ll-athens-*"
  ],
  "priority": 1,
  "template": {
    …
```

Fig. 6 LL Athens Elasticsearch index template

4. Conclusions and Perspectives

Our approach consists in the definition of flexible metadata models and data models which can be used by the various partners of a LL (and by extension by the various partners intervening in Logistics) to describe the data to be collected. This approach allows different levels of data sharing while at the same time, ensures compliance with data protection regulations by providing the LL with the mechanism to only share the data needed for a given purpose such as the evaluation. Therefore, we have created a process by using human and machine-readable languages (for e.g., YAML) to describe the data independently from a particular implementation and translate them in the specific format needed by the database implementation which is in our case Elasticsearch. Thus, our approach allows several data models to be used if they are described into YAML and support various databases (SQL and NoSQL).

In 5G-LOGINNOV, our approach form adapted ECS was tested in 3 LLs which and facilitated data exchange between the distinct partners of the ports. It resulted in the designing the metadata according to the requirements from the evaluation tasks and the trials description, provided a harmonized interface for project-level data sharing between the LLs and collaboration with scientific partners using the collected data for project evaluation. Our perspectives are to support and test: (i) other metadata and data models particularly the FESTA data models; (ii) generation of other databases for e.g., PostgreSQL and MongoDB and (iii) finally published our work as an open-source project.

Acknowledgements

5G LOGINNOV is co-funded by the European Commission, Horizon 2020 research and innovation programme under grant agreement No. 957400 (Innovation Action).

References

5G Observatory. (2022). *European 5G Observatory*. Récupéré sur https://5gobservatory.eu/

5GGROWTH project. (2022). Retrieved from 5GGROWTH: https://5growth.eu/

5G-LOGINNOV. (2021). *D3.1 Trial methodology, planning and coordination.*

BBC. (2022). *The cost of the Suez Canal bloackade*. Retrieved from https://www.bbc.com/news/business-56559073

BOOSTLOG project. (2021). *Cloud report – Coordination and Collaboration.*

DataCite Metadata Working Group. (2021). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs. doi:https://doi.org/10.14454/3w3z-sa82

Economic Commission for Europe. (2017). *White Paper on a Reference Data Model.* Retrieved from https://unece.org/DAM/cefact/cf_plenary/2017_Plenary/ECE_TRADE_C_CEFACT_2017_11E__White_Paper_on_RDM_.pdf

European Commission. (2022). *A European Green Deal*. Retrieved from https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en

European Commission. (n.d.). *AEOLIX*. Retrieved from https://cordis.europa.eu/project/id/690797

FOT-NET. (2022). *Welcome to the wiki of Field Operational Tests*. Retrieved from fot-net: https://wiki.fot-net.eu

Selis project. (2022). *SELIS project*. Retrieved from SELIS: https://selisproject.eu/

Sorkun, M. F. (2020). Digitalization in Logistics Operations and Industry 4.0: Understanding the Linkages with Buzzwords. doi:10.1007/978-3-030-29739-8_9

The BE OPEN project. (2022). *Be OPEN project*. Retrieved from beopen: https://beopen-project.eu/

Wilkinson, M., Dumontier, M., & Aalbersberg, I. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data, 3*. doi:https://doi.org/10.1038/sdata.2016.18