Ref. Ares(2023)1471386 - 28/02/2023

Twin Transition for The Timber Industry

Deliverable 1.5 Initial 5G-TIMBER data flow architecture for the circular material and production process management

5G-TIMBER | Work Package 1, Task 1.3
Final Delivery Date: 28-02-2023

Iftikhar, Ahmad

Document Information

Project Acronym & Full Title	5G-TIMBER Secure 5G-Enabled Twin Transition for Europe's Timber Industry Sector
Project Number	101058505
Type of Action	HORIZON Innovation Actions
Topic	HORIZON-CL4-2021-TWIN-TRANSITION-01-08
Project start date/duration	01 June 2022 / 36 months
Work Package/ Result	WP1 / D1.5 initial 5G-TIMBER data flow architecture for the circular material and production process management
Result Type/ Dissemination level	Report/ PU
Lead partner	Tieto Sweden AB
Due date	M6
Responsible Author/Authors	Stefan Westlund (Tieto-SE)/Ann-Louise Hallding, David Buffoni (Tieto-SE) /Petteri Mussalo, Iftikhar Ahmad, (Tieto-FI)/ Muhammad Mahtab Alam, Yannick Le Moullec (TalTech), Tord Kaasa (Jotne), Francesco Berardinucci (Polimi), Marco Rossoni (Polimi), Marcello Urgo (Polimi)
Project URL	https://www.5g-timber.eu/

Review Process Detail

Date	Responsible	Action
19 October 2022	Jari Halme (VTT)	Quality Assurance and Risk Manager (QARM) sends a request for the deliverable to Deliverable Responsible Person (DRP). Request is also sent to the WP Leader and the Reviewers.
26 October 2022	Iftikhar Ahmad (Tieto-FI)	DRP initially freezes the Table of Contents (ToC) in agreement with the WP Leader.
2 November 2022	Iftikhar Ahmad (Tieto-FI)	DRP sends the draft for the named reviewers (2 or 3 persons: WP leader unless DRP & deliverable consumer(s)). Includes both Word version and PDF version. PDF version is needed to check e.g., broken cross-references, fonts, layout, etc.
12 December 2022	POLIMI (Marcello Urgo), Harmet (Mari Emmus) and Jotne (Tord Kaasa)	Reviewers sends the deliverable back for revision. A discussion to be follow in January on Data Management.
06 January 2023	Iftikhar Ahmad (Tieto-FI)	DRP addresses the comments in two weeks.
17 January 2023	POLIMI (Marcello Urgo), Harmet (Mari Emmus) and Jotne (Tord Kaasa)	Reviewers send the deliverable back with additional comments after internal discussions.
31 January 2023	Iftikhar Ahmad (Tieto-FI)	DRP addresses the comments on the UC figures based on the discussion in two weeks.
15 February 2023	POLIMI (Marcello Urgo), Harmet (Mari Emmus) and Jotne (Tord Kaasa)	Re-review: reviewers send the deliverable back with additional comments.
16 February 2023	Iftikhar Ahmad (Tieto-FI)	DPR addresses the comments in one week
24 February 2023	POLIMI (Marcello Urgo), Harmet (Mari Emmus) and Jotne (Tord Kaasa)	Reviewers approves the deliverable and send it coordinator/technical manager. QRM is also informed
27 February 2023	Jari Halme (VTT)	Quality review done by QRM

28 February 2023	Anet Tammets (TalTech)	Project Coordinator approves and saves the deliverable as the final version. Project Coordinator submits he deliverable to the European Commission.

History of changes

Version	Issue Date	% Complete	Changes	Contributor(s)
V0.1	10/10/2022	1	Initial Deliverable Structure	Iftikhar Ahmad (Tieto-FI)
V0.2	14/10/2022	10	Initial ToC	Stefan Westlund (Tieto-SE) /Iftikhar Ahmad (Tieto-FI)
V0.3	21/11/2022	35	document updated with UCs	Iftikhar Ahmad (Tieto-FI)/ Francesco Berardinucci, Marco RossoniMarcello Urgo (Polimi),
V0.4	22/11/2022	40	details added about use-cases, digital twin, and data management	Tord Kaasa (Jotne)/ Iftikhar Ahmad (Tieto-FI)/ Francesco Berardinucci, Marco RossoniMarcello Urgo (Polimi)
V0.5	29/11/2022	45	use cases related communication added	Tord Kaasa (Jotne)/ Iftikhar Ahmad (Tieto-FI)/ Francesco Berardinucci, Marco RossoniMarcello Urgo (Polimi)
V0.6	04/01/2023	55	updated use case	Petteri Mussalo and Iftikhar Ahmar (Tieto-FI)
V0.7	13/01/2023	65	updated use case	David Buffoni (Tieto-FI)
V0.8	26/01/2023	75	Use cases and Data management diagram	Petteri Mussalo (Tieto-FI) and Iftikhar Ahmar (Tieto-FI)

V0.9	10/02/2023	85	Review and feedback	Muhammad Mahtab Alam and Yannick Le Moullec (TalTech)
V1.0	17/02/2023	95	Data management	Petteri Mussalo (Tieto-FI)
V1.1	27/02/2023	95	Updated Version Based On Re-Review Comments	Petteri Mussalo (Tieto-FI)
V1.2	28/02/2023	100	Quality Review	Jari Halme (VTT)
V1.3	28/02/2023	100	Version for Delivery	Anet Tammets (TalTech)

Disclaimer

The content of this document reflects only the author's view. Neither the European Commission nor the HaDEA are responsible for any use that may be made of the information it contains.

While the information contained in the documents is believed to be accurate, the authors(s) or any other participant in the 5G-TIMBER consortium make no warranty of any kind concerning this material, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose.

Neither the 5G-TIMBER Consortium nor any of its members, their officers, employees, or agents shall be responsible or liable in negligence or otherwise howsoever in respect of any inaccuracy or omission therein.

Without derogating from the generality of the preceding neither the 5G-TIMBER Consortium nor any of its members, their officers, employees, or agents shall be liable for any direct or indirect, or consequential loss or damage caused by or arising from any information advice or inaccuracy or omission herein.

Copyright

© 5G-TIMBER Consortium. This deliverable contains original, unpublished work except where indicated otherwise. Acknowledgment of previously published material and of the work of others has been made through appropriate citation, quotation, or both. Reproduction is authorised the source is acknowledged.

Table of Contents

List of Figures

List of Tables

Glossary of terms and abbreviations

Abbreviation / Term	Description
5G	5th Generation
AR	Augmented Reality
CA	Consortium Agreement
DAG	Directed Acyclic Graph
DoA	Description of Action
DRP	Deliverable Responsible Person
DT	Digital Twin
EC	European Commission
eMBB	Enhanced Mobile Broadband
EU	European Union
GA	Grant Agreement
GDPR	General Data Protection Regulation
GeA	General Assembly
KPI	Key Performance Indicator
LCS	Location based services
mMTC	Massive Machine-Type Communications

MDM	Master Data Management
PI	Performance Indicator
PT	Project Target
QRM	Quality Assurance & Risk Manager
SON	Self-Organizing Networks
TIM	Technical and Innovation Manager
UCs	Use-Cases
UI	User Interface
URLLC	Ultra-Reliable Low Latency Communications
WVC	Wood Value Chain
ZSM	Zero touch network Service Management

1. Executive Summary

This document is the deliverable "D1.5 – 5G-TIMBER data flow architecture for the circular material and production process management" of the Horizon Europe Innovation Action project "5G-TIMBER - Secure 5G-Enabled Twin Transition for Europe's TIMBER Industry Sector" (hereinafter also referred to as "5G-TIMBER", project reference: 101058505).

This deliverable defines the principles for data exchange, with respect to different flows, i.e., (a) material (b) digital twin assisted production, and (c) production monitoring, based on openness (open standards and opensource), trust, semantic interoperability, and connectivity limitations. Considering the different case categories with all use-cases and the involved actors, principles will be drafted with respect to parts/materials/processing information within the whole life cycle by the example of wood processing - from the sawmill to reusable wooden construction elements. It will first address the definition of requirements to support the implementation and integration of technical solutions in WP2, and WP3. Successively, it will also address the validation of these principles in the implementation of the case pilots (WP4) and use the results to support standardization (T5.5) and dissemination (T6.4) activities.

The intended audience of D1.5 includes all members of the 5G-TIMBER consortium and the Project Officer.

The final version of this deliverable, due on November 30th 2022, has been submitted with a delay on February 28th 2023, to take into consideration the technical and operative details related to different use-cases from Harmet and Hekotek along with the current status of the availability of 5G technologies, hardware components and related capabilities, emerged in the continuous advancement in the data sharing architecture.

2. Introduction

Data is the basic building block of the new economy and play a central role in the digital technology. Access to large amounts of data plays a vital role in the understanding of digital economy. Therefore, it is important to have a proper data sharing architecture that shares it with all the stakeholders. Data flow will add the wood value chain from forest to the sawmills to building blocks of wooden houses and throughout the lifespan of wooden structure. It will also contribute to the project objectives by sharing wood quality information to sawmills, whereas sawmills output with wooden house producer. Equipped with the data information the wooden house maintenance will also be benefited.

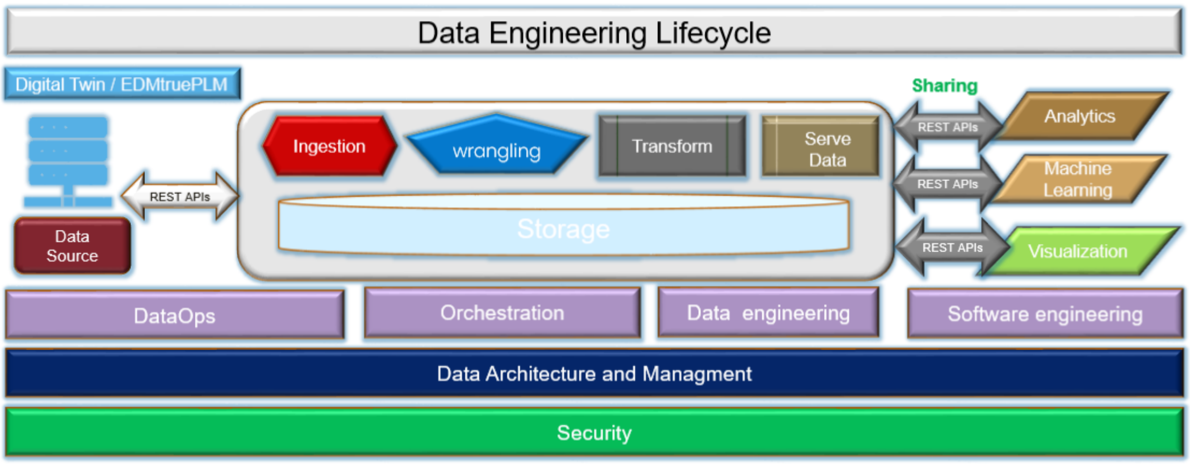


Figure 1: Data engineering lifecycle

Large information technological companies are working in data sharing and management for some time and there are several frameworks, tools and systems that are in development from the last one decade. [MS Azure](), [Google](), and [Amazon]() are offering Managed cloud services for the data collection from data sources to storage and access for different usage. **Data engineering** is used to develop, implement, and maintain a framework that takes in raw data from different sources for ingestion and produce high-quality information (data) by data-wrangling as consistent information. It supports different usages for example analysis, monitoring, visualization, and machine learning etc. It (data engineering) also includes the security, data management, DataOps, data architecture, orchestration, software engineering and data lifecycle management. The data source can be physical system, or a Digital Twin (DT) as shown in **Figure 1**. A DT is a set of virtual information mimicking a physical system/asset. Thus, the DT requires

data from the physical system to function. A full definition of DTs can be found in the AIAA Position Paper on the subject [23].
Figure 1 shows the data engineering life cycle management for **data-sharing-architecture-solution** from data generation (source) to utilization including wrangling (cleaning, removing noise or privacy related information) and transformation. Later data will be served for different services. **REST APIs** will be used for the information exchange from edge to server. They (REST APIs) will be defined further as use-cases evolve in different phases of the project as defined in the project application.
In the 5G-TIMBER project, data engineering and sharing will have the following attributes.

- **Data interoperability:** for the exchange of data with different stakeholders, it should be based on state-of-the-art developments and standardization for the Timber industry. This implies a secure environment & anonymization of the data for the processing of it including personal data.
- **Data as a service** ([EU draft Data Act](#)) **[6]**: As EU is envisioning to share the product data with end-users, it is essential to collect & store data of different components and manufacturing for later usages.
- Support for **heterogeneous data types** and formats (encodings)
- **Data strategy** helps how to use data and for which purpose. In the project, data access profiles will be created in each use-case for sharing information.
- **Data Governance model** [7]: here we follow the **EU draft data governance model** as listed below,
    - Common data space for sharing data.
    - Data altruism: meaning people or organization donating data for the public good. This is a part of the [EU Data Governance Act](#). For altruistic purposes of data sharing, the type of data and range of data subjects should clearly be specified.
    - Altruistic data sharing: Only public and non-profitable organization voluntarily sharing data under a rulebook.
- **Data modelling** covers the intra and extra relationships of data items and how they co-relate with each other. It will be helpful for merger and demerger of data streams. For example, how to combine temperature data with humidity data. Furthermore, how sensor data archived in a digital twin relates to simulated data.

- **Data architecture:** will be cost effective, agile (for element insertion and deletion), real time data access & update, inherit support of security and privacy, scalability, manageability, and fault tolerance.

The main goal of the **data-sharing-architecture** is to achieve the objectives defined in the project proposal and agreed in the GA and Deliverable 1.3, as listed below:

- Increase wood-based materials recycling by 50% by digital twin transition and data driven solutions.
- Increase manufacturing productivity by 15% with twin transition and 5G-supported manufacturing.
- Reach 99% of the work done in the factory (vs. 85% today) by using digital twin transition, 5G connectivity and data driven solution.
- Reduce on-site work by 10% with better AR/XR support combined data driven solution and proven in different use-cases.
- Reduce product nonconformities by 10% by using data driven solution along with digital twin transition along with AR/VR solution.
- Increase the safety of workers in wooden houses production and onsite assembling by using 5G based location services along with machine learning to provide support for safety of workers.

Above targets will be achieved by efficient sharing of data among different stack holders. An efficient data sharing architecture is required for the uses-cases along the categories are listed below. For more details check **Deliverable 1.1**.

- **UC1 category**: Data driven sawmill woodworking machines (UC1.1, UC1.2, UC1.3)
- **UC2 category**: Modular wood-house factory (UC2.1, UC2.2, UC2.3)
- **UC3 category**: Construction and renovation with wooden elements, valorization of composite waste (UC3.1, UC3.2, UC3.3)

Given the above, the objective of **T1.3** is to define the architecture of data sharing solution for above UCs while keeping in mind the overall PTs. The task T1.3 span from M2 to M24. This deliverable (D1.5) is due on M6. Its summaries the work done from M1 to M4, focusing on the initial definitions of the architecture of data sharing solution. The final version therefore will be documented in D1.6 due on M24.

Taking the above as a starting point, T1.3 focuses on defining the architecture of data sharing solution among the feedback material processing chain and related use-cases:

1. Proposing different data sharing architectures matching the needs and requirements of different use-cases.
2. The architecture will be based on openness, trust, semantic interoperability and the connectivity limitation of different devices and networks.
3. Data models will include the whole lifecycle of wood processing from sawmill to reusable wooden construction material.
4. Efficient data sharing and data streams aggregation.

There are different options (opensource, open standards or combined opensource and open standards) for data flow architecture for the IoT solution, we will discuss three options having different advantages:

- Opensource and support network layer (5G-connectivity): Intel® Smart Edge Open: is an opensource edge computing software toolkit for building platforms optimized for edge, it has services like 5G-RAN, 5G core, AI and media processing, security workloads and data sharing [1].
- Eclipse Arrowhead: is an opensource community driven project consisting of systems and services for real time data handling, data management (ingestion, storage and sharing) and system security. It also supports automation system engineering, scalability of automation of systems [2]
- Proprietary solutions: MS Azure IoT hub: is a commercial (proprietary) solution for data ingestion, cleaning, aggregation, sharing, along with AI and security services [3].
- The virtual data, (CAD, SIM, PLM, logistics), will feed into a DT archiving solution. Such a solution built on open available international standards, governed by a standards development organization (SDO), such as ISO 10303 is preferred to facilitate interoperable and accessible data.

2.1. Mapping 5G-TIMBER Outputs

This section maps the components (elements) described in the 5G-TIMBER's Description of Action (DoA)/Grant Agreement (GA), i.e., the description of the deliverable and task description in the DoA/GA, against the chapters included in this deliverable and the justification of the work conducted in the project. See Table 1.

Table 1: Adherence to 5G-TIMBER's GA Deliverable & Tasks Descriptions

5G-TIMBER GA Element	5G-TIMBER GA Element Outline	Respective Document Chapter(s)	Justification
DELIVERABLE			
D1.5 Initial definition	5G-TIMBER data flow architecture for the circular material and production process management	Chapter 3	Defined the requirements and data flow architecture
TASKS			
T1.3 Architectural planning of data sharing solution among the feedbacked material processing chain and related UCs	The principles for data exchange will be defined, with respect to different flows, i.e., (a) material (b) DT assisted production, and (c) production monitoring, based on openness, trust, semantic interoperability and connectivity limitations.	CHAPTER 3, 4, 5	DATA FLOW ARCHITECTURE AND RELATED TECHNOLOGIES
	Considering the different case pilots and the involved actors, principles will be drafted with respect to parts/materials/processing information within the whole lifecycle by the example of wood processing - from the sawmill to reusable wooden construction elements.	CHAPTER 3, 4, 5 AND 6	DEFINE ARCHITECTURE IN CHAPTER 3 AND USE CASE RELATED TOPICS ARE CONSIDERED IN CHAPTER 4, 5, AND 6
	The tasks will first address the definition of requirements to support the implementation and integration of technical solutions (WP2, WP3). Successively, it will also address the validation of these principles in the implementation of the case	CHAPTER 4, 5 AND 6	DISCUSSED IN CHAPTER 4, 5 AND 6

	pilots (WP4) and use the results to support standardization (T5.5) and dissemination (T6.4) activities.	

2.2. Deliverable Overview and Report Structure

This deliverable is organized as follows:

- Chapter 3 presents the architecture of data sharing solution along with related services and frameworks
- Chapter 4 describes the data sharing architecture according to the transport layer for UC1 Category
- Chapter 5 describes the data sharing architecture according to the transport layer for UC2 Category
- Chapter 6 discusses the UC3 category and transport layer and data flow architecture
- Chapter 7 system requirements
- Chapter 8 concludes the deliverable
- Appendix I Definitions
- Appendix II: An example of JSON information for different use-cases

2.3. Other project outputs

This section describes the interdependencies with the other project tasks and activities.

a) Interdependencies between the task(s) of this deliverable and the other WPs (inter-WP dependencies):

- WP2: the federation, integration and orchestration of the technological developments that will be carried out in WP2 will consider the details of the *initial* 5G-TIMBER data flow architecture for the circular material and production process management defined in T1.3/D1.5 to drive and tune the technical work. The experience that will be gained in WP2 (mainly via the cycles of lab trials) will feedback to WP1 to elaborate the *final* definitions of the 5G-TIMBER data flow architecture for the circular material and production process management in T1.3/D1.6.

- WP3: the details of the *initial* 5G-TIMBER data flow architecture for the circular material and production process management in T1.3/D1.5 will further propagate to WP3 where they will be considered for carrying out the system level and operational readiness for UC pilots' activities. The experience that will be gained in WP3 (mainly via the cycles of lab and field trials) will feedback to WP1 to elaborate the *final* 5G-TIMBER data flow architecture for the circular material and production process management in T1.3/D1.6.
- WP4: the deployment of 5G-TIMBER applications into demonstration sites in WP4 will consist of the cycles of field trials during which the KPIs will be measured and evaluated; the experience that will be gained during the first months of these field trials will feedback to WP1 to freeze the *final* 5G-TIMBER data flow architecture for the circular material and production process management in T1.3/D1.6.

b) Interdependencies between the task(s) of this deliverable and the other similar WP Tasks (intra-WP dependencies):

- The *initial* UCs and general requirements defined in T1.1/D1.1 are the foundation for refining the requirements and defining the *initial* KPIs in T1.3/D1.5.
- T1.3/D1.5 will elaborate the *initial* version of the architecture for data sharing and exchange of feedbacked material processing chain and related UCs; this architecture will have to have the provisions for measuring, storing, and exchanging the data needed for collecting and analyzing the *initial* KPIs defined in T1.1/D1.1.
- T1.3/D1.5 will provide foundation for the data pre-processing and transformation for different use cases (WP2 & WP3).
- T1.3/D1.5 will also provide guideline for the data privacy and security for different use-cases.
- T1.3/D1.5 will provide foundation for cross-enterprise data exchange in task T3.1.

3. Architecture of Data sharing

A data sharing architecture solution should have the following attributes:

- Security & Privacy: Security and privacy are the basic building blocks and should always be implemented in the default setting or in the architecture. We will focus on zero-trust, end-to-end security. National and EU level regulation compliance such as EU-Data and EU-AI Act will be in the focus in addition to security and privacy.
- Interoperability and meta-data management: Middle-wares and applications should be data (semantic) interoperable in 5G-TIMBER industry.
- Common component: consist of IoT device firmware to connected pipe (protocols), Serializable component & format, storage, message brokers, compute, Data streams transform (merger/enrich) and consumption. Furthermore, the DT repository for sharing, archiving, and exchange of data.
- Scalability: It should handle significant variation in data streaming and variation from numbers of users requesting the data.
- Robustness and stability: Anything that can fail will fail. No system is failproof. It takes lots of effort to make a failproof system. Rather it is cost-effective to make risk tolerant with risk measurement in different risk perspectives. But the architecture must be fault tolerant and can cope with failures. There should be a recovery time objective with a recovery point.
- Leadership or ownership: Make the decision to upscale or downscale resources efficiently and prioritize between the amount of data to handle or number of clients to handle. In addition to project/service/ecosystem leadership the data ownership and data sharing ownership along with leadership are essential. Data producer (Controller) are the data owners if not agreed otherwise.
- Couple systems: Think bigger and connect many systems with different capabilities and needs.
- Multi-door (reversible) decision: Streams from data source can merge and will generate enriched stream. Architecture must be able to merge, demerge and again merge the data streams to create a variety of data-streams for different data consumers.
- DataOps: Data driven management of infrastructure spending. It is a superset of scalability with clear objectives such as cost-efficiency, profitability, performance etc.

Section 3.1 describe the data management and data sharing framework whereas section 3.3 explains the network protocols. Section 3.4 details the data model for the data sharing architecture. Serialization & deserialization is the crucial part of a data sharing architecture is described in Section 3.5. And section 3.6 list different data schema types for the different communication needs.

3.1. Data management and sharing framework architecture

Edge is the important bridge between the real & digital world. IoT devices are referred as edge devices in this document. At the Edge, the digital (or virtual) world meets the physical system (or environment). Edge conveys the physical system information to the virtual (or software) world. Edge devices have different capabilities and can collect different values from physical world. In the 5G-TIMBER project it will also collect the location information from multiple sources such as 5G-network, Global positioning satellite, WIFI, and by location tags. But in this project, 5G based location will be preferred.

Different types of IoT devices will collect data and route it to their respective IoT-hub. IoT-hub is the on premises first data collection point where IoT devices using similar connection types such as 5G or WIFI to send data. Data will be verified for quality, security, and privacy at the IoT-hub. Later IoT-hub will route the data to the gateway that is connected to the Internet and will pass the data to private/public cloud according to the network configuration.

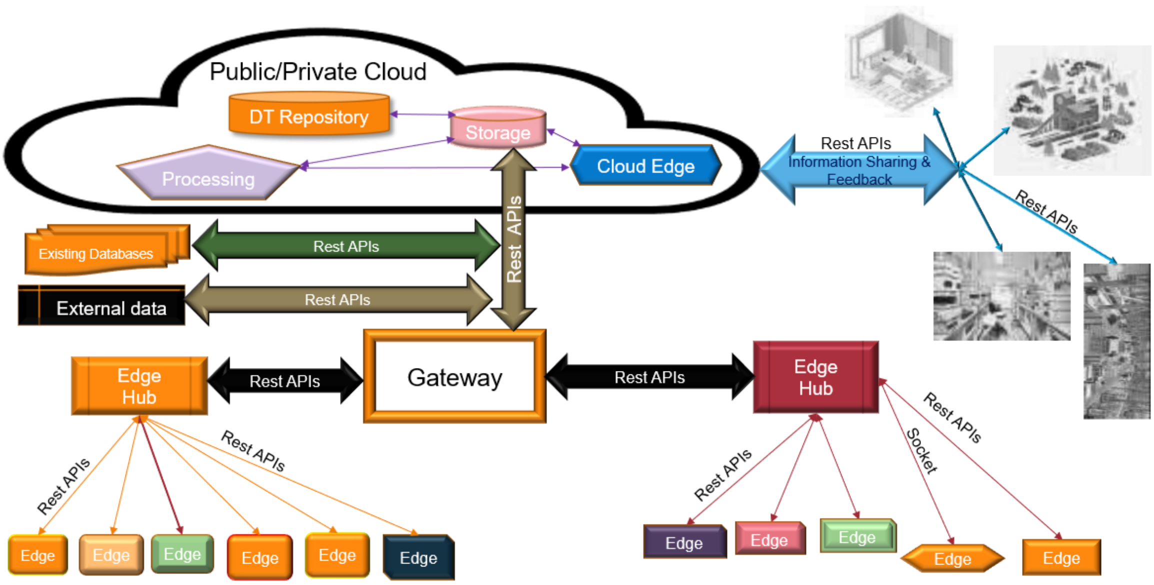


Figure 2: Data propagation architecture

Data propagation from IoT devices to Edge-Hub and later to gateway architecture is shown in **Figure 2**. It also shows the external data that can consist of Supervisory Control and Data Acquisition (SCADA) systems, old legacy databases or external data sources such as JSON, CSV files or etc. Information sharing with external partners (companies, vendors etc.) will use

REST APIs. REST APIs will be defined during the project execution. External partners can provide feedback and additional information across blue arrows in the **Figure 2**, that can be used for different purposes in the framework.
For learning purposes in the feedback information (data) flow there are usually two types of drift, one is data drift and other is the concept drift. Both are handled differently, and appropriate measures are taken to handle it. As shown in the **Figure** *3*, in case of concept drift a new machine learning (ML) model is trained from scratch with new data. Whereas in case of data drift machine learning model is re-trained (incrementally updated). New model training is demanding and is usually done on the server side, but incremental training can either be done on IoT-hub or gateway.

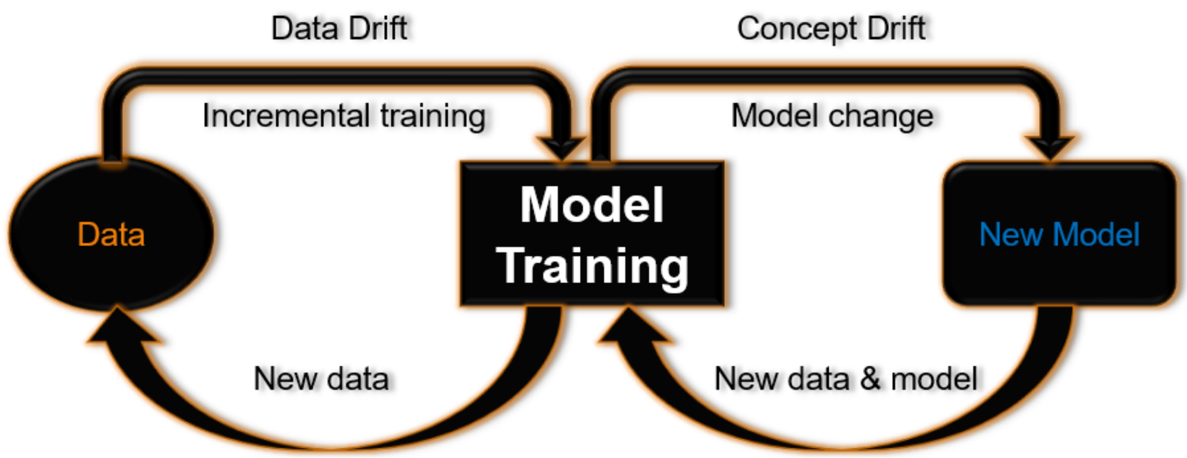


Figure 3: Data and concept drift

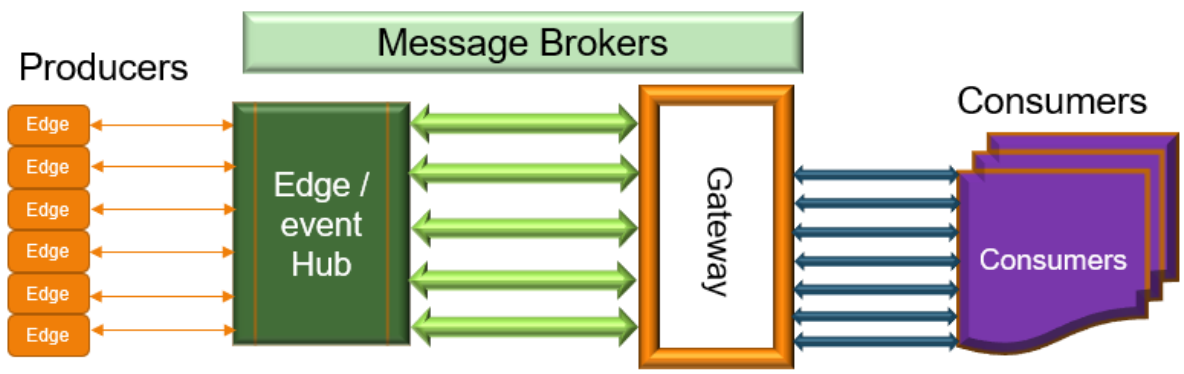


Figure 4: Framework architecture for producers and consumers

Different internal or external data consumers (visualization, analytics, machine learning etc.) can have different requirements, and these requirements can change dynamically. Therefore, message brokers are used to merge and demerge different data streams. In **Figure 4** architecture of data producers and consumer along with message broker is shown. Information (data streams) can be aggregated in different ways for more comprehensive analysis of different scenarios in a variety of use-cases. Event processing can be initiated from IoT devices, IoT hub, gateways or in the cloud. We propose IoT hub for one of the event generators for efficient message brokering.

3.2. Data management

Comprehensive data governance and data management are key elements for beneficial complex information system delivery. Data governance basic idea is to get valuable and high-quality data available for the user organizations. Data management is the operations and architectures how the data governance is utilized.

Regulations to be notified in 5G-TIMBER are European Union Data Act, AI Act and General Data Protection Regulation (GDPR). The EU Data Act is a key measure for making more data available for use in line with EU rules and values. The EU draft AI Act is a proposal to harmonize the rules on artificial intelligence used in EU. GDPR imposes obligations onto organizations collecting data related to people in the EU. It is notifiable, a video is seen also as personal data.

Data drives more and more the overall industrial development. Fast digitalisation increases the data growth rate and the available data sources. Modern and innovative technologies provide new possibilities to support both individuals and enterprise operations. Enterprises improve the operations [23]. The enterprises depend on the available data [24]. Developing technologies and paradigms, like I4.0 [25] emphasize data and analytics.

Data quality depends on data generation design and production processes [26]. Data quality has four categories: intrinsic, accessibility, contextual and representational data quality [27]. The notion of data quality should be well understood when designing information systems [26].

Data value is the difference between cost and benefit derived from the object [28]. Data value correlates with the data quality and the high rate increases the value [29].

Data valuation is a process when the data value for the data users increases. The data value depends on the context where the data is being used [30]. Data governance and data curation increase the data value [32]. The data valuation determines the asset suitable for task-dependent or task-independent assignments.

The 5G-TIMBER project lies on data from different sources with different kind of response and latency times, data formats, data streams, adaptability, and localization. Response times are required to be short to ensure the application fluent usability. Latency times are related to camera and live stream assisted use cases (Chapter 6.2) where the user is warned about the threatening danger in the risky sawmill environment. Different data formats from industrial devices, cameras, IoT devices, and more traditional information systems.

Because the aim is to develop solution to support use in different companies and environments (Chapter 2) the localization, adaptability and data governance are the starting points for successful development.

Ontologies enable the concept and terminology inside and between different industries. Different organisations are developing and providing the ontologies and terminologies.

Data governance (Figure 5) with reasonable meta data management and data management are essential methods to support complex systems to use the data in primary but also secondary purposes. Primary data use is the operational use. Secondary data use where the operational data is used to analytical, forecasting, statistical and administrative purposes.

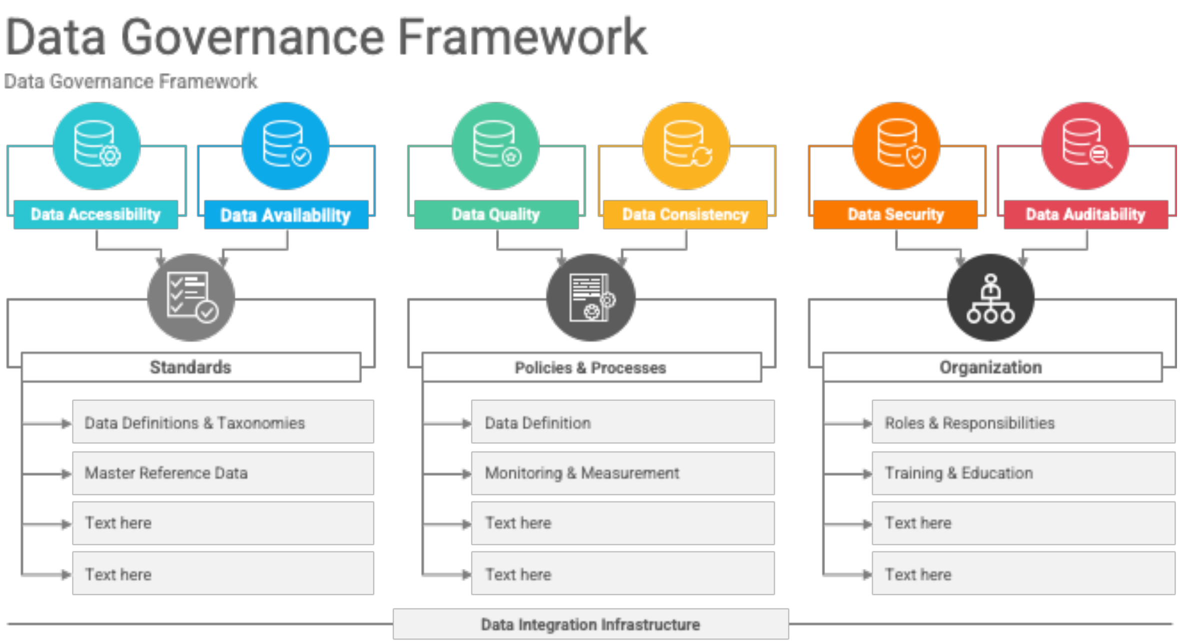


Figure 5: Conceptual presentation of Data Governance Framework

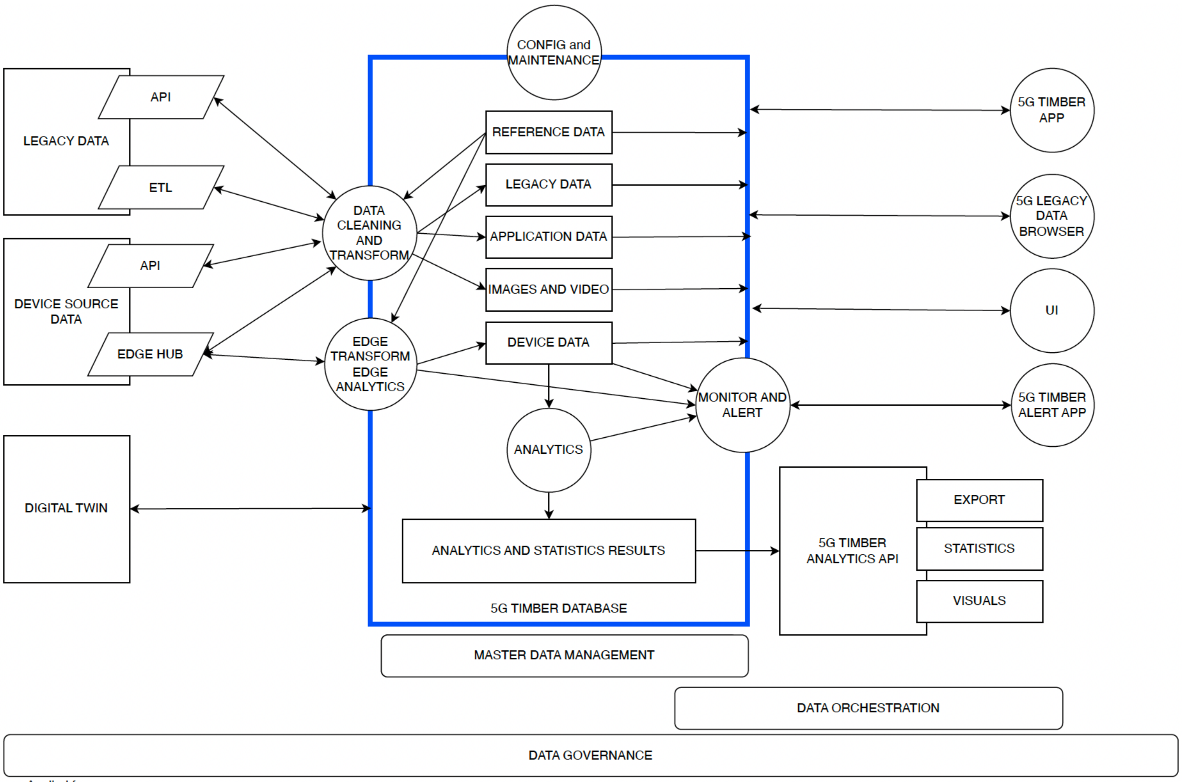
Adaptability is a key concept where the information system is productized and delivered to different users but also to different market areas. Adaptability requires services for setup configuration and providing services for different customers. IN 5G-TIMBER case configuration is required for different users, devices, variables but also analytical services and data export to external targets.

The data controller (Chapter 7.1) owns the data and is responsible for the data sharing following the regulations and enterprise practices. The data processors (Chapter 7.2) process the data on behalf of data controller.

Currently the software development is based on agile, iterative or hybrid methodologies. The current methods starting point is the changing operation environment, developing technology and changing user needs. What near the development is the final solution, that detailed the definitions are possible to have.

3.2.1. Conceptual presentation of 5G-TIMBER Data management

5G-TIMBER data management **Figure 6** is based on managerial operations (rounded squares), data elements (squares), interfaces (parallelograms), processes (circles) and data flows (arrows between the elements).



Applied from
https://learn.microsoft.com/en-us/azure/architecture/reference-architectures/data/profisee-master-data-management-data-factory
https://research.aimultiple.com/edge-analytics/

Figure 6: Conceptual presentation of 5G-TIMBER data management illustrating the managerial concepts (rounded squares), interfaces (parallelograms), data elements (squares), processes (circles) and data flows (arrows between the elements).

Data management cover managerial concepts (Chapter 3.2.2), data elements (Chapter 3.2.3), interfaces different kind (Chapter 3.2.4), processes (Chapter 3.2.5) and data flows between the elements (Chapter 3.2.6). The data trailing enables the data insert and change trailing (Chapter 3.2.7).

3.2.2. Managerial concepts

.

Managerial concepts are the policies and practices affecting to the data management and it is listed below.

Data governance	Data governance secures availability of high-quality information system data enabling the decision making
Data orchestration	Data orchestration is an automated process for bringing data together from multiple sources, standardizing it, and preparing it for data analysis
Digital twin	A digital twin is a virtual model designed to accurately reflect a physical object.
Master data management	Master data management (MDM) is a technology-enabled discipline in which business and IT work together to ensure the uniformity, accuracy, stewardship, semantic consistency, and accountability of the enterprise's official shared master data assets. Master data is the consistent and uniform set of identifiers and extended attributes that describes the core entities of the enterprise including customers, prospects, citizens, suppliers, sites, hierarchies, and chart of accounts.

3.2.3. Data elements

Data elements are always identified with the object identifiers identifying the objects, variables, and users.

Application data	Data used by the applications and transformed to application supported structures identifiers following the MDM guidelines
Analytics results and statistics	Analytics process results and statistical data transformed in easy-to-use format and structures following the MDM guidelines
Device data	Data collected by different devices and transformed to format supported by the

	applications and processes following the MDM guidelines
Device source data	Source data directly from devices
Export	Routine to export the data to external use I.e., research, statistics, and reporting
Image and video data	Images, photos, X-ray images, videos
Legacy data	Data from previous systems, I.e., wood sample and house drawing archive.
Location data	Object location in common and supported coordination system.
Reference data	Organized data of phenomena features, i.e. users, lists, units, transforms. Andrew White:" *While MDM provides solutions around key business concepts such as customer, product, location and employee, reference data management provides solutions around data that is not mission critical across an enterprise. This might include a list of codes and descriptions for software used in an organization, a list of cost centre codes and descriptions, etc."*
Statistics	Operational data will be organized for statistical use either in the database, data warehouse or data lake. Usually, the data is modelled to follow the functional entities.
Visuals	Data will be transformed to support data visualisation with high response performance.

3.2.4. Interfaces

The interfaces encapsulate the application logic and enable the application modules communicate with each other.

5G-TIMBER ANALYTICS API	API for the user applications to utilize the statistical, advanced analytics and machine learning results with low latency and high response rate.
API	Application programming interface, key element for applications and database interoperability
EDGE HUB	Collects the smart device dataflows and moves the data to transformation process.

REST	Representational state transfer sometimes mentioned as rest full APIs.

3.2.5. Processes

The processes execute different parts of the application, produce, use an delete the data.

5G-TIMBER alert app	Application alerting the users for different issues. Use case XX.XX alerts the user based on threatening situations by using video, object recognition and forecasting the object movements. Alerting app gets inputs from monitor and alert processes
5G-TIMBER APP	5G-TIMBER APP consists of several application user modules needed for the 5G TIMER realization.
5G-TIMBER LEGACY DATA BROWSER	5G-TIMBER LEGACY DATA BROWSER provides possibilities to browse legacy data and documents, which are not converted for the novel solution structures and are not compatible with the existing solution document standards or data structures and data items.
Analytics	Analytic process is responsible for analytics all kind from simple aggregated data to advanced analytics and complex machine learning modules. The produced data is stored to results database to enable fast response times.
Config and maintenance	The entire 5G-TIMBER information system configuration and maintenance module.
Data cleaning and transform	Process monitors the incoming data flow, detects the data anomalies, and cleans the data. System administrators are alerted about the anomalies. Process transforms and harmonizes the data to be 5G-TIMBER compatible.
IOT Transform and edge analytics	Process monitors the incoming data flow and detects the anomalies before sending the data to edge analytics.
Monitor and alert	Process monitors device analytics data streams; edge analytic stream and analytics results and alerts the user needed in the role and use case.

UI	5G-TIMBER end user application providing the end user role specific services. End user application supports the users in different roles and consists of various end user services.

3.2.6. Data flows

Data flow contents are depending on the underlying use cases and used devices. In all cases there are object identifiers, user identification or device identifier, and date and time details.

3.2.7. Data trailing

All human produced data fulfil the change/audit trail documentation (**Error! Reference source not found.**).

Table 2: Conceptual representation of data structure consisting of the audit trail realization.

```
Define DATA_OBJECT as
      OBJECTIDENTIFIER,    -- Object identifier, reference data
      VALUE_NUM,              -- Numeric OR
      VALUE_STR,             -- String OR
      VALUE_DATETIME,      -- Datetime
      STATUS,              -- Status 1 = valid, 2=corrected
      AUDIT_TIME,          -- Time, when status changed
      AUDIT_USER           -- User/device ID
      etc.
```

3.2.8. Data management summary

1. Short latency times lead to a situation where the cloud solution consists of risks related to the response and processing times and the processing should be based on the EDGE technology.
2. Common goals of the solution on hand leads to situations where the contents should be configurable. Configuration is a key to harmonize the data from different sources and from different devices.
3. Data quality is essential for near real time analytics and alerting. Continuous data flow monitoring and database scanning helps to improve the used data quality.
4. Industrial ontologies and terminologies are promising for ecosystem-based development. However, the industrial terminologies need more attention about their maturity and how they can be utilized. The lack of international ontologies set requirements for the system and data configurability. Anyhow, it is essential to follow the international

development and begin to utilize the ontologies and terminologies. The data configuration module should adapt in future to one or several ontologies.

3.3. Network protocols

Messaging over network is important and plays an important role for efficient data transfer, error recovery and other network related issues.
MQTT and AMQP run over TCP/IP. MQTT is basically designed for Internet-of-Things (IoT) whereas AMQP is for enterprise messaging. There are several other methods such as websocket, COAP, ZeroMQ, HTTP, HTTPS and data formats such as XML, comma separated values (CSV) and JSON.
We propose REST-APIs over HTTPS, due to secure communication. REST-APIs will be defined later as the project progress. JSON format will be used for serialization/deserialization of the data.

3.4. Data model

There are different options for data models and relationship between data elements. We propose the vectors of vectors for the data models. It is shown in **Figure 8**. In addition to the IoT data, data coming directly from the machines that will be part of the UCs, such as PLM (including sensor data), sim, CAD, etc., will be archived in a DT solution based on the open standard ISO 10303.

EDM Council: "A model that presents the declarative structure of data through a defined scope and perspective, and to a given level of abstraction.
"

Data Virtualization: It is related to query-pushdown regardless of data pipeline and storage for different operations. We propose an event driven publish and consumer paradigm. Data source will publish self-defined data with complete context information, without un-necessary details.

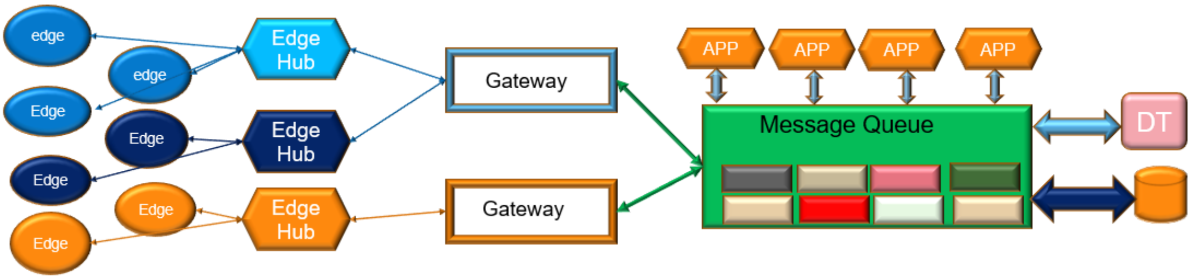


Figure 6 Publish and consumer data message queue

Stream Transformation: fusions of multiple IoT devices data to enrich the data stream and generate more meta-data. Streams directed acyclic graph (DAG) is an example where multiple IoT data streams are merged and create and enrich the streams. Stream transformation is shown in **Figure 6**

Message broker: handle different streams of data from a variety of data-sources and register the interest of different consumers and provide appropriate access to different streams for different usage. Message broker is also shown in **Figure 4**. Whereas producer (published) and consumer queue are shown in **Figure 6**.

Data wrangling: Data wrangling takes messy, malformed data and turns it into useful, clean harmonises and restructures data. This is generally a batch transformation process. This is usually done automatically and mostly remove garbage from the income data from data source due to connectivity (network) problem, sensor malfunction, intrusion, etc.

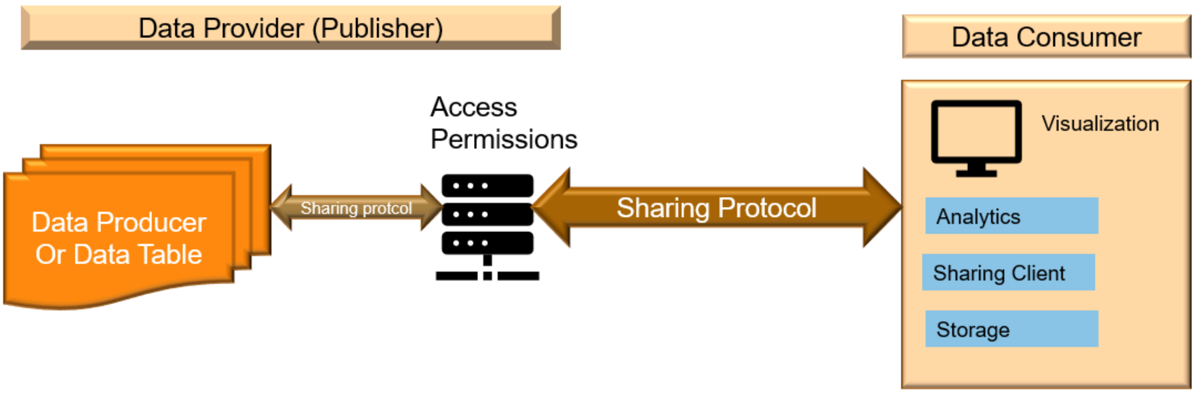


Figure 7: Secure data sharing

Secure data sharing is shown in **Figure 7**. Data provider or publisher also provide the access control of the data that is integrated part of the data sharing. Access control is managed by the data publisher. Ownership of data goes to the data publisher that will also control the access rights of the data.

Data schema:

Data schemas consist of vector of vectors with a depth of 1-M and on each level a width of 1-N. The first vector contains the header information that include permission, level of anonymization of the data. For sharing of the data with multiple parties, it is important that each element must contain mime-type and interpretation of the mime-type. For example: Value X represent size of wood, everyone is using the same analogy that X has the units in meters or any other units. If possible, it also includes all well-known variant mime-types such as Mime-type X-image/jpg or X-image/JPEG. Data schema is shown in **Figure 8**.

Figure 8: Data schema (Vector of vectors)

3.5. Serialization

XML

Extensible Markup Language (XML) was one of first standard for the data exchange at the beginning of Internet era. It is generally slow to deserialize and serialize for data engineering applications. XML is still used to interact with the exchange of data with legacy systems. Nowadays JSON has largely replaced XML for plain-text object serialization & deserialization.

JSON

JavaScript Object Notation (JSON) has emerged as the new standard for data exchange over APIs, and it has also become an extremely popular format for data storage. In the context of databases, the popularity of JSON has grown apace with the rise of MongoDB [19] and other document stores. Databases such as Snowflake, BigQuery [17], and SQL [18] Server also offer extensive native support, facilitating easy data exchange between applications, APIs, and database systems.

Google Protobuf [20]**:** Protocol buffers are Google's language-neutral, platform-neutral, extensible mechanism for serializing structured data. It is a kind of XML but in binary format.

Apache Avro**:** is a binary data serialization framework and data schemas can either be defined in JSON or IDL.

In 5G-TIMBER we will use human readable JSON format over HTTPS for data serialization and deserialization.

3.6. Data schema types

In 5G-TIMBER, there will be four different types of data schema, which are listed below:

1. **Normal schema** for normal data transfer, it is Vector of vector type schema shown in **Figure 8**. The normal Schema is a vector of vectors, and it is highly flexible for the data exchange between a variety of devices and services.
2. **Slim schema** for real time data transfer that needs high bandwidth and low latency. Although 5G-network is providing inherent support for it. Therefore, we are proposing a slim schema for real time communication.
3. **Alert schema** is a slim schema type by nature for real time communication, but it has an even smaller load and should carry different types of alerts that require urgent attention across the network.
4. **Debug schema** is a kind of normal schema with lots of extra information such as network and device health, workload, and other housekeeping activities. It helps to regulate and optimize network usage.

3.7. Current state-of-art of data sharing framework

Data sharing is an active area of research for industry and academia. Social media services have made it an important topic among big players such as Microsoft, Google, Meta (Facebook), Twitter, Amazon etc. They are providing several state-of-art services and frameworks with different features. Below is the list of few examples:

- MS Azure IoT hub [3]: Public cloud services, for the management of massive IoT devices and data collection.
- Amazon IoT hub [3]: Public cloud service, from Amazon for the management of IoT data.
- Manage cloud services from Azure [9], Google [10], Amazon [8]: Public cloud services for data management (collection, storage, transform and sharing (serving)).
- Casandra [11]: Apache Cassandra is an opensource NoSQL distributed data storage and for scalability and high availability platform.
- Apache Airflow [12]: is an opensource workflow management platform for data engineering pipelines.
- Eclipse Arrowhead [13]: is an opensource, efficiently integrate hardware and software of Internet of Things (IoT) devices platform. It provides IoT devices management, security data collection and storage.
- Elasticsearch [14] and Kibana [14]: are data storage and access for private and public cloud.
- Snowflake [15]: Managed software as a service to provide data warehouse, data lakes and it also decouple data storage and processing.

Selection of the appropriate tools or framework depends upon the three Vs (Volume, Velocity and Variety) of the data. With our current understanding of the three Vs, we propose Eclipse Arrowhead for IoT device and data management (sharing). Elasticsearch can be used for storing logs and textual data whereas Kibana can be used for visualization or dashboarding. Data controller are the data owner and data-processors are processing Information on behalf of data controllers. Data controller is responsible for the regulatory compliance and information sharing for data processors.

3.8. Research innovation

With the recent EU regulation proposal, EU is empowering the end-user to have access to the product and services data for different usage such as performance, experience, repair etc. This is a kind of data sharing with a wider set of users with diverse interest. Most of the manufacturers are not making the products end-to-end inhouse, and some parts and components are always outsourced. Therefore, it is important to get the relevant data from relevant parties and include it for any future information (data) sharing with the end user.

EU draft Data Act [6], draft Data Governance Act [7], combined with the EU draft AI Act require manufacturers to preserve some data for a longer period of time and share it with the product user and related parties. This opens-up a new area of research and gives a competitive edge to those who have these capabilities on time.

4. UC Category 1: Data driven sawmill and woodworking machine setup

UC Category 1 shown in **Figure 9**, it leverages 5G network technologies in combination with digital twin models, sensorisation, and edge/cloud computing to validate and assess their benefits in sawmills with respect to sawmill machinery and resulting wood products. It will combine data & material models along with digital twin and AI driven machine fault prediction to improve productivity and reduce the waste.

REST APIs details will be defined during the project and will be included in the next release of this document (D1.6).

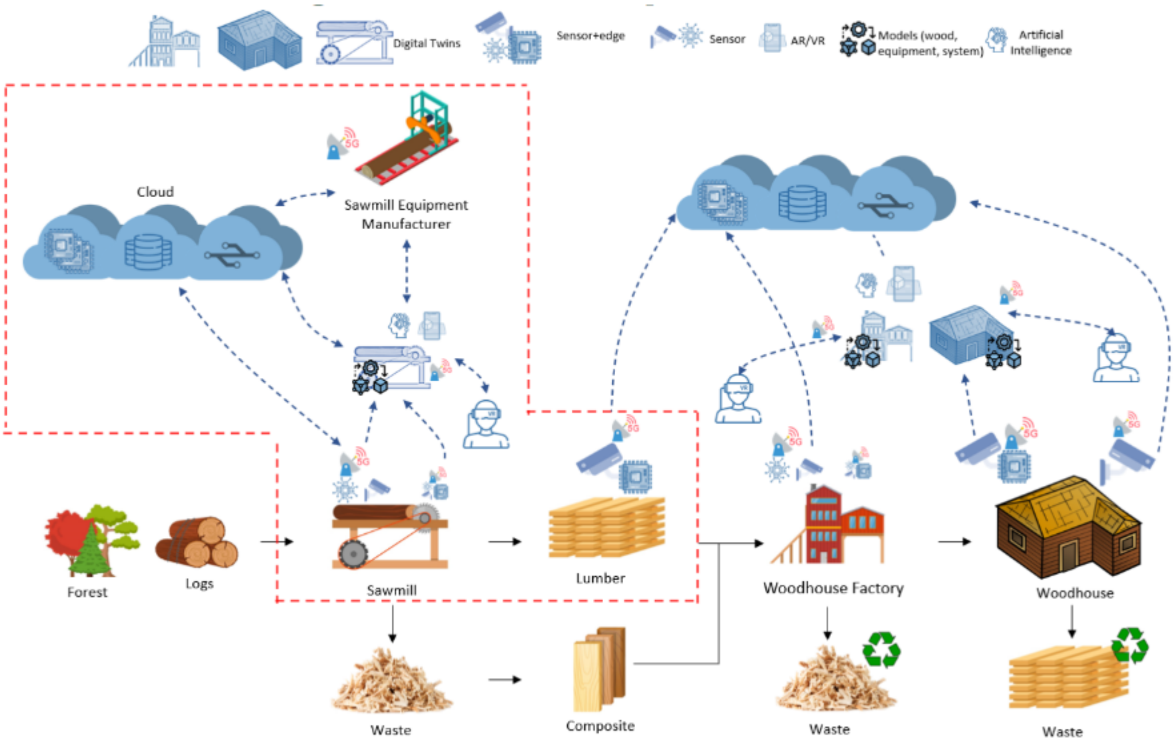


Figure 9: Use Case Category 1

Enhanced Mobile Broadband (eMBB) is used in Use Case 1.1 and Use Case 1.2 for efficient data transfer. Whereas Massive Machine-Type Communications (mMTC) is used in Use Case 1.1, 1.2 and 1.3 for massive data transfer. Ultra-Reliable Low Latency Communications (URLLC) is only used in Use Case 1.2 for digital twin assisted assembly. Self-Organizing Networks (SON) and Location based services are not used in Use Case Category 1.

4.0. UC Cat 1 - Data flow architecture and data schema

All IoT devices will use Rest APIs to transfer data from edge to the server and information will be encoded in UTF-8. The basic data flow architecture is shown in **Figure 2** in 3. For communication Rest APIs will be used using JSON over HTTPS to transfer data to the server and the information will be encoded in UTF-8. JSON information related fields are shown in Appendix II. REST APIs based communication will used UC1.1, UC1.2 and UC1.3 across the green arrow in the **Figure 10**, **Figure 11** and **Figure 12**.

Data Controller (Hekotek and Harmet) shown in **Figure 10**, **Figure 12** and **Figure 13** will ingest the data, wrangle it, and transform if required, later serve it to the data processor. Data processors shown in green blocks shown in **Figure** ***10***, **Figure** ***11*** and **Figure** ***12*** will process the data and later serve the generated output (meta-data) back to the data controllers. Data processors must also wrangle the output data before publishing it to data controllers to avoid transmitting any proprietary information. In the context of data architecture, 5G-connectivity shown in yellow colour will only be used for information transfer and will not process the information directly.

4.1. UC1.1 Data and material models driven lumber production and handling

As shown in Figure 11, Use Case 1.1 will focus on data and material models driven lumber production. Use Case 1.1 will be based on data and knowledge transfer. According to information material hardness testing based on image/X-RAY/CT will be conducted for optimal material selection. Data exchange will focus on images and forest information related to wood. Normal, alert and debug data schema will be used for exchange of information.

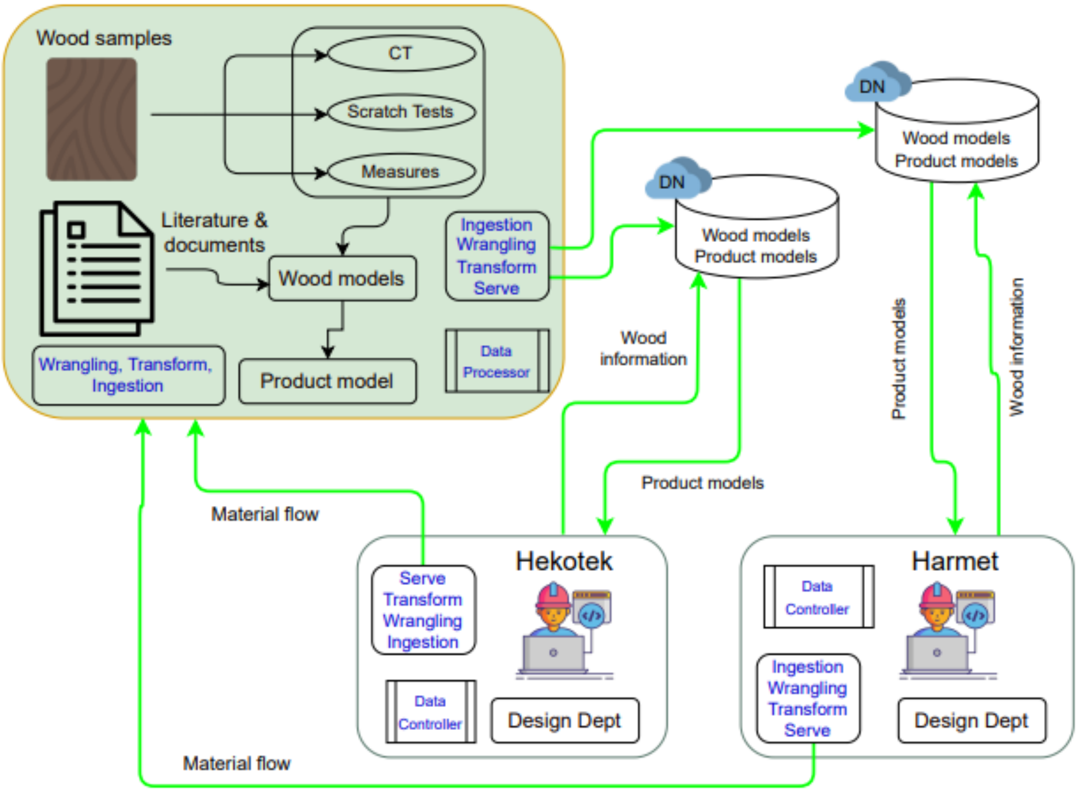


Figure 10: Use Case 1.1, Data and material models driven lumber production and handling

4.2. UC1.2 DTs assisted assembly, remote operation, and servicing of woodworking machinery

The objective of UC1.2 is to support the operation and control of the timber packing machine, leveraging a Digital Twin (DT) of the machine and Machine Learning (ML) models to process data acquired in the operating environment (i.e., sensor data, video streams, data streams).

As shown in **Figure** *11* high speed data transfer is required for demonstration via trials on how augmented reality software/tools (compliant with BIM tools of Hekotek) can be used for machine service operations and assembly processes. It requires two-way real-time communication. Normal, alert and debug data schema will be used for exchange of the information.

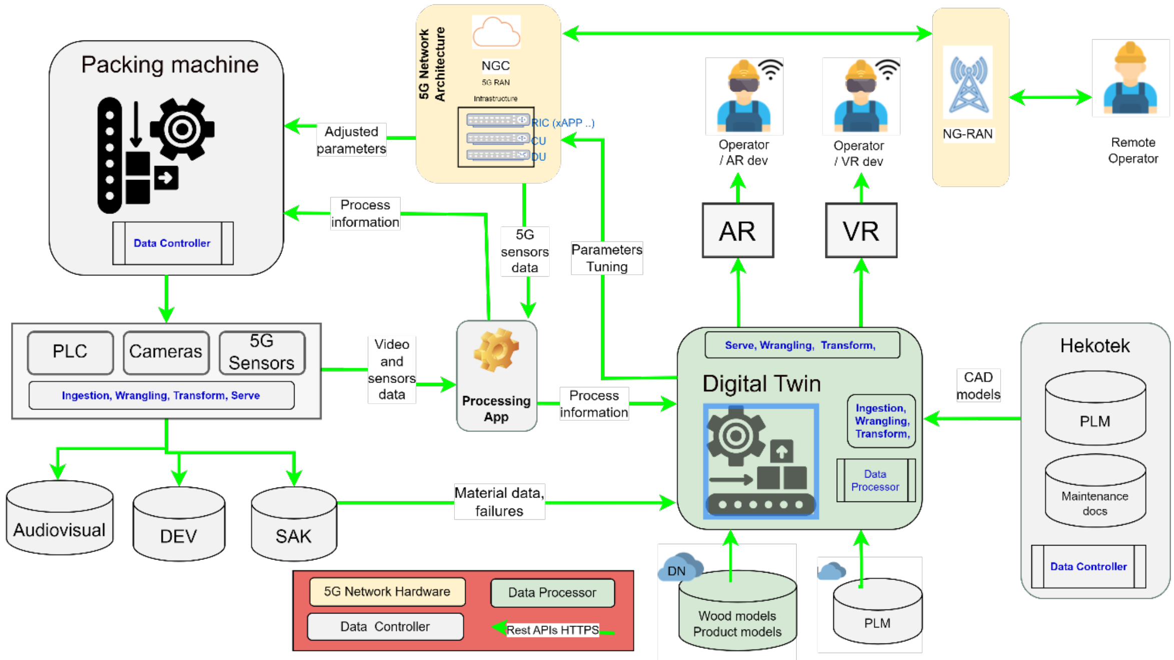


Figure 11: Use Case 1.2, Digital twin assisted assembly, remote operations, and servicing for wooden working machinery

4.3. UC1.3 Data and AI driven industrial machine's fault prediction and design support

As shown in **Figure 12**, Use Case 1.3 is intended to provide a general approach to support state-of-art maintenance in small-volume machinery production. Specifically, the use of IoT data for predictive maintenance and design enhancement of sawmill and woodworking machines will be the focus of the UC, demonstrating this in live operating conditions over the 5G.

The data schema of Use Case 1.3 is shown in appendix II. It uses low bandwidth & low latency in massive machine-type communications for data transfer and data will be saved on server side for machine learning and validation. Normal, alert and debug data schema will be used for the exchange of information. In this use case relevant data of a machine will be used for the predictive maintenance by using machine learning methods.

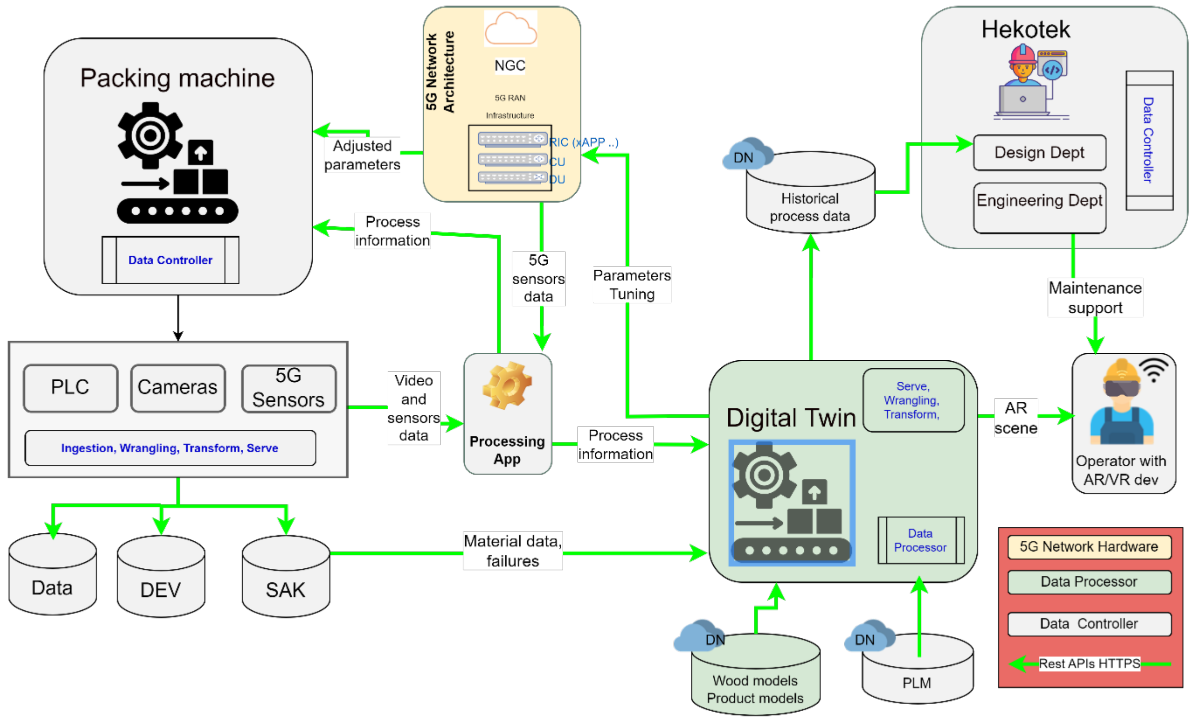


Figure 12: Use Case 1.3 Data and AI driven industrial fault prediction and design support

5. UC Category 2: Modular wood-house manufacturing

UC Category 2: as shown in **Figure 13**, these use cases focus on the manufacturing process of the wooden house module, from when the lumbers arrive at the woodhouse factory to when the modules leave the factory for the construction site. UC2 category addresses deploying a 5G private network within a production environment and how it can support integrating different applications and tools.

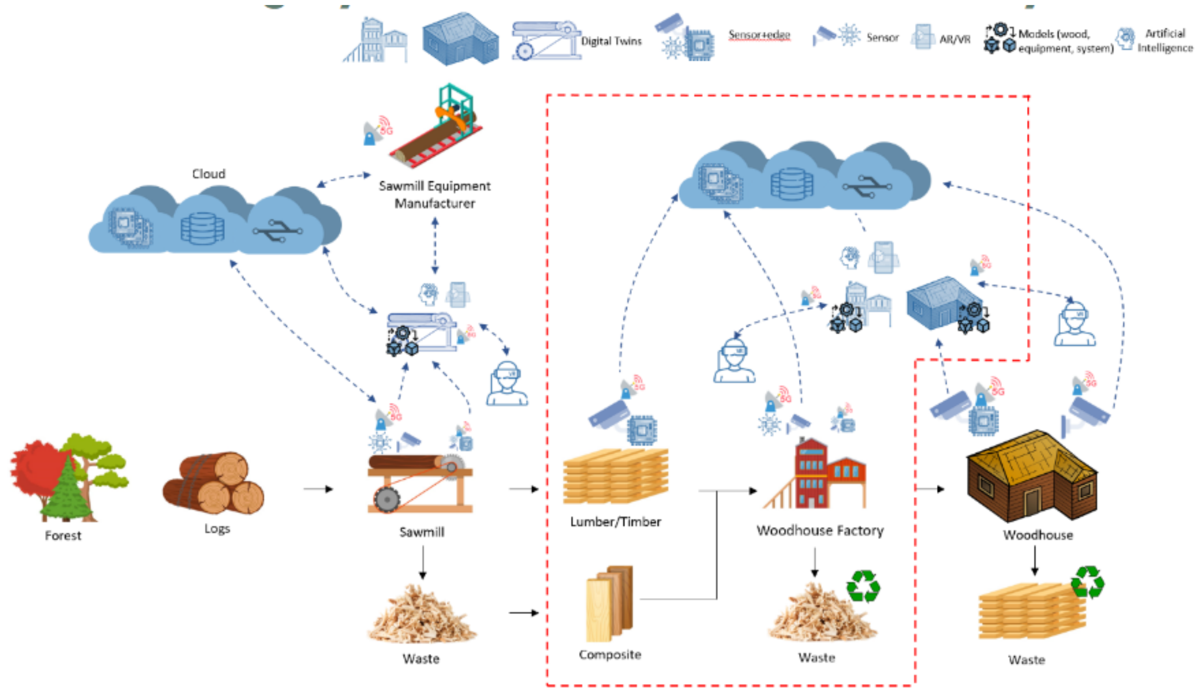


Figure 13: Use Case category 2

Use Case 2.1 uses Enhanced Mobile Broadband (eMBB) and Ultra-Reliable Low Latency Communications (URLLC) for communication and will also employ location-based services (LCS). Whereas in Use Case 2.2 Enhanced Mobile Broadband (eMBB) & Ultra-Reliable Low Latency Communications (URLLC) will be used. Use Case 2.3 Enhanced Mobile Broadband (eMBB) and Ultra-Reliable Low Latency Communications (URLLC) for communication. Self-Organizing Networks functionality will not be used in Use Case category 2.

5.1. UC Cat 2 - Data flow architecture and schema

For communication Rest APIs will be used using JSON over HTTPS to transfer data to the server and the information will be encoded in UTF-8. JSON information related fields are shown in Appendix II. REST APIs based communication will used UC2.1, UC2.2 and UC2.3 across the green lines in the **Figure 14**, **Figure 15** and **Figure 16**. It will be defined during the project execution and the details will be included in the next release of this document (D1.6).
In multi-organization settings, data-controllers own the data & publish it for processing. Whereas data-processors process the data on behalf of data-controllers for different applications or use-cases. Data-controllers in white blocks, shown in **Figure 14**, **Figure 15** and **Figure 16** will start from ingesting and wrangling the data, and transform data if required, later serve (publish) it to the data processor. Data-processors shown in green blocks will process the

data and later serve (publish) the generated output back to the data controllers. Data processors must also wrangle the output data before publishing It to data controllers to avoid transmitting any proprietary information. In the context of data architecture, 5G-connectivity shown in yellow colour in in **Figure 14**, **Figure 15** and **Figure 16** will only be used for information transfer and will not process the data directly.

5.2. UC2.1 Precise localisation for production processes optimization and human safety

As shown in **Figure 14**, precise localisation of human is important to avoid accidents in Woodhouse factories. Therefore, 5G localisation-based testing will be used alongside human detection in visual content, that would optimize material and human movement trajectories. It will therefore increase the production efficiency and improve the open space safety. Video data will be streamed to the IoT-hub, gateway to the server for visual content analysis according to General Data Protection Regulation (GDPR) compliance.

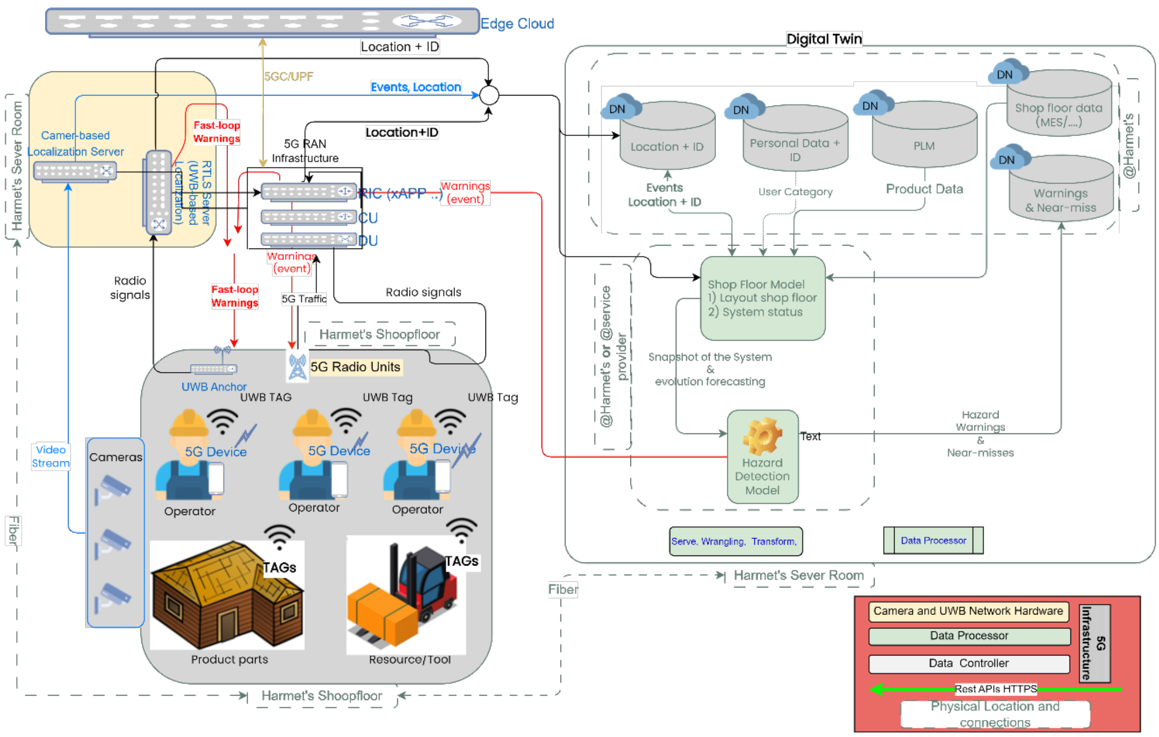


Figure 14: Use Case 2.1, Precise localization for production processes optimization and human safety

As shown in **Figure 14**, video data can be either processed on the edge or on the gateway/cloud-edge. We will use alert, normal, slim and debug data schema for the exchange of information. Generated alert will move on the red arrow to the destination using 5G connection.

5.3. UC2.2 Digital Twin, augmented reality, and Industry 5.0 UX assisted production supported by ZSM 5G data exchange

The UC2.2 aims to improve woodhouse assembly quality, efficiency, and safety by implementing Augmented Reality (AR) and Industry 5.0-assisted production.

As shown in **Figure 15** High-speed low latency two-way data transfer is required for this DT/AR Use Case 2.2. There will be 5-10 AR devices (smart glasses, tablets) across the factory that will be used for experimentation.

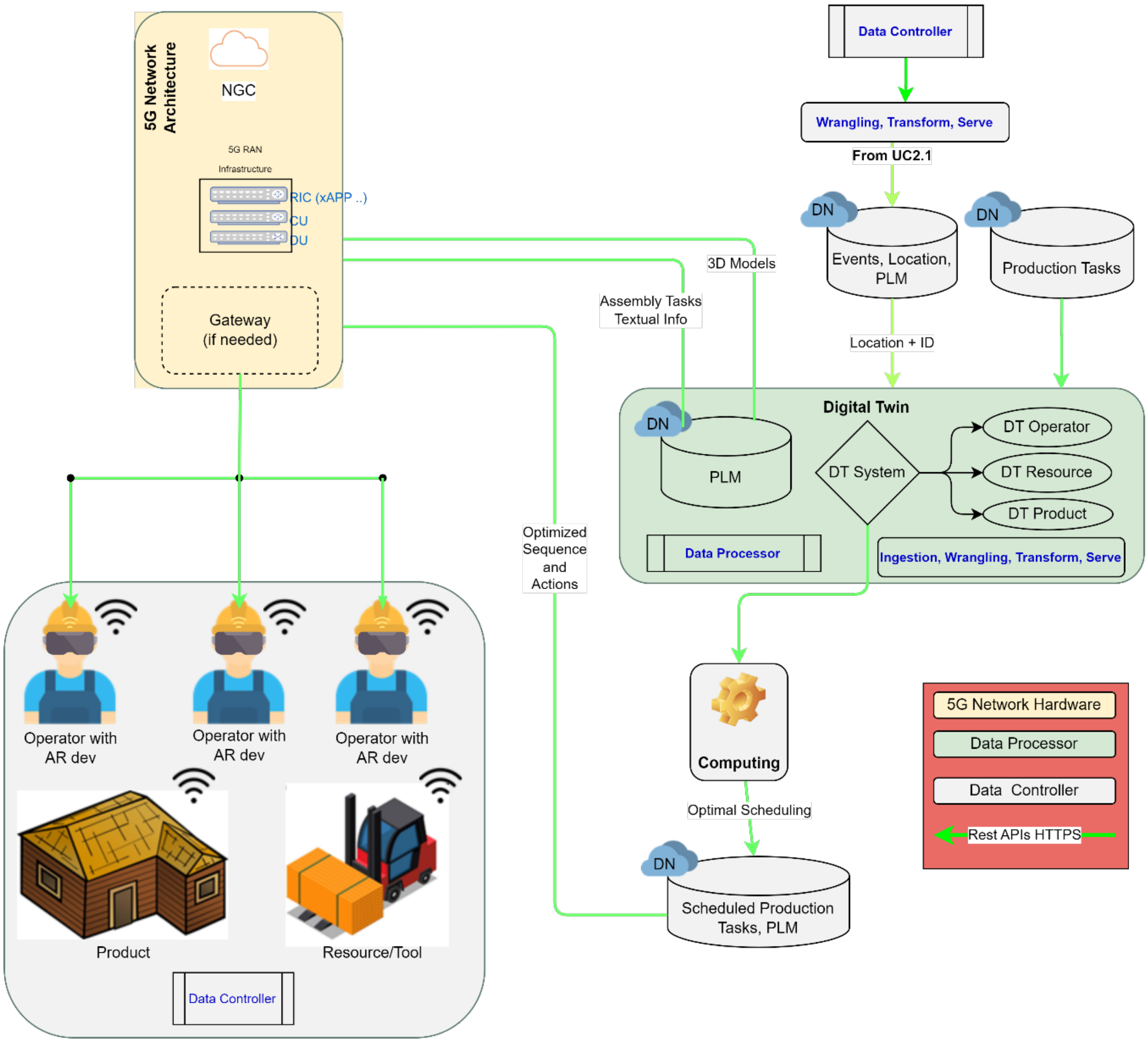


Figure 15: Use Case 2.2, DT, AR and Industry 5.0 UX assisted production supported by ZSM 5G data exchange.

In Use Case 2.2, we will use alert, normal, slim and debug data schema for the exchange of information.

5.4. UC2.3 Data-driven design to production cycle optimisation

As shown in Figure 16, UC2.3 aims to shorten the design to the production cycle and provide automated feedback of discovered non-compliances to initial design and manufacturing steps.

Massive machine-type communications & Ultra-reliable low latency communications will be used in UC2.3 for data exchange. Use Case 2.3 will also use data from other Use Case.

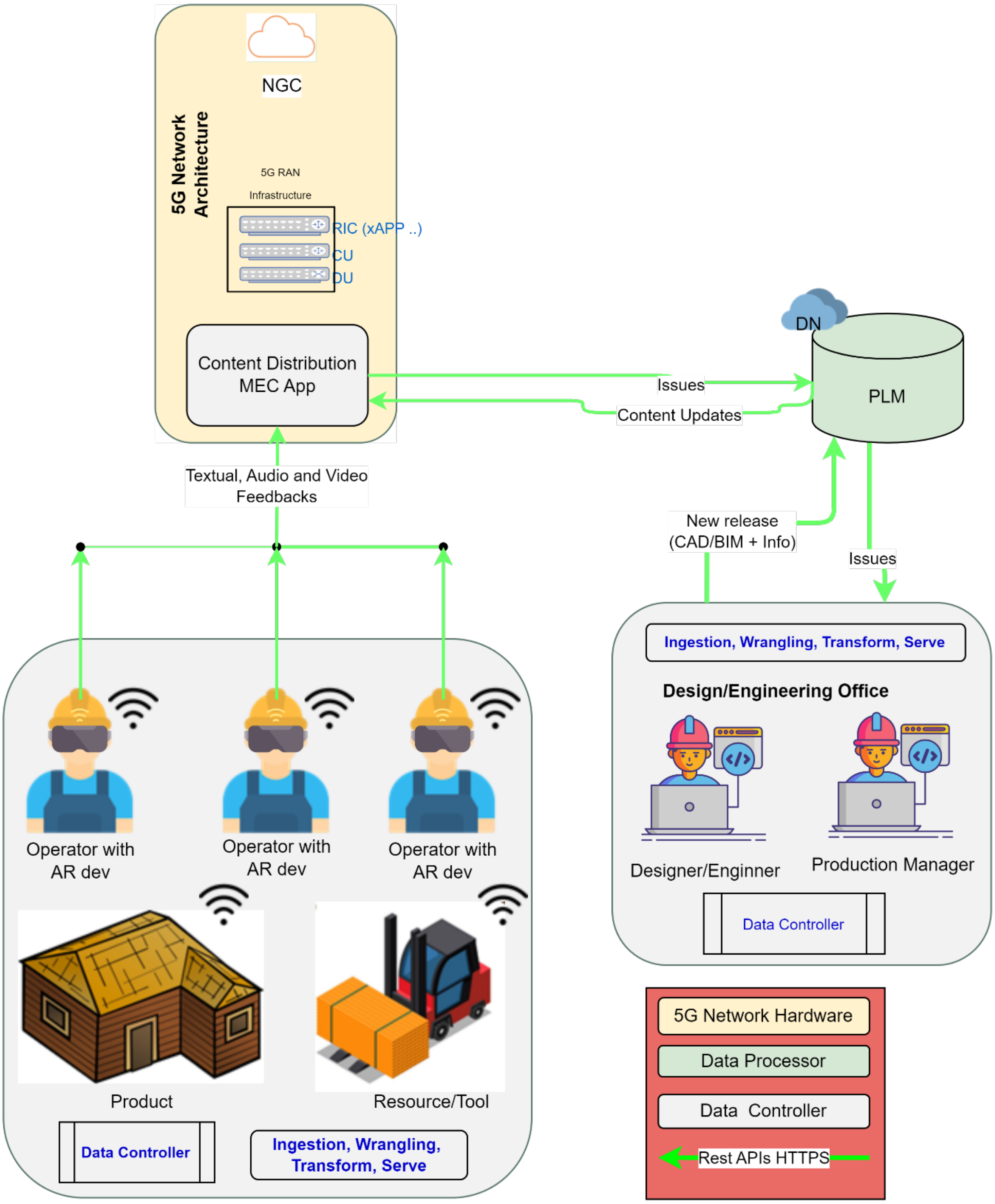


Figure 16: Use Case 2.3, Data driven design to production cycle optimization

Alert, normal, slim and debug data schema for the exchange of information will be used.

6. UC Category 3: Construction and renovation with wooden elements, valorisation of composite waste

UC Category 3 shown in **Figure 17** addresses selected steps in the lifecycle of wooden elements: wooden houses' lifecycle and analysis of alternative wooden composite materials will be studied in category 3 Use Cases.

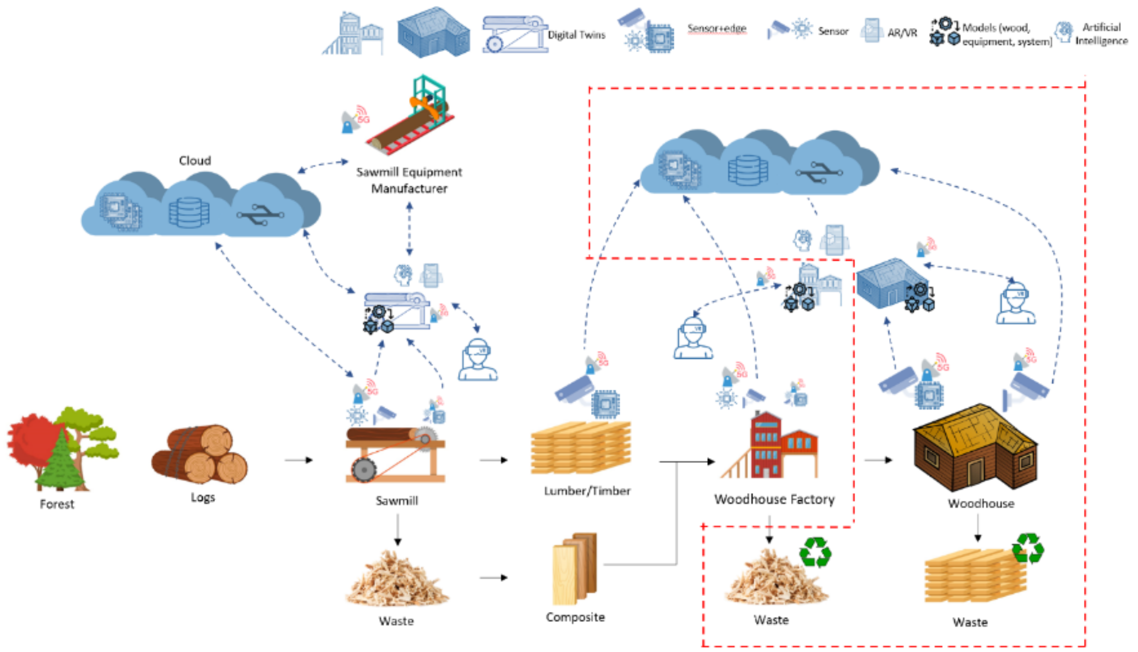


Figure 17: Use Case category 3

In Use Case 3.1 Enhanced Mobile Broadband (eMBB) and Ultra-Reliable Low Latency Communications (URLLC) will be used for communication and will also target 5G Self-Organizing Networks (SON) functionality. Whereas in Use Case 3.2 Massive Machine-Type Communications (mMTC) will be used along with SON. Location based services will not be used in Use Case category 3.

6.1. UC Cat 3 - Data flow architecture and schema

For communication Rest APIs will be used using JSON over HTTPS to transfer data to the server and the information will be encoded in UTF-8. JSON information is shown in appendix II. REST APIs based communication will used UC3.1, UC3.2 and UC3.3 across the green arrows in the **Figure 18**, **Figure 19** and **Figure 20**. It will be further detailed during the project execution and the details will be provided in the next release of this document (D1.6).
As stated earlier, in multi-organization settings, data-controllers own the data & published it for processing. Whereas data-processors process the

data on behalf of data-controllers for different applications or use-cases. Data-controllers in white blocks, shown in Figure *18*Figure *18*, Figure *19*Figure *19* and Figure *20*Figure *20* will start from ingesting and wrangling the data, transform data if required, and later serve (publish) it to the data processor. Data-processors shown in green blocks will process the data and later serve (publish) the generated output back to the data controllers. Data processors must also wrangle the output data before publishing It to data controllers to avoid transmitting any proprietary information. In the context of data architecture, 5G-connectivity shown in yellow colour in Figure *18*Figure *18*, Figure *19* Figure *19*and Figure *20* Figure *20*will only be used for information transfer and will not process the data directly.

6.2. UC3.1 AR and Industry 5.0 UX-assisted assembly and renovation of green buildings

As shown in Figure 18, The objective of UC3.1 is to test and validate XR technology to support the assembly of the modules of wooden houses at the construction site. This UC will leverage the tools, technologies and tested in UC2.2 during the construction of the modules, extending their use on construction sites. Enhanced mobile broadband & ultra-reliable low latency communications will be used along with self-organizing networks functionality of 5G network.

Two-way high speed and low latency communication is required. We will use alert, normal and debug data schema for the exchange of information.

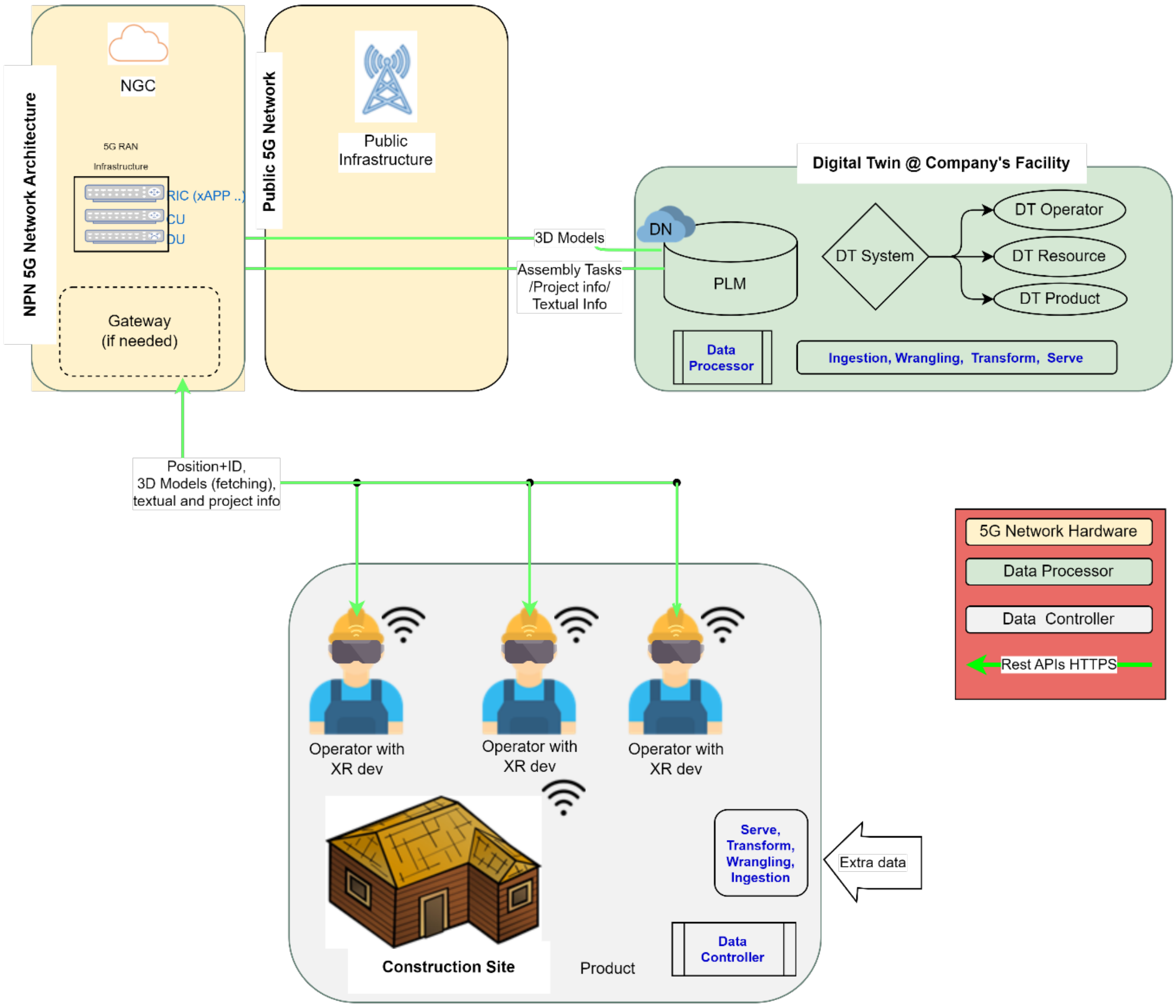


Figure 18: Use Case 3.1, AR and Industry 5.0 UX assisted assembly and renovation of green buildings

6.3. UC3.2 Long term SHM of wooden construction using IoT and embedded artificial intelligence

In Use Case 3.2 shown in **Figure 19**we will exploit the capability of 5G networks and 5G based IoT devices to support the structural health management of wooden constructions. UC3.2, covers the use phase of the house, thus addressing 5G IoT devices, tools and methodologies supporting the monitoring and maintenance. Massive machine-type communications are required. We will use normal & debug data schema for exchange of Information. Data will be generated from the onsite wooden home and can use private and public 5G network. Therefore, it is important to use the encryption for incoming data and outgoing alerts events.

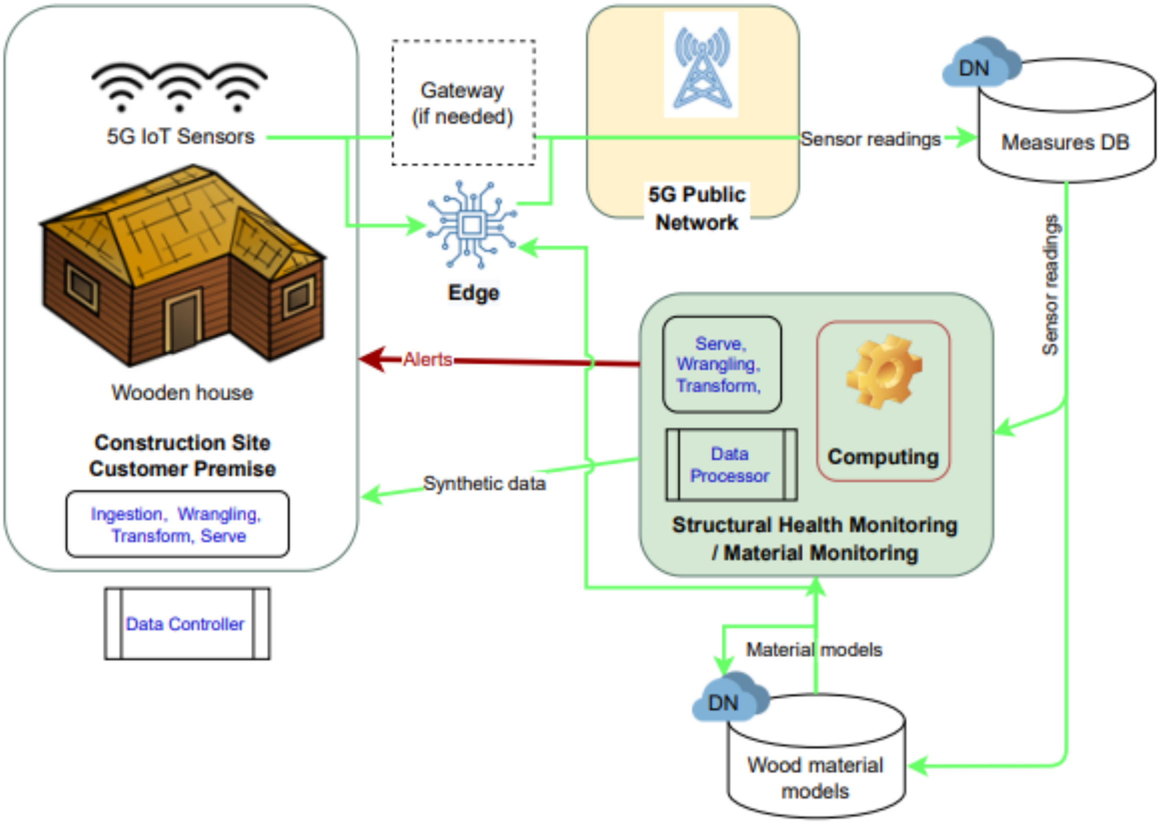


Figure 19: Use Case 3.2, Long term SHM of wooden constructions using IoT and embedded AI

6.4. UC3.3 Wooden composites waste recycling

The objective of UC3.3 as shown **Figure 20** is to validate the feasibility of material models used to drive the selection of suitable wood composites for woodhouse manufacturing and design.

This is the only Use Case that will not use any 5G-specific connectivity. We will use normal & debug data schema for exchange of information.

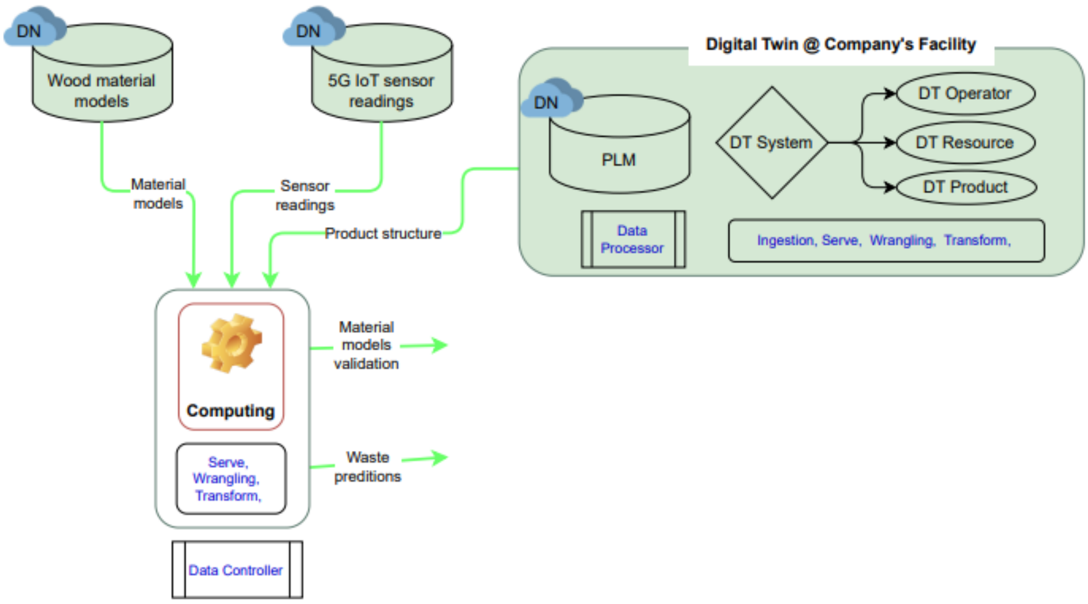


Figure 20: Use Case 3.3, Wooden composites waste recycling

7. System Requirements

General Data Protection Regulation (GDPR) requirement combined with EU draft Data Act [6] and EU Data Governance Act [7] requirements are considered in creating the data architecture. We also make it flexible and extendable for different usage in the project. GDPR draft requirements are listed below in section 7.1 and 7.2.

7.1. Data Controller

The data controller owns the data and gives permission to the Data processors to processes information and extract/generate useful information. All data processors are register with Data controller.

7.2. Data Processor

The data processors process the information on behalf of the data controller and generate metadata.

7.3. Risks and Mitigation plans

The data Controller should inform the Data controller about any data breach within the number of pre-agreed days.

8. Conclusions

Data sharing and management is an important task in projects management especially in the current dynamic IT environment. Therefore, it is important to select flexible data schemes, and an appropriate data engineering lifecycle to work with different tools frameworks along with a variety of national and international regulations. Security & privacy is the basic building block along with an efficient and flexible schema as per EU set regulations.

We propose to use Eclipse Arrowhead for IOT data management including data acquisition, storage and sharing. Its architecture is based on open standards and opensource platforms.

8.1. Recommendations

Extensive use of opensource tools, framework combined with third party audit are the best practices for high quality products. More compliance with national and international regulation and standards will provide better interoperability that will in turn improve the product quality and reduce the cost.

8.2. Next steps

Get more information about different use cases and provide better support to all use cases. Define the REST APIs for the data exchange according to the use cases evolution during the project.

9. References

[1] Intel smart edge open: https://github.com/smart-edge-open
[2] Eclipse Arrowhead: https://projects.eclipse.org/projects/iot.arrowhead
[3] Azure IoT: https://learn.microsoft.com/en-us/azure/architecture/reference-architectures/iot
[4] ISO 10303: https://www.iso.org/standard/66654.html
[5] ISO 27001: https://www.iso.org/isoiec-27001-information-security.html
[6] EU draft data Act: https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113
[7] EU Data governance Act: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020PC0767&from=EN
[8] EU Data governance Act: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020PC0767&from=EN
[9] Amazon reference architecture https://learn.microsoft.com/en-us/azure/architecture/aws-professional/services
[10] Azure cloud services https://azure.microsoft.com/en-us/get-started/
[11] Google Cloud services: https://cloud.google.com/docs/open-tutorials
[12] Apache Cassandra: https://cassandra.apache.org/
[13] Apache airflow: https://airflow.apache.org/
[14] Eclipse Arrowhead: https://projects.eclipse.org/projects/iot.arrowhead
[15] Elastic: https://www.elastic.co/
[16] Snowflake: https://docs.snowflake.com/en/user-guide/intro-key-concepts.html
[17] Bigquery: https://cloud.google.com/bigquery
[18] SQL: https://www.mysql.com/
[19] Mongodb: https://www.mongodb.com/home
[20] Protobuf: https://developers.google.com/protocol-buffers
[21] Apache Avro: https://avro.apache.org/
[22] Data Governance Framework Image: https://www.slidesalad.com/product/data-governance-powerpoint-template-ppt-slides/
[23] digital-twin-institute-position-paper-(december-2020).pdf, aiaa.org
[24] Attard, J. , Brennan, R. 2018. A semantic data value vocabulary supporting data value assessment and measurement integration. In:, Cordeiro,J., Filipe,J. (Eds).ICEIS 2018 - Proceedings of the 20th International Conference on Enterprise Information Systems. 21-24

March, 2018. Vol. 2. Polytechnic Institute of Setúbal / INSTICC, Portugal, Funchal, Madeira, Portugal. doi:10.5220/0006777701330144. pp. 133-144.

[25] Attard, J., Orlandi, F., Auer, S. 2017. Exploiting the value of data through data value networks. ACM International Conference Proceeding Series. 7 March 2017 10th International Conference on Theory and Practice of Electronic Governance, ICEGOV 2017. March 7. - March 9. 2017. Vol. Part F128003. ACM, New Delhi, India. doi:10.1145/3047273.3047299. pp. 475-484.

[26] Viscusi, G., Batini, C. 2014. Digital information asset evaluation: Characteristics and dimensions. Lecture Notes in Information Systems and Organisation 7: pp. 77-86. doi:10.1007/978-3-319-07040-7_9.

[27] DAMA International 2017. DAMA-DMBOK : Data management body of knowledge. 2. ed. Technics Publications, USA

[28] Ghorbani, A. , Zou, J. 2019. Data shapley: Equitable valuation of data for machine learning. In:, Chaudhuri,K., Salakhutdinov,R. (Eds).36th International Conference on Machine Learning, ICML 2019. June 9-15, 2019. Vol. 2019-June. Proceedings of Machine Learning Research, Long Beach, California, USA. Available from http://proceedings.mlr.press/v97/ghorbani19c/ghorbani19c.pdf, pp. 2242-2251.

[29] Lasi, H., Fettke, P., Kemper, H.-., Feld, T., Hoffmann, M. 2014. Industry 4.0. Business and Information Systems Engineering 6(4): pp. 239-242. doi:10.1007/s12599-014-0334-4.

[30] Shukla, M., Manjunath, S., Saxena, R., Mondal, S., Lodha, S. 2015. WinOver enterprise dark data. Proceedings of the ACM Conference on Computer and Communications Security. October 12-16, 2015. Association for Computing Machinery, Denver, Colorado, USA. doi:10.1145/2810103.2810131. pp. 1674-1676.

[31] Strong, D.M., Lee, Y.W., Wang, R.Y. 1997. Data quality in context. Communications of the ACM 40(5): pp. 103-110. doi:10.1145/253769.253804.

[32] Wand, Y., Wang, R.Y. 1996. Anchoring data quality dimensions in ontological foundations. Communications of the ACM 39(11): pp. 86-95. doi:10.1145/240455.240479.

10. Appendix I: Definitions

- '**data**' means any digital representation of acts, facts or information and any compilation of such acts, facts, or information, including in the form of sound, visual or audio-visual recording.
- '**re-use**' means the use by natural or legal persons of data held by public sector bodies, for commercial or non-commercial purposes other than the initial purpose within the public task for which the data were produced, except for the exchange of data between public sector bodies purely in pursuit of their public tasks.
- '**non-personal data**' means data other than personal data as defined in point (1) of Article 4 of Regulation (EU) 2016/679.
- '**metadata**' metadata is a type of data that describes other data, for example, describing the origin, structure, or characteristics of computer files, webpages, databases, or other digital resources.
- '**data holder**' means a legal person or data subject who, in accordance with applicable Union or national law, has the right to grant access to or to share certain personal or non-personal data under its control;
- '**data user**' means a natural or legal person who has lawful access to certain personal or non-personal data and is authorised to use that data for commercial or non-commercial purposes;
- '**data sharing**' means the provision by a data holder of data to a data user for the purpose of joint or individual use of the shared data, based on voluntary agreements, directly or through an intermediary;
- '**access**' means processing by a data user of data that has been provided by a data holder, in accordance with specific technical, legal, or organisational requirements, without necessarily implying the transmission or downloading of such data.
- '**data altruism**' means the consent by data subjects to process personal data pertaining to them, or permissions of other data holders to allow the use of their non-personal data without seeking a reward, for purposes of general interest, such as scientific research purposes or improving public services
- '**secure processing environment**' means the physical or virtual environment and organisational means to provide the opportunity to re-use data in a manner that allows for the operator of the secure processing environment to determine and supervise all data processing actions, including to display, storage, download, export of the data and calculation of derivative data through computational algorithms.
- '**representative**' means any natural or legal person established in the Union explicitly designated to act on behalf of a provider of data sharing services or an entity that collects data for objectives of general interest made available by

natural or legal persons on the basis of data altruism not established in the Union, which may be addressed by a national competent authority instead of the provider of data sharing services or entity with regard to the obligations of that provider of data sharing services or entity set up by this Regulation.

11. Appendix II: An example of JSON information for different use-cases

Name	Type	count	Description
Date	Date	1	Date
Version	Text	1	Describe the schema version
Time	Time	1	Time stamps in milli seconds
Permission header	text	1	Permission module
Location	Text	1	location
Orientation	Text	1	Angle in degree
Battery level	Text	1	In percentage
Network-ID	Text	1	Identity name
Gateway-ID	Text	1	URL or IP
HUB-ID	text	1	URL or IP
MAC	text	1	MAC
Data	vector	1	Product background data
Data	Vector	1	Performance data
Data	Vector	1	Digital Twin data
Data	Vector	1	Processes data
Data	vector	1	Container for data
Data	vector	1	System data
Data	Vector	1	Sensor data
Sensor-Data-1	int	1	Humidity
Sensor-Data-2	int	1	Temperature in centigrade
Transformation-Event	text	1	Reason why data is transmitted