Ref. Ares(2025)1540766 - 26/02/2025

Biological Resources Certifications Schemes

Type of Action (Horizon-IA)

Call identifier: HORIZON-CL6-2021-ZEROPOLLUTION-01

Start date of project: 09–01-22 | **Duration:** 3 years (36 months)

D3.2 The cloud based software application hosting the operations and functions of full T&T of BioReCer

Deliverable reference number

D3.2

Work package number

WP3

Due date of deliverable

28.02.2025

Actual submission date

26.02.2025

Authors

Name	Organisation	Email
Sebelan Danishvar	Brunel University	sebelan.danishvar@brunel.ac.uk
Alireza Mousavi	Brunel University	alireza.mousavi@brunel.ac.uk

Type

☒	**R**	Document report
☐	**DEM**	Demonstrator, pilot, prototype
☐	**DEC**	Websites, patent fillings, videos, etc.
☐	**OTHER**	

Dissemination Level

☒	**PU**	Public, fully open, e.g. web
☐	**SEN**	Sensitive, limited under the conditions of the Grant Agreement
☐	**CI**	Classified, information as referred to in Commission Decision 1001/844/EC

The BioReCer Project receives funding from the Horizon Europe Framework Programme under the Grant Agreement Number 101060684.

Funded by the European Union

Lead beneficiary

Brunel University

Contributing beneficiaries

Brunel University, Anfaco

Internal reviewer(s)

Name	Organization	Email
Dusica Banduka	NOVA Institute	dusica.banduka@nova-institut.de
Sofía Estévez Rivadulla	USC	sofia.estevez.rivadulla@usc.es
Lucía González-Monjardin	CETAQUA	lucia-jimena.gonzalez@cetaqua.com
Monica Pandey	MCS	pandey@meo-carbon.com

Change log

Date	Issue/Version	Reason for change

Report

Action Title	Biological Resources Certifications Schemes
Action Acronym	BioReCer
Action Number	101060684
Deliverable Identifier	D3.2
Deliverable Title	The cloud based software application hosting the operations and functions of full T&T of BioReCer
Document Status	Final version
Version	1.0
Authors	Sebelan Danishvar; Alireza Mousavi
Lead Beneficiary	Brunel University
Deliverable Type	Report
Dissemination Level	Public
Format	pdf
Due Month	February
Date	26.02.2025
DOI	
Keywords	BioBased, BioEconomy BioWaste, Tracking and Traceability, Software, Blockchain

Document History

Version	Description	Date
1.0	Version submitted to the EU portal	25/02/2025

Table of contents

1. Introduction 9
1.1 Purpose and scope of the deliverable 9
1.2 Key features of the T&T software application 10
1.3 Contribution to BioReCer's vision 12
1.4 Structure of the deliverable 12

2. Core functionalities of the T&T system 14
2.1 End-to-End traceability mechanisms 15
2.1.1 Framework overview 15
2.1.2 Functional workflow 17
2.1.3 Addressing key challenges and solutions 19
2.2 Real-time data capture and processing 20
2.2.1 Technological foundations 20
2.2.2 Data acquisition infrastructure 22
2.2.3 Data transmission and storage 26
2.2.4 Dynamic data processing 35
2.3 Reporting on reliability, conformity, and performance 42
2.3.1 Reliability reporting 43
2.3.2 Conformity reporting 44
2.3.3 Overall performance reporting 46

3. Technical specifications 50
3.1 Software architecture and design 50
3.1.1 Overview of the architecture 50
3.1.2 Integration with external systems. 51
3.1.3 Data processing workflow 52
3.1.4 Security and compliance considerations 54
3.2 Data handling and storage protocols 56
3.2.1 Efficient data organization and management 56
3.2.2 Storage infrastructure 59
3.3 Scalability and optimization 61

3.4 Cloud-based infrastructure overview 63

4. Hyperspectral imaging (HSI) integration 67

4.1 Introduction to HSI technology 67

4.2 Role of HSI in the BioReCer framework 68

4.3 Data collection and processing with HSI 70

4.4 Benefits of HSI for biological feedstock traceability 74

5. Implementation and performance 75

5.1 Deployment methodology 75

5.2 Performance metrics and benchmarking 79

5.3 Results from initial validation and testing 85

5.3.1 Stress testing: 86

5.3.2 Load testing: 86

5.3.3 Functional Testing 88

5.3.4 End-to-End system validation 88

5.3.5 Automated error detection 89

6. Conclusion 90

7. List of Abbreviations 91

8. References 93

List of figures

Figure 1: Key features of the WP3's T&T software application 12
Figure 2: Standardization of data 16
Figure 3: Bio-based supply chain traceability workflow 18
Figure 4: Enhancing traceability with technology 22
Figure 5: Data acquisition modes..24
Figure 6: Hash functions..28
Figure 7: Centralized blockchain enabling smart contract automation and 30
Figure 8: Data processing in centralized blockchain 32
Figure 9: CAD design for transmittance setup 71
Figure 10: Fresh algae harvesting on the beach 72
Figure 11: Algae spectral data acquisition with transmittance setup 72
Figure 12: Sample preparation for laboratory analysis...73
Figure 13: Mean spectrum of the samples for each algae species...............................74

Publishable executive summary

Deliverable D3.2 represents the advancements made in WP3 (Product tacking and traceability (T&T) for circular value chain integration), particularly in Task 3.2 (Design, build and implementation of the T&T software and hardware apparatus), in terms of the delivery of a study track and traceability (T&T) system through the BioReCer project addressing the specific bio-based supply chains' challenges. This deliverable focuses on transforming the theoretical framework outlined in D3.1 (Complete and approved theoretical framework for full T&T as the modus operandi of biological feedstock labelling demonstrating reliability, conformity and performance) into a practical, scalable, and functional application that integrates seamlessly with the Information and Communication Technology (ICT) tools. Initial deployments have already demonstrated improvements in traceability accuracy and supply chain data validation, reinforcing regulatory compliance across bio-based industries.

The core achievement of this deliverable is the creation of a centralized blockchain based T&T infrastructure that ensures data unchangeability and traceability. This system is designed to address the complexities of bio-based supply chains by securely storing compliance data, circularity indicators (CIs), and critical sustainability metrics.

By developing Application Programming Interfaces (API) driven data integration with ICT tool, he T&T system has successfully enhanced data exchange efficiency, allowing stakeholders to access real-time sustainability metrics. Initial results indicate that stakeholders benefit from faster compliance reporting, reduced administrative overhead, and increased trust in the traceability of bio-based materials.

A key contribution of Task 3.2 was the operationalization of CIs developed collaboratively with WP2. WP3’s implementation ensured these indicators were dynamically computed within the T&T system, empowering stakeholders to measure and improve their environmental performance. Advanced algorithms and formula integration enabled the system to compute over 40 circularity metrics accurately and efficiently, making it a vital tool for achieving BioReCer’s sustainability goals.

Comprehensive testing and validation processes were conducted to evaluate the system’s scalability, reliability, and accuracy. Stress and load testing demonstrated the system’s capacity to handle large data volumes and simultaneous API queries from ICT tool without performance degradation. Pilot implementations confirmed the practicality of the T&T system in real-world scenarios, showcasing its readiness to support stakeholders across diverse bio-based sectors such as fisheries, agriculture, and forestry.

Task 3.2's contributions also included the development of modular system architecture and cloud-based infrastructure, enabling scalability and adaptability for future expansions. These advancements prepare the system to incorporate emerging technologies, such as machine learning models and predictive analytics, to further optimize performance and stakeholder engagement.

Deliverable D3.2 expressed the commitment of WP3 to the development of the objects of the BioReCer project through innovative technological solutions and collaborative efforts. By providing stakeholders with a reliable and transparent T&T system, WP3 plays a pivotal role in driving sustainability, compliance, and efficiency within bio-based supply chains. This deliverable establishes a strong foundation for future developments and reinforces BioReCer's mission to lead the transformation toward a sustainable bio-based economy.

1. Introduction

Deliverable D3.2 is partly about developing the theoretical groundwork we initiated in D3.1 (Complete and approved theoretical framework for full T&T as the modus operandi of biological feedstock labelling demonstrating reliability, conformity and performance), moving from framework-level design to practical application. It encapsulates the implementation of a cloud based Track and Traceability (T&T) system along with innovative technologies to address the sophisticated requirements of bio-based domain areas including the fishery, forestry, agricultural, and urban bio-waste sector.

This deliverable tackles the multifarious features/requirements of bio-based supply chain integration including the real-time monitoring of data, trustworthiness, robustness and interoperability with sector-specific software tools and certification schemes. Building on advanced digital technologies, such as centralized blockchain, Internet of Things (IoT) enabled devices and Hyperspectral Imaging (HSI), the T&T system aims to enforce compliance with sustainability standards and regulatory requirements. In D3.2, we assessed the viability of the framework via case-specific applications. In addition, we lay the groundwork for deployment at scale across all BioReCer case studies.

Reflecting BioReCer's ambition to support the bio-based economy by providing actionable insights and developing collaboration between stakeholders, this deliverable is a step towards that goal. Here we present the results of a practical application, whose guidelines strengthen the strategic goals of the project and provide a replicable model: the solution ensures traceability, accountability, and environmental compliance in bio-based supply chains.

1.1 Purpose and scope of the deliverable

The purpose of D3.2 is to transition from theory to practice by deploying a working T&T software application. The deliverable addresses the design, implementation, and evaluation of the software and tools necessary for the compatibility of the software with sectoral tools and ICT tool interface managed by WP5 (BioReCer ICT tool). The primary scope encompasses integrating real-time monitoring, blockchain validation of data and calculation of environmental indicators corresponding to the requirements found in BioReCer of case studies. This deliverable aims to create a replicable model for bio-based supply chain traceability and compliance through a focus on interoperability and scalability.

1.2 Key features of the T&T software application

The T&T software application introduced in this deliverable integrates advanced functionalities tailored to fulfil the diverse needs of bio-based supply chains. Designed with adaptability and user-centricity in mind, the application supports seamless data management, monitoring, and reporting. Key features include:

Centralized blockchain model

The centralized blockchain infrastructure serves as the secure backbone of the T&T framework, providing immutable data storage and supporting scalability and efficient system management. It underpins all supply chain activities, providing traceability and accountability across diverse sectors.

Application Programming Interface (API) driven interoperability

Seamless integration between the centralized blockchain managed by WP3 (Product tacking and traceability (T&T) for circular value chain integration) and the ICT tool handled by WP5 (BioReCer ICT tool) is achieved through robust API connections. This enables real-time data exchange, facilitating compatibility with sector-specific technologies and fostering cross-sector collaboration without direct interaction between stakeholders and the blockchain.

Real-Time monitoring and dashboards

Real-time dashboards, accessible via the ICT tool, deliver actionable insights into supply chain activities. These dashboards display critical metrics such as inventory status, environmental indicators, and compliance performance, enhancing decision-making and operational efficiency while maintaining system integrity.

Sector-specific tools and adaptability

The T&T framework incorporates specialized tools tailored to sector-specific needs, including HSI for fisheries, vessel monitoring systems, and environmental sensors. These tools assist precise data collection and validation that enhance the adaptability and effectiveness of the system across bio-based industries.

Calculation and visualization of CIs

Automated CIs computation is a core functionality of the application. Using data collected from stakeholders via ICT tool, the system enables real-time tracking of sustainability metrics and align compliance with circular economy principles.

Compliance monitoring

Integrated compliance mechanisms to validate data related to sustainability standards and regulatory requirements. Automated checks streamline validation of certification data, reducing manual workload and minimizing errors while ensuring system reliability.

Scalability and flexibility

The T&T application is designed to evolve with the needs of bio-based sectors. Its scalable architecture supports the addition of datasets, functionalities, and case studies, aligning with BioReCer's vision for sustainability and transparency.

Data analytics and visualization capabilities

Advanced analytics tools offer deep insights into supply chain performance and sustainability. By identifying bottlenecks and analysing trends, these capabilities optimize operations, improving efficiency and environmental outcomes.

Role-based access control (RBAC)

RBAC guarantees stakeholders interact with the system exclusively through the ICT tool managed by WP5. This secure configuration assigns permissions tailored to each stakeholder's role:

- Bio-producer: Input data such as batch numbers and origin details.
- Bio-industry: Access validated data for product transformation and compliance documentation.
- Certification body: Retrieve summary reports and compliance data for monitoring.

This approach safeguards system-wide confidentiality and data integrity while maintaining a seamless and user-friendly experience. The centralized blockchain managed by WP3 acts as the secure repository for all validated data and provides traceability and accountability across the supply chain.

This comprehensive set of features underscores the T&T application's pivotal role in achieving BioReCer's objectives, providing a reliable and innovative solution for managing bio-based supply chains (Figure 1).

1.3 Contribution to BioReCer's vision

Deliverable D3.2 advances BioReCer's vision by transforming the T&T framework from D3.1 into a functional, cloud-based application that enhances traceability, sustainability, and compliance across bio-based supply chains. By integrating seamlessly with the ICT tool managed by WP5, this system facilitates real-time monitoring, automated Circularity Indicators (CIs) calculations, and data-driven decision-making.

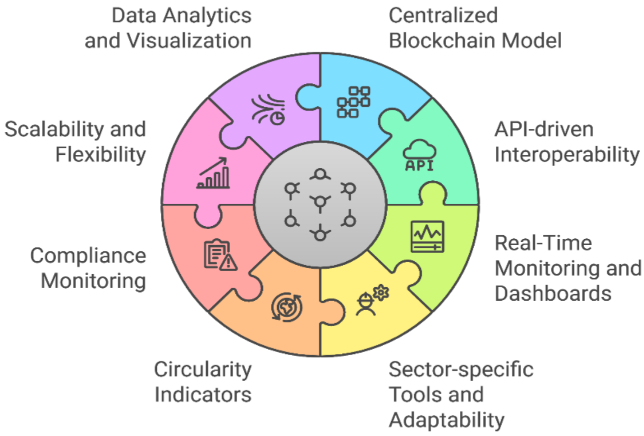

Figure 1: Key features of the WP3's T&T software application

A key contribution of D3.2 is the adaptation of T&T tools to sector-specific needs, ensuring effective implementation across fisheries, forestry, agriculture, and urban/industrial waste. Technologies such as HSI and automated compliance reporting enhance traceability and accountability within each supply chain.

Furthermore, this deliverable strengthens stakeholder collaboration by allowing biological feedstock producers, industry players, and certification bodies to interact efficiently within the system. With its scalability and interoperability, D3.2 establishes a replicable model that supports BioReCer's mission of driving sustainability, transparency, and regulatory alignment in the bioeconomy.

1.4 Structure of the deliverable

This deliverable is structured into nine sections that provide insights on key aspects of the T&T software application developed in the context of BioReCer project.

The **Core functionalities of the T&T system** section delves into the operational mechanisms of the software, including end-to-end traceability, real-time data capture, and reporting functionalities that indicate reliability, conformity, and performance across the supply chain.

In **Technical specifications**, the deliverable outlines the software architecture and design, data handling protocols, and the cloud-based infrastructure supporting the centralized blockchain model. This section provides the technical foundation for the framework's operation.

HSI integration introduces the role of HSI technology within the framework, focusing on its application for biological feedstock traceability. It discusses the benefits, data collection processes, and the added value HSI brings to the T&T system.

The **Implementation and performance** outlines the development methodology, including agile practices employed to iteratively refine the system. It also addresses challenges encountered during deployment and the solutions implemented to overcome them. In this section the deliverable evaluates the system's effectiveness through benchmarking and initial validation results. Key performance indicators (KPIs) are presented to assess the software's traceability and operational performance.

Finally, the **Conclusion** summarizes the achievements of the deliverable and its contribution to BioReCer's objectives. It reflects on the broader impact of the T&T framework in advancing sustainability, transparency, and accountability in bio-based supply chains.

This structure provides a logical flow, enabling readers to grasp the comprehensive scope, technical underpinnings, and practical applications of the T&T framework within the BioReCer project.

2. Core functionalities of the T&T system

The T&T system is the cornerstone of WP3's efforts to enhance transparency, sustainability, and accountability across bio-based supply chains. Designed and implemented, T&T system is a robust and dynamic platform engineered to address the complexities of modern supply chains, particularly within the bio-based sectors. This section provides a comprehensive overview of the system's fundamental features, emphasizing its ability to enable seamless traceability, facilitate real-time data capture and processing, and deliver insightful reporting on critical parameters such as reliability, conformity, and performance.

At its core, the T&T system embodies a multi-functional framework that combines advanced digital technologies with sector-specific tools to promote the continuous and accurate monitoring of supply chain activities. The system takes advantage of a centralized blockchain infrastructure as its backbone, providing a secure, transparent, and immutable repository for all data collected from stakeholders. This centralized model simplifies data management and enhances scalability, which makes the system adaptable to varying supply chain requirements.

To operationalize traceability, the T&T framework employs API-driven interoperability and provides seamless data exchange between WP3's centralized blockchain and WP5's ICT tool. This integration bridges the gap between stakeholders and the blockchain, offering a user-friendly interface while maintaining data integrity and security. The system's architecture supports real-time data synchronization, enabling stakeholders to monitor supply chain events as they occur, developing timely decision-making and proactive issue resolution.

In addition to its technical underpinnings, the T&T system emphasizes user-centric functionalities, including dynamic dashboards and interactive reporting tools. These features allow stakeholders to visualize key metrics such as inventory status, environmental indicators, and regulatory compliance performance. The inclusion of sector-specific tools, such as HSI, vessel monitoring systems, and environmental sensors, further enhances the system's applicability and relevance, representing that it meets the unique needs of diverse bio-based industries.

The T&T system also incorporates comprehensive validation mechanisms to confirm data accuracy and reliability. Automated checks are conducted at every stage of data collection and processing, reducing errors and inconsistencies. Furthermore, the system's design adheres to rigorous security protocols, including RBAC as part of the T&T system, promoting that data is accessible only to authorized users and safeguarding confidentiality.

By addressing the interconnected challenges of traceability, data accuracy, regulatory compliance, and stakeholder engagement, the T&T system positions itself as an indispensable tool for advancing BioReCer's mission. Its ability to provide actionable insights and ensure accountability across the supply chain not only supports operational efficiency but also aligns with the project's broader goals of fostering sustainability and innovation in the bio-based economy.

2.1 End-to-End traceability mechanisms

The implementation of end-to-end traceability mechanisms in the BioReCer project ensures that biological feedstocks are tracked and traced seamlessly throughout their lifecycle. These mechanisms, central to the BioReCer T&T system, provide a transparent, secure, and scalable framework to meet the challenges of modern bio-based supply chains, including variability, sustainability, and compliance.

2.1.1 Framework overview

End-to-end traceability in BioReCer is achieved through a synergistic integration of digital technologies, centralized data systems, and sector-specific methodologies. This approach addresses the complexity of biological feedstock supply chains, characterized by diverse stakeholders, fragmented data, and stringent sustainability requirements. Key components of this mechanism are as below.

Event-driven traceability:

This is a cornerstone of the WP3 end-to-end traceability system, enabling the precise monitoring and recording of critical supply chain events such as harvesting (supplier), processing, transportation, and distribution in real-time. This approach certifies that every significant action within the supply chain is logged and associated with a unique digital identifier. These identifiers act as immutable records, providing a clear and comprehensive history of each product's journey. By capturing essential metadata, such as timestamps, geolocations, and batch numbers, event-driven traceability establishes a high level of accuracy and accountability [1]. This mechanism enhances operational transparency and empowers stakeholders to validate sustainability claims, monitor compliance with regulatory standards, and make informed decisions at every stage of the supply chain. By bridging gaps in data sharing and integration, event-driven traceability forms the backbone of a robust and reliable T&T system that supports circular economy goals.

Data integration and standardization:

Both are critical components of the WP3 end-to-end traceability mechanisms, certifying seamless communication and compatibility across diverse data systems [2]. This process harmonizes information from multiple sources by employing robust APIs, which enable interoperability with sector-specific tools and existing certification schemes. By standardizing data exchange protocols, the system overcomes the inherent challenges of fragmented supply chains, allowing for a unified approach to traceability and sustainability monitoring (Figure 2).

Furthermore, the data integration framework is designed to align with regulatory and policy frameworks, such as the European Union's Circular Economy Action Plan and International Organization for Standardization (ISO) standards. This verifies that all data collected and processed adhere to stringent sustainability and circularity requirements, facilitating compliance across diverse sectors. Whether it involves fisheries, forestry, or agricultural operations, the system's ability to integrate data from various stakeholders enhances its utility and scalability.

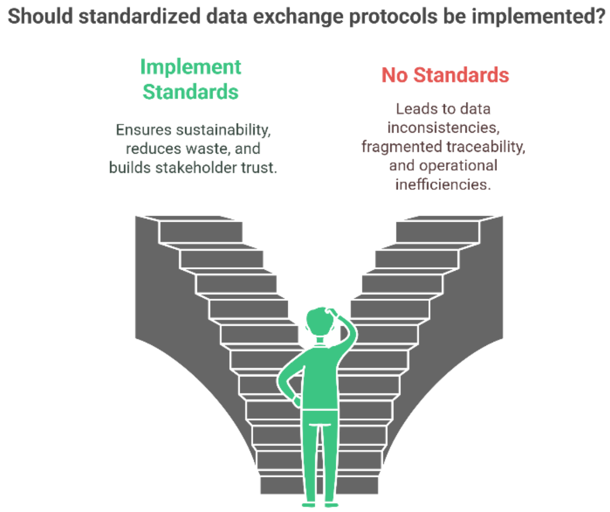

Figure 2: Standardization of data

This capability streamlines operations and strengthens stakeholder confidence by providing a transparent and verifiable record of compliance with environmental and sustainability criteria.

Blockchain for data security and transparency:

This part is a fundamental feature of the WP3 T&T system and confirms that all supply chain data is securely stored and safeguarded against tampering. Utilizing the centralized blockchain infrastructure introduced in **D3.1**, the system provides an unchangeable and transparent ledger that records every transaction and event within the supply chain. This

makes certain that data integrity is maintained throughout, offering stakeholders a reliable and verifiable source of information.
In addition to secure data storage, the blockchain uses smart contracts to automate the verification of data accuracy. These self-executing contracts validate and process data according to predefined rules, reducing the potential for human error and aligning compliance with regulatory and sustainability standards. By integrating these capabilities, the blockchain enhances trust and accountability among all stakeholders, fostering collaboration and confidence in the traceability framework. This robust approach positions blockchain technology as a key enabler for achieving transparency and security in bio-based supply chains.

Sector-specific adaptations:
They play an important role in the BioReCer T&T system to establish that the framework meets the unique requirements of diverse bio-based sectors. By integrating tailored tools and technologies, the system addresses the specific challenges and operational needs of each sector. For instance, HSI is employed for biomass analysis in fishery, agricultural, forestry and urban bio-waste applications, offering detailed insights into the composition and quality of feedstocks [3]. These sector-specific adaptations enhance the precision and applicability of the traceability framework and confirm that it is robust and versatile. By tailoring tools and methodologies to each sector, the system strengthens its capacity to support sustainability, transparency, and accountability across the bio-based supply chain.

2.1.2 Functional workflow

The end-to-end traceability mechanisms follow a systematic workflow that integrates real-time data capture, analysis, and reporting (Figure 3):

Data capture:
The functional workflow begins with comprehensive data collection using IoT-enabled sensors and manual input mechanisms. These technologies record essential attributes of bio-based feedstocks, including physical, chemical, and logistical properties such as batch numbers, timestamps, and geolocation details. This meticulous metadata tagging certifies that all relevant information is captured at the source, forming the foundation for reliable and seamless traceability.

Data transmission:

Once captured, the data is securely transmitted via robust APIs to a centralized blockchain infrastructure. This clinches that all information is accurately integrated and consistently maintained across the various stages of the supply chain. By utilizing secure transmission protocols, this step guarantees data integrity while enabling interoperability between diverse stakeholders and systems.

Real-time monitoring and event logging:

The captured data is then processed for real-time monitoring and event logging. This stage is critical in providing stakeholders with actionable insights into key supply chain metrics, including inventory levels, environmental impact data, and quality compliance. Such continuous logging facilitates dynamic and timely decision-making, assuring that supply chain events are visible and auditable as they occur.

Data analytics and reporting:

Advanced analytics tools integrated into the system compute vital CIs and generate compliance reports. These outputs are visualized through dynamic dashboards accessible via an ICT tool. The dashboards allow stakeholders to evaluate sustainability metrics, assess operational efficiency, and verify regulatory compliance. This layer of reporting enhances accountability and provides a comprehensive view of the bio-based supply chain's performance.

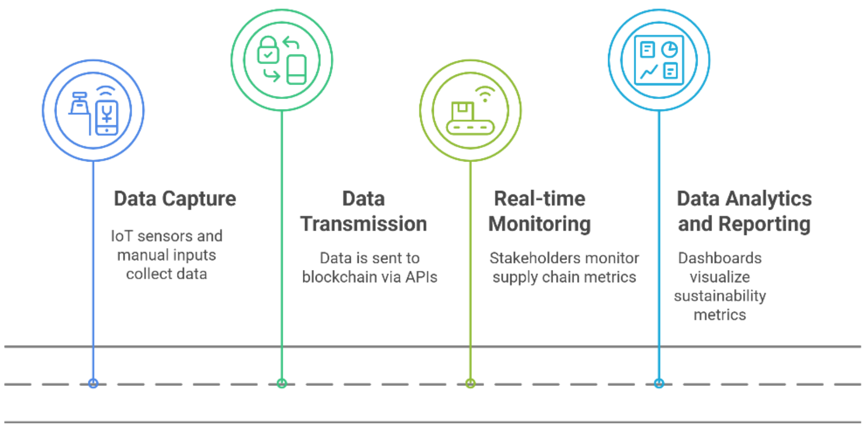

Figure 3: Bio-based supply chain traceability workflow

2.1.3 Addressing key challenges and solutions

The WP3 framework, as outlined in D3.1, provided a theoretical foundation for addressing core challenges in bio-based traceability systems, focusing on interoperability, data integrity, and scalability. Building upon this groundwork, D3.2 operationalises these concepts, demonstrating their implementation within the T&T system. This transition from theoretical framework to practical execution demonstrates that technical advancements are functional, scalable, and responsive to industry needs.

Interoperability in practice

One of the most significant barriers in bio-based supply chains is the lack of seamless communication between diverse systems and stakeholders. While D3.1 explored the necessity of interoperability, D3.2 implemented it through real-time API-driven data exchange, integrating IoT-enabled sensors and HSI to ensure sector-specific compatibility. This high-level interoperability connects fragmented systems into a cohesive framework that supports real-time data exchange and ensures alignment with certification schemes.

Strengthening data integrity

Ensuring that supply chain data remains secure and tamper-proof is critical for fostering trust among stakeholders. D3.1 introduced blockchain-based security mechanisms as a theoretical solution, while D3.2 implements a centralized blockchain infrastructure, automating data validation through smart contracts.

Blockchain technology provides an immutable record of transactions, while smart contracts automate data validation, reducing manual errors and enhancing accountability throughout the supply chain.

Scalability through modular system design

Bio-based supply chains require systems that can grow and adapt alongside increasing data volumes, stakeholder networks, and regulatory requirements. D3.1 emphasized the need for scalability, while D3.2 delivers a modular, cloud-based architecture that supports real-time data processing and integration of emerging technologies. This implementation ensures long-term sustainability, enabling the system to evolve without disrupting existing operations.

By translating the theoretical framework of D3.1 into a fully functional T&T system, D3.2 demonstrates tangible progress in traceability, regulatory compliance, and operational efficiency. The implemented API integrations, blockchain security mechanisms, and scalable architecture validate the framework's effectiveness in real-world applications.

2.2 Real-time data capture and processing

Real-time data capture and processing are at the heart of WP3's T&T system, enabling seamless and dynamic integration of supply chain activities. This section outlines how real-time systems eliminate delays in data flow, allowing stakeholders to monitor and act on critical supply chain events as they happen. It highlights the transformative potential of real-time systems in addressing challenges such as data fragmentation, compliance validation, and operational transparency in bio-based sectors.

2.2.1 Technological foundations

The practical implementation of the T&T system in WP3 builds upon the technological framework outlined in D3.1. This section focuses on the operational aspects of integrating key technologies, including IoT devices, APIs, blockchain, and HSI within the T&T system, specifically addressing traceability, compliance, and decision-making in bio-based supply chains.

Internet of Things (IoT) devices

IoT devices serve as touchpoints for real-time data collection from the physical environment, significantly bolstering the traceability and management of bio-based supply chains. These devices monitor essential parameters such as temperature, humidity, and geolocation of perishable bio-feedstocks.

While no physical IoT devices have been directly connected to the WP5 ICT tool at this stage, WP3 conducted extensive testing using dummy and simulated data that accurately emulated IoT device outputs.

- **Testing approach:** Simulated IoT data streams were created to mimic real-world sensor readings, such as temperature fluctuations during transport, humidity variations in storage, and geolocation tracking of bio-based materials. This allowed WP3 to evaluate the system's ability to process, validate, and store IoT-derived data within the blockchain-based T&T system.
- **Integration with WP5's ICT Tool:** While physical IoT integration is not yet implemented, the simulated data was structured and processed as if it were coming from live IoT sensors. This enabled testing of data ingestion, validation, and compliance checks within the T&T framework.
- **Challenges & future prospects:** The use of simulated data highlighted key considerations for future IoT integration, such as ensuring compatibility with WP5's API architecture, refining data synchronization methods, and addressing potential connectivity issues.

- **Stakeholder impact:** Even with simulated data, the tests demonstrated how IoT-enabled monitoring can enhance traceability, improve supply chain decision-making, and support regulatory compliance. The results provide a solid foundation for future deployment and real-world IoT integration.

By using simulated IoT data, WP3 effectively tested and validated the T&T system's ability to process real-time traceability data, bridging the gap between theoretical concepts in D3.1 and practical applications in D3.2. These tests ensure that the system is ready for future integration with live IoT devices when they become available in later project phases.

Application Programming Interfaces (APIs)

APIs form the spine of interoperability, connecting BioReCer's T&T system with external tools and data sources. Key functionalities include:

- Data security and validation: APIs employ encryption protocols to secure communication between stakeholders and prevent data breaches.
- Cross-sector integration: APIs standardize data formats for compatibility across diverse tools like HSI and certification platforms. For example, if an IoT device provides data in JSON while a certification platform uses XML, the API converts these formats into a unified structure before sending the data to the centralized system.

Blockchain technology

BioReCer employs a centralized blockchain model to log immutable records of all supply chain events. Blockchain is extensively discussed in D3.1.

HSI

HSI is a cutting-edge technology integrated into the T&T system to provide in-depth insights into the composition and quality of biological feedstocks. This powerful tool enables precise analysis of feedstock characteristics, maintaining compliance with quality and sustainability standards. A detailed discussion of HSI technology, its applications, and benefits will be presented in Section 4 of this deliverable.

Synergies for a perfect T&T system

The integration of IoT devices, APIs, blockchain, and HSI creates a robust, real-time T&T system that delivers:

- Precision: Real-time, granular data facilitates accuracy in monitoring and decision-making.

- Efficiency: Automated workflows and edge computing reduce latency and resource usage.
- Transparency: Blockchain's immutable records build trust and accountability among stakeholders.
- Scalability: APIs and modular technologies enable seamless expansion as new tools or data sources are incorporated.

By anchoring WP3's T&T system on these advanced technologies, the project establishes a highly scalable, transparent, and efficient framework for real-time traceability across bio-based supply chains. Figure 4 illustrates how the WP3 T&T system leverages advanced technologies to enhance traceability, ensuring scalability, transparency, efficiency, and precision across bio-based supply chains

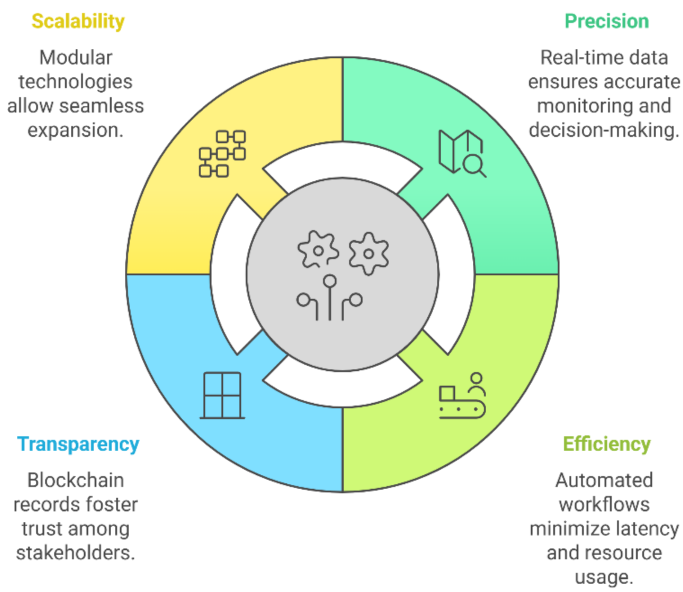

Figure 4: Enhancing traceability with technology.

2.2.2 Data acquisition infrastructure

A robust and adaptive data acquisition infrastructure integrates advanced automatic technologies and manual methodologies to ensure seamless data capture across diverse supply chain stages. The combination of IoT-enabled sensors, radio frequency identification (RFID) and barcode systems, and manual questionnaires guarantees data completeness, accuracy, and adaptability in real-world conditions (Figure 5).

Automated data acquisition technologies

- **IoT-enabled sensors**:
  IoT sensors monitor environmental parameters such as temperature, humidity, and pH levels, helping to preserve bio-based feedstocks under optimal conditions during transportation and storage. These sensors provide continuous data streams, allowing stakeholders to track the quality and safety of materials in real time. Additionally, advanced sensors equipped with geolocation capabilities capture precise transit data, enabling dynamic decision-making to address potential inefficiencies or delays in the supply chain. This seamless integration of real-time monitoring confirms operational efficiency and maintains the integrity of bio-feedstocks throughout their journey [4].
- **RFID and barcode systems**:
  RFID tags and barcodes simplify the identification and tracking of materials by assigning unique identifiers to each product or batch. These technologies enable seamless data capture, indicating that all relevant information is accurately recorded and linked to specific supply chain stages. By maintaining consistency in data management, RFID and barcode systems provide end-to-end traceability, from raw material acquisition to the final delivery of products, thereby enhancing accountability and transparency throughout the lifecycle. While not continuously active like IoT sensors, these technologies provide discrete tracking and facilitate seamless data capture at critical checkpoints, such as during material processing or delivery [5].
- **HSI systems**:
  Automatic data acquisition technologies integrate seamlessly with HSI systems to enable precise quality assessments and detailed traceability at a granular level. These systems capture and process critical data related to the composition and quality of bio-based feedstocks, supporting accurate monitoring and compliance with sustainability standards. The collected data is then securely transmitted to the blockchain using robust APIs, which guarantee data immutability and provide real-time accessibility for stakeholders. This integration strengthens the transparency and reliability of the T&T system, enhancing decision-making and accountability across the supply chain.

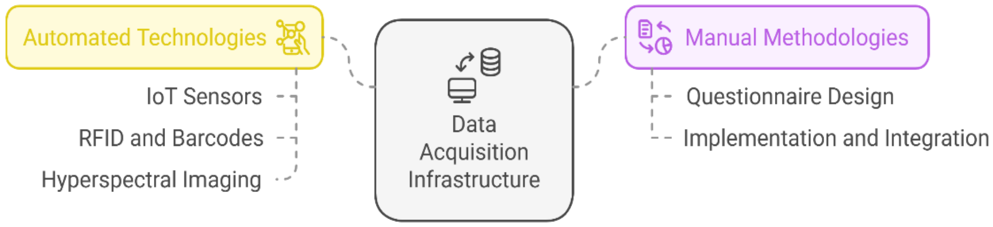

Figure 5: Data acquisition modes

Manual data collection through questionnaires

In the BioReCer project, manual data collection via questionnaires serves as an important method for gathering essential information, especially in scenarios where automated data acquisition technologies are not yet accessible to all stakeholders. This approach enables comprehensive data capture across diverse participants, facilitating the integration of varied data sources into the centralized T&T system.

- **Design and structure of questionnaires**
  The questionnaires are meticulously crafted to extract detailed information pertinent to each stage of the supply chain. They encompass sections that solicit data on company registration, product specifications, processing methods, logistics, and transportation details. This structured design enables the collection of both quantitative and qualitative data, providing a holistic view of the supply chain dynamics [6].
- **Implementation and data integration**
  To streamline the data collection process, all questionnaires are hosted on the BioReCer ICT tool. This centralized platform allows stakeholders to input their responses efficiently, promoting uniformity and ease of access. The WP3 team queries the required data directly from the ICT tool, integrating the manually collected information into the centralized blockchain database. This approach maintains data integrity and helps seamless interoperability with other components of the T&T system, enabling efficient tracking, analysis, and reporting.

Best practices and lessons from external experiences:

Drawing from established best practices in questionnaire-based data collection, several key considerations have been incorporated to enhance the effectiveness of this approach:

- **Questionnaire design**: Maintaining clear and simple question wording to reduce respondent confusion and enhance data quality. Questions are structured to be easily understandable, avoiding complex terminology and leading questions [6].
- **Pilot testing**: Conducting pilot tests with a small subset of respondents to identify potential issues in the questionnaire design and to refine questions for better clarity and relevance [7]
- **Training for enumerators**: Providing comprehensive training to individuals administering the questionnaires to make consistency in data collection and to address any respondent queries effectively [8].
- **Data quality assurance**: Implementing checks and validation mechanisms within the electronic application to detect and rectify inconsistencies or errors in the responses [9]

By integrating these best practices, the WP3 aimed to enhance the reliability and validity of the data collected through questionnaires, making manual data collection a robust component of the overall T&T system.

Innovative practices in data acquisition

In cases where advanced infrastructure is available, WP3 anticipated leveraging several innovative practices to enhance data acquisition efficiency and accuracy.

- Real-time monitoring systems:

One key approach is the implementation of real-time monitoring systems that combine IoT sensors with real-time analytics. These systems allow stakeholders to continuously monitor feedstock conditions and receive automated alerts for any deviations, such as temperature breaches or delays, enabling prompt corrective actions [10].

- Dynamic questionnaires:

Dynamic questionnaires are designed to adapt based on specific supply chain requirements. This flexibility shows that even when advanced technology is unavailable, essential data is consistently captured, filling critical information gaps in the T&T system.

- Edge computing for localized processing:

The use of edge computing at data acquisition nodes is proposed to process information locally, reducing latency and facilitating immediate responses to anomalies [11].

These practices collectively enhance the robustness and scalability of the data acquisition infrastructure, confirming comprehensive coverage across diverse operational scenarios.

2.2.3 Data transmission and storage

In a bio-based supply chain characterized by diverse stakeholders and complex workflows, safeguarding the integrity and accessibility of data is paramount. By adopting a centralized blockchain model, the BioReCer framework addressed challenges such as data tampering, unauthorized access, and inconsistent record-keeping. This model provided robust data protection while maintaining transparency and ensuring that all system components—ranging from IoT-enabled sensors to sector-specific tools—are seamlessly interconnected. This subsection majorly focuses on the mechanisms and protocols implemented to guarantee secure data transmission and reliable storage. It explains how advanced cryptographic techniques and blockchain functionalities facilitate data integrity, immutability, and accessibility, providing a dependable foundation for supply chain traceability and compliance. The discussion progresses through secure data transmission protocols, the role of centralized blockchain, and the integration of advanced storage infrastructure.

2.2.3.1 Secure data transmission protocols

In the WP3 T&T framework, secure data transmission plays a critical role in maintaining the integrity and reliability of the centralized blockchain system. Although the primary responsibility for managing secure transmission lies with WP5's ICT tool, WP3 gives assurance that the data retrieved from this platform is validated, securely stored, and protected from tampering during its lifecycle within the T&T system. By leveraging robust integration practices and blockchain-enabled mechanisms, WP3 guarantees that the transmitted data contributes to a transparent and trustworthy supply chain.

Interaction with the ICT Tool

WP5's ICT tool handles the secure transmission of data from various sources, including IoT devices, manual questionnaires, and sector-specific tools. This platform employs industry-standard encryption protocols, access control mechanisms, and authentication systems to maintain data security during transmission. WP3 interacts with the ICT tool to retrieve processed and structured data for further validation and storage within the centralized blockchain.

Data validation and integration

Once data is retrieved from the ICT tool, WP3 employs cryptographic techniques to have its authenticity and integrity:

- Hash verification: All incoming data is hashed using cryptographic algorithms, such as SHA-256, to detect any modifications during transmission.
- Integrity Checks: The system verifies that data conforms to predefined formats and accuracy thresholds before it is logged in the blockchain.

These steps make the blockchain only records accurate and trustworthy data, maintaining the reliability of the entire T&T framework.

Secure storage within the blockchain

After validation, data is immutably stored in the centralized blockchain, providing long-term security and transparency:

- Immutable records: The blockchain architecture confirms that once data is logged, it cannot be altered retroactively.
- Role-based access: Data stored in the blockchain is accessible only to authorized stakeholders, with permissions controlled by WP3's integration mechanisms.

Collaborative protocols with WP5

WP3's reliance on WP5's secure transmission capabilities creates system-wide security. Key collaborative practices include:

- Seamless data exchange: APIs facilitate efficient and secure transfer of validated data between the ICT tool and the blockchain.
- Alignment with standards: The system adheres to ISO/IEC 27001, GDPR guidelines, and NGSI-LD, ensuring compliance with international data security and interoperability standards.

By taking advantage of WP5's secure data pipelines and focusing on validation and storage, WP3 guarantees that the T&T system remains robust, reliable, and aligned with BioReCer's overarching goals of sustainability and transparency.

2.2.3.2 Centralized blockchain for data integrity and immutability

The adoption of a centralized blockchain model offers a robust mechanism for maintaining data integrity, immutability, and accessibility, tailored to meet the unique challenges of bio-based supply chains. This section delves into the core functionalities, practical implementation, and benefits of the centralized blockchain model in ensuring the reliability of the T&T system, while addressing sector-specific challenges and operational requirements.

Key features of the centralized blockchain:

A) Immutability of data records

The centralized blockchain framework employed by WP3 facilities all data entries are immutable, meaning they cannot be altered or deleted once logged. This is achieved through cryptographic hashing algorithms such as SHA-256, which create unique digital fingerprints for each data entry (Figure 6). Any modification to the data would produce a completely different hash, immediately flagging inconsistencies [12].

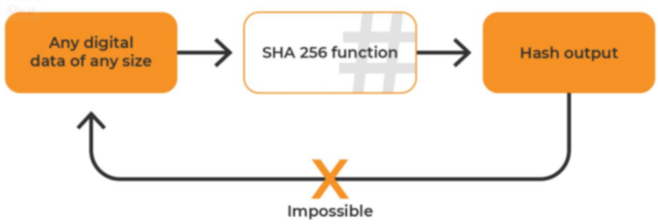

Figure 6: Hash functions

The blockchain's design incorporates a sequential, linked record structure where each block references the hash of the previous block. This demonstrates the integrity of the entire chain, as altering one block would invalidate the entire chain. Such immutability is critical for compliance with regulatory standards and for developing trust among stakeholders by guaranteeing that data remains authentic and verifiable.

B) Automated data validation with smart contracts

Smart contracts, pre-programmed with specific conditions, automate data validation processes within the blockchain. Smart contracts, deployed on a blockchain, play a transformative role in automating and improving supply chain processes. These computer programs execute automatically when predefined conditions are met, ensuring reliability and efficiency without requiring intermediaries for validation. In a supply chain context, smart contracts offer significant advantages by enhancing traceability, compliance, and operational efficiency. For instance, temperature readings from IoT sensors can be automatically checked against predefined thresholds. By eliminating manual validation steps, smart contracts reduce the risk of human error, increase operational efficiency, and provide real-time assurance of data accuracy [13].

Adapting smart contract interactions to a centralized blockchain

In the centralized blockchain model, the blockchain acts as the single authoritative hub for hosting and executing all smart contracts. Instead of a decentralized network where each node runs smart contracts independently, the centralized blockchain operates under controlled administration (WP3 or a designated authority in BioReCer).

- Blockchain node: A centralized node is responsible for hosting the blockchain and validating smart contract executions.
- Centralized ledger: All transactions and smart contract outcomes are recorded on a single ledger, which assure data integrity and transparency for all stakeholders.

Figure 7 shows smart contract automation and integration across supply chain stakeholders.

Smart contracts across the supply chain

- **Supplier interactions**:
  Smart contracts govern the submission of product details, such as certifications, batch numbers, and quality parameters. A supplier uploads a batch of biomass data, triggering a smart contract that verifies compliance with sustainability criteria.
- **Manufacturer interactions**:
  Other side, smart contracts automate the exchange of raw material data, allowing the manufacturer to only process compliant inputs. The manufacturer can receive a notification when required documentation (e.g., certifications) has not been validated by the smart contract.
- **Logistics/storage interactions**:
  Contracts manage data related to transportation, such as geolocation, storage conditions, and transit times. As an example, if temperature sensors detect a breach in cold chain requirements, a smart contract can automatically alert stakeholders and log the event.
- **Consumer interactions**:
  Final products are tagged with immutable data records, enabling consumers to trace their origins and quality attributes through the blockchain. A smart contract validates the final product’s conformity with sustainability standards before it is logged for consumer access.

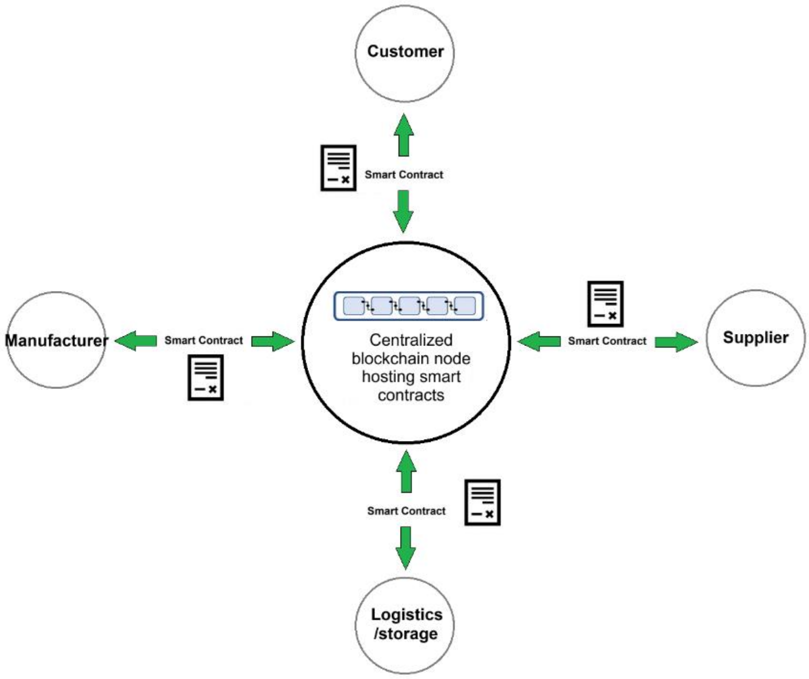

Figure 7: Centralized blockchain enabling smart contract automation and integration across supply chain stakeholders.

C) Traceability and transparency

The centralized blockchain offers end-to-end traceability by maintaining an immutable ledger of all supply chain events. Each transaction is time-stamped and linked to specific stakeholders, enabling the reconstruction of the product's journey from origin to end use. This transparency supports certification processes, simplifies audits, and enhances consumer confidence in bio-based products.

Example in case in fisheries:

In fisheries, the blockchain tracks data such as the vessel's location, the fishing method used, and the storage conditions of the catch. When a batch of fish is caught, its details are immediately logged on the blockchain, including the date, time, and exact coordinates of the catch. As the fish moves through processing, transportation, and retail stages, additional data such as temperature control logs and packaging details are appended. If the final product is certified under the certification scheme, then the information related to the process and validity are included. This assists consumers to verify the sustainability of their seafood by tracing its journey from ocean to table, fostering trust and meeting certification requirements as well as regulatory compliance standards like the EU's Common Fisheries Policy.

D) RBAC

To confirm data confidentiality, the blockchain system incorporates RBAC mechanisms. Stakeholders such as suppliers, manufacturers, and regulators are granted access to specific data segments based on their roles and permissions. Therefore, sensitive information is accessible only to authorized parties, maintaining a balance between transparency and privacy.

Practical implementation of the centralized blockchain

Figure 8 illustrates the sequential stages of data processing within the centralized blockchain, from initial data capture through sensors and questionnaires to secure storage, ensuring data integrity and traceability throughout the supply chain.

- **Data logging and cryptographic hashing**

  Data from diverse sources, including IoT sensors, manual questionnaires, and sector-specific tools, is collected via the BioReCer ICT tool. Before being logged into the blockchain, each data entry undergoes a hashing process using algorithms such as SHA-256. The resulting hash serves as a unique identifier, ensuring the integrity of the data.

  The implementation involves the following steps:
  - Data capture: Sensors and questionnaires transmit data to the ICT tool.
  - Hash generation: Each data entry is hashed and linked to metadata, such as timestamps and stakeholder IDs.
  - Blockchain entry: The hashed data is appended to the blockchain, preserving its immutability.

    This process creates a secure and tamper-proof audit trail, essential for meeting regulatory and sustainability requirements.
- **Smart contract workflows**

  Smart contracts are integral to the blockchain's functionality, automating critical processes such as compliance checks and data validation. Examples include:
  - Real-time compliance monitoring: Verifying that environmental parameters (e.g., CO_2 emissions, water usage) fall within acceptable ranges.
  - Automated notifications: Triggering alerts when anomalies are detected, such as deviations in storage conditions.

  These workflows streamline operations, reduce administrative burdens, and confirms that data logged in the blockchain is both accurate and compliant.

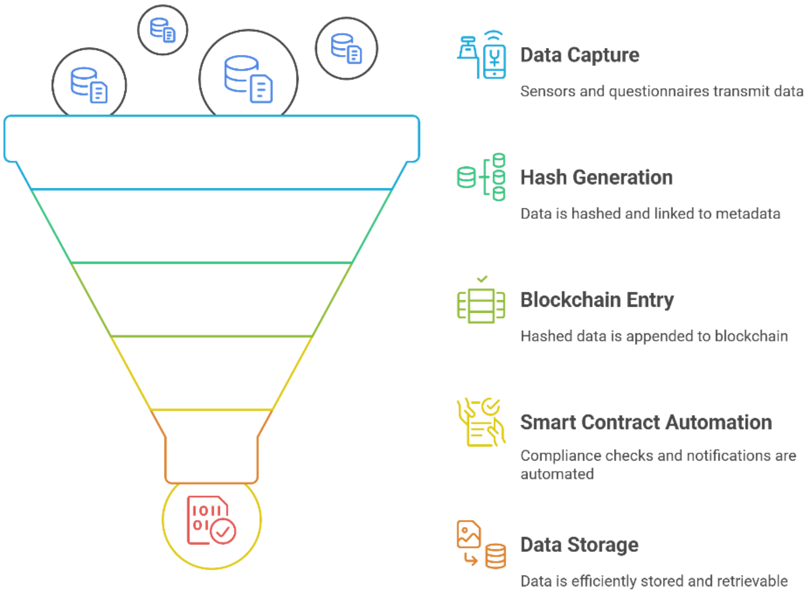

Figure 8: Data processing in centralized blockchain

- **Data storage and retrieval**
  The blockchain's architecture supports efficient storage and retrieval of data. Key features include:
  - Indexed search capabilities: Stakeholders can quickly locate specific records using metadata such as timestamps or transaction IDs.
  - Scalable storage: The centralized model allows for seamless integration of additional datasets as the T&T system expands.
  - Backup and recovery protocols: Regular backups keep data availability and resilience against potential system failures.

Advantages of the centralized blockchain model

- **Efficiency**
  Compared to decentralized models, the centralized blockchain employed by BioReCer offers faster transaction processing and reduced computational overhead. This is particularly advantageous for real-time applications, such as monitoring perishable bio-feedstocks.
- **Security**
  By centralizing data management while maintaining blockchain's core security features, the model minimizes attack vectors. Cryptographic hashing and smart

contract validations further enhance the system's resilience against tampering and cyberattacks.

- **Cost-effectiveness**
  The centralized approach reduces the resource-intensive consensus mechanisms required in decentralized systems, lowering operational costs without compromising on transparency or integrity.
- **Scalability**
  The centralized blockchain is designed to accommodate the growing data demands of the BioReCer project. Its modular architecture supports the integration of new datasets, tools, and stakeholders, for long-term applicability.

2.2.3.3 Storage infrastructure

The storage infrastructure of WP3's T&T system must be secure, efficient, and scalable for management of supply chain data and we need to design one to support long-term data retention and rapid access. The system integrates centralized blockchain technology with modern storage methodologies while maintaining compatibility with sector-specific tools.

Centralized blockchain integration for secure storage

The local centralized blockchain, hosted on a secure server, underpins BioReCer's storage infrastructure, providing an immutable and transparent record of all supply chain events. Each transaction is permanently logged on the blockchain, therefore, data cannot be altered or deleted, a feature crucial for regulatory compliance and fostering stakeholder trust.

- Data immutability: Records are protected through cryptographic hashing, maintaining their integrity and authenticity.
- Centralized efficiency: The centralized model simplifies data retrieval and ensures optimized performance for real-time operations.

Cloud-based redundancy and scalability

To support interoperability with WP5's ICT tool, the system leverages cloud-based solutions exclusively for secure data exchange and redundancy in API-driven interactions.

- Redundancy: Data is backed up across multiple data centres, minimizing the risk of loss during system failures.
- Scalability: The architecture accommodates new datasets and stakeholders as the BioReCer system expands.

- Regulatory compliance: Cloud solutions align with GDPR and ISO standards, keeping data security and privacy.

Data retrieval and search optimization

Efficient access to stored data is vital for stakeholders to perform timely analyses and meet compliance requirements. Advanced indexing and search capabilities ensure seamless data retrieval.

- Metadata indexing: Each record is tagged with key metadata (e.g., timestamps, batch numbers, and origins) for quick searches.
- RBAC: Stakeholders can access only authorized data, maintaining confidentiality and security.

Integration with sector-specific tools

The storage infrastructure is built to integrate with tools like IoT devices, HSI systems, and environmental monitoring sensors. This integration supports accurate data collection and facilitates comprehensive traceability.

Backup and disaster recovery protocols

The system incorporates robust protocols to safeguard data against loss or corruption.

- Automated backups: Periodic backups are created and securely stored offsite.
- Recovery testing: Regular simulations verify that recovery processes are effective and reliable.

Future-proofing the storage infrastructure

WP3's storage system is designed for adaptability, maintaining its effectiveness amid evolving technologies and regulatory requirements.

- Modular design: Supports new analytics tools and data formats.
- Energy efficiency: green data centre technologies align with BioReCer's sustainability goals. Our server infrastructure incorporates advanced liquid cooling technology, which is significantly more energy-efficient compared to traditional air-based cooling systems. By using liquid as a medium to transfer heat away from high-performance components, the system minimizes the energy required for temperature regulation. This innovation reduces overall energy consumption and aligns with BioReCer's commitment to sustainability, ensuring that our storage infrastructure operates efficiently while maintaining environmental responsibility.

- Interoperability: Accommodates emerging technologies like AI-based analytics and blockchain enhancements.

By combining centralized blockchain technology with cloud-based solutions, the system is secure, efficient, and future-ready data management. This infrastructure supports BioReCer's overarching objectives of transparency, traceability, and sustainability in bio-based supply chains.

2.2.4 Dynamic data processing

Dynamic data processing is enabling real-time analysis and decision-making throughout the supply chain. By integrating advanced methodologies and leveraging innovative technologies, the system ensures that stakeholders can monitor KPIs such as sustainability metrics, CIs, reliability of information, efficiency levels, overall eco performance measurement of supply chain and compliance in real-time. This section introduces the methodologies employed for real-time analytics and outlines how the system calculates and visualizes critical sustainability indicators.

2.2.4.1 Real-time analytics methodologies

The BioReCer T&T system employs advanced methodologies for real-time data processing and utilises Apache Spark to handle the dynamic and diverse nature of bio-based supply chains. These methodologies assist efficient analysis of large-scale data streams, ensuring timely insights for stakeholders and robust decision-making capabilities.

Apache Spark and stream processing frameworks

Apache Spark serves for stream processing in the T&T system, efficiently handling real-time data retrieved from the ICT tool via APIs. Its Structured Streaming module provides a scalable and fault-tolerant stream processing engine built on the Spark SQL engine, allowing for seamless computation on streaming data in a manner similar to batch operations. This approach enabled WP3 to process and analyse pre-structured data provided by WP5, helping rapid insights and decision-making [14].

Integration with APIs

WP3's T&T system integrates multiple data sources from WP5's ICT tool using API-driven data exchange. This ensures that both real-time automated data streams (IoT, sensors, and imaging systems) and manually entered datasets (questionnaires, certification data, and stakeholder inputs) are incorporated into the system for sustainability tracking and compliance verification.

How the integration works:

1. Automated Data Retrieval (IoT & Sensors):
    - The ICT tool provides structured datasets from IoT sensors, RFID, barcodes, and HSI imaging systems through API endpoints.
    - WP3's system queries these APIs at predefined intervals or receives pushed data from the ICT tool.
    - Upon retrieval, Apache Spark's Structured Streaming processes the data in near real-time, performing validation, transformation, and aggregation before storing it in the blockchain infrastructure.
2. Manual Data Entry & Questionnaires:
    - Not all data in the T&T system is sensor driven. Stakeholders manually input key traceability data through the ICT tool, including self-reported sustainability metrics, conformity assessments, and regulatory compliance information.
    - These datasets are then transferred via APIs from the ICT tool to WP3's system for further processing.
    - WP3 conducted tests using existing questionnaire responses to verify the system's ability to process both automated and manual data entry methods.
3. Processing and storage:
    - Regardless of whether the data comes from automated sensors or manual input, WP3's system performs standardization and validation to prevent discrepancies.
    - The processed data is then stored in the blockchain.
4. Future scalability and adaptability:
    - Although IoT devices are not yet directly connected to the ICT tool, WP3 tested integration using dummy and simulated IoT data streams to emulate real-world conditions.
    - Similarly, for manual data, WP3 collaborated with WP2 to validate the handling of questionnaire responses, ensuring that the system can accommodate both structured sensor data and unstructured human inputs.

This hybrid approach allows the T&T system to integrate automated real-time monitoring with stakeholder-reported sustainability data, ensuring comprehensive traceability across the bio-based supply chain.

Real-time data processing

Using Apache Spark's Structured Streaming, WP3 was able to process large-scale, high-throughput data streams provided by the ICT tool while maintaining fault tolerance and reliability. The processed data can be stored or used to generate actionable insights for compliance monitoring, supply chain optimization, and dynamic visualization dashboards [15].

The primary goal of this implementation is to leverage Apache Spark's Structured Streaming module for:

- Retrieving pre-processed data from the ICT tool via APIs.
- Performing real-time analytics on incoming sensor data (e.g., temperature and humidity).
- Computing dynamic metrics such as running averages for environmental parameters, which are critical for compliance monitoring and operational optimization.

Below is a Python programming language example demonstrating how to use Spark Structured Streaming to query data from an API:

```
from pyspark.sql import SparkSession
from pyspark.sql.functions import col, avg, from_json
from pyspark.sql.types import StructType, StructField, StringType, FloatType
import requests
import json

# Initialize Spark session
spark = SparkSession.builder \
    .appName("BioReCerAPIBasedStreaming") \
    .getOrCreate()

# Define schema for incoming JSON data
schema = StructType([
    StructField("sensor_id", StringType(), True),
    StructField("temperature", FloatType(), True),
    StructField("humidity", FloatType(), True),
    StructField("timestamp", StringType(), True)
])

# API endpoint for querying the ICT tool
```

```
API_URL = "http://ict-tool.example.com/api/v1/sensor-data"

# Function to fetch data from ICT Tool API
def fetch_data_from_api():
    response = requests.get(API_URL)
    if response.status_code == 200:
        return response.json()
    else:
        raise Exception(f"API request failed with status code {response.status_code}")

# Fetch data from API and convert to RDD
def get_streaming_data():
    raw_data = fetch_data_from_api()
    return spark.sparkContext.parallelize([json.dumps(record) for record in raw_data])

# Read API data as a streaming DataFrame
stream_data = spark.readStream.schema(schema).json(get_streaming_data())

# Parse the data for analysis
parsed_data = stream_data.select(
    col("sensor_id"),
    col("temperature"),
    col("humidity"),
    col("timestamp")
)

# Calculate running averages for temperature and humidity
aggregated_data = parsed_data.groupBy("sensor_id") \
    .agg(
        avg("temperature").alias("avg_temperature"),
        avg("humidity").alias("avg_humidity")
    )

# Output the results to the console
query = aggregated_data.writeStream \
    .outputMode("complete") \
    .format("console") \
    .start()

query.awaitTermination()
```

Key features of the code

1. **API integration**:
    - The script fetches pre-processed data from the ICT tool using HTTP GET requests.
    - This approach manages a seamless integration with WP5's infrastructure and eliminates redundant data ingestion layers.
2. **Real-time data processing**:
    - Apache Spark's Structured Streaming processes incoming data as it is retrieved, enabling near-instantaneous computations and insights.
    - Example: Real-time monitoring of average temperature and humidity per sensor ID.
3. **Scalability and fault tolerance**:
    - Spark's distributed architecture possesses scalability to handle large volumes of data, making it suitable for BioReCer's growing dataset requirements.
    - Built-in fault tolerance guarantees data integrity and consistency even during system disruptions.

Advantages of the approach

The use of Apache Spark for real-time analytics in the WP3 T&T system offers several advantages. First, it enhances decision-making by providing real-time insights that allow stakeholders to respond proactively to anomalies, such as deviations in temperature or humidity during transport. This capability ensures that potential issues are addressed promptly, reducing the risk of any quality degradation.

Operational efficiency is another key benefit, as the integration of data from the ICT tool and its subsequent processing through Spark minimizes latency. This streamlined approach ensures that actionable insights are delivered in a timely manner, enabling informed decisions across supply chain operations. Additionally, the system strengthens compliance and traceability by preparing processed data for storage in the centralized blockchain. This helps to have transparent and immutable traceability records, ensuring alignment with regulatory standards and fostering stakeholder trust.

2.2.4.2 Computing CIs

CIs are quantitative metrics designed to assess the efficiency of resource use within a system and its alignment with circular economy principles. These indicators evaluate the extent to which materials are reused, recycled, and retained within the production lifecycle,

minimizing waste and dependency on non-renewable inputs. By providing actionable insights, CIs support the development of sustainable practices and compliance with environmental and regulatory standards.
WP3 is responsible for the digitalisation of the framework developed by WP2 (Environmental sustainability assessment: Diagnosis and mapping of current indicators and data collection of biological resources), enabling real-time and historical data processing for circular indicators. The formulas provided by WP2, as presented in **D2.3 - Circular indicators**, form the basis for these calculations. This framework integrates seamlessly into the T&T system and enables the visualization of approximately 40 indicators critical to assessing circularity.

CIs framework

WP3 is tasked with calculating a wide array of CIs, grouped by roles within the supply chain. The database (in an Excel file format) consists of three primary sheets including questions for primary producers, processor manufacturer and waste managers which focus on gathering foundational data from stakeholders and address data related to the supply chain and concentrate on post-consumer processes and end-of-life handling of materials.
This database contains structured questions designed to identify the stakeholder's role and gather data essential for indicator computation. For example, the question, "What is your role in the supply chain for the product you are assessing?" requires the user to select an option from a dropdown menu, such as:

- Primary Producer
- Processor (Intermediate Pre-Consumer Producer)
- Manufacturer (Final Pre-Consumer Producer)
- Waste Managers/End-of-Life Processors (Post-Consumer Producer)

This structured approach confirms the standardization of input data and provides a robust foundation for calculating CIs.

Role of WP5 and the ICT tool

To streamline data collection, all questions and dropdown response options from the Excel file are integrated into the ICT tool managed by WP5. Stakeholders interact with this tool to input their responses, ensuring data consistency and accessibility. WP3 then queries this pre-validated data via APIs, minimizing manual intervention and enabling seamless integration into the T&T application.

Implementation of circularity calculations

WP3 developed an advanced computational tool capable of extracting, interpreting, and applying the formulas provided in D2.3 for CIs. These indicators, as defined in D2.3, are incorporated into the T&T system, allowing automated assessment of circularity performance across the supply chain. These CIs, computed and calculated through WP3's advanced computational tool, play an important role in aligning BioReCer's supply chain with circular economy principles. By providing real-time insights into CIs, the tool ensures that all stakeholders can make informed decisions to enhance the sustainability of bio-based supply chains.

Tool adaptability and future-proofing

The computational tool developed by WP3 incorporates features that verify its adaptability and scalability:

- **Dynamic formula integration**:
  The tool is designed to "soak" the formulas provided by WP2. This architecture allows for future updates to the formulas and calculation with minimal effort. WP2 can directly update the formulas within the shared framework, and the T&T system will adapt automatically without significant reprogramming.
- **Efficient code structure**:
  The computational tool developed by WP3 has a highly efficient code structure, utilizing modular and nested coding practices to manage the intricate dependencies present in the circularity formulas provided by WP2. Extracting nested formulas from complex databases like Excel files is inherently challenging due to their layered nature. Often, formulas in one cell reference other cells, which in turn contain their own formulas. This creates a cascading chain of dependencies, making the extraction and computation of results not only labour-intensive but also prone to errors and difficulties.
  To address and solve this complexity, WP3 engineered a solution that automates the parsing, interpretation, and execution of these nested formulas. By breaking down the dependencies into smaller, manageable modules, the tool ensures that the calculations are both computationally efficient and error-resistant. The modular approach allows the system to follow each formula's chain of references systematically, ensuring all related calculations are completed accurately before consolidating the final result.
  A key feature of the tool is its ability to adapt to future updates. Since the formulas are nested and interdependent, any changes made to a cell or its references in the

Excel file could potentially disrupt the entire calculation flow. WP3's tool overcomes this challenge by dynamically detecting and incorporating such updates without requiring extensive manual reprogramming. This adaptability helps that the system remains resilient and scalable as WP2 modifies or adds new indicators to the circularity assessment framework.

By automating the interpretation of nested formulas and ensuring compatibility with potential updates, WP3 has significantly enhanced the computational efficiency of circularity calculations. This innovation not only streamlines the process for current data but also future-proofs the system, providing a robust foundation for ongoing sustainability assessments within BioReCer's T&T system.

– **Real-time updates**:

As stakeholders input data into the ICT tool, the computational tool processes it in real-time, updating CIs dynamically and allows stakeholders to have access to the most accurate and up-to-date information.

Visualization and stakeholder access

The calculated CIs are transmitted back to the ICT tool, where they are displayed on interactive dashboards. These dashboards allow stakeholders to:

– Monitor compliance with circular economy goals.
– Identify areas for improvement within the supply chain.
– Access detailed reports for regulatory submissions and internal assessments.

By using WP2's formula-based framework and integrating it into the T&T system, WP3 verifies that BioReCer's CIs are calculated accurately and efficiently. This collaborative approach supports compliance with environmental standards and enhances the transparency and accountability of bio-based supply chains. The adaptability of the computational tool further ensures that the system remains robust and responsive to evolving requirements, reinforcing BioReCer's commitment to sustainability and innovation.

2.3 Reporting on reliability, conformity, and performance

With T&T, all stakeholders will have access to comprehensive reporting mechanisms that are designed to provide transparency, accountability, and informed decision-making. Through these reporting tools, stakeholders can monitor the system's reliability, compliance with regulatory standards, and overall performance. This section outlines the methodologies, key indicators, and tools used to generate dynamic reports, ensuring the system aligns with BioReCer's sustainability and traceability objectives.

2.3.1 Reliability reporting

Reliability reporting evaluates the robustness and dependability of data contributions across the bio-based supply chain. This framework ensures that the system remains effective, consistent, and transparent, even as data sources and stakeholder engagement vary. Reliability is measured by assessing the completeness, accuracy, and timeliness of data provided by stakeholders and the system's ability to deliver dependable insights.

Key components of reliability reporting

1. Data Availability:

   The reliability of the T&T system is heavily influenced by the availability of required data. Data must be retrieved in a timely and complete manner from stakeholders and ICT tools.

   Metrics include:
   - Percentage of completed questionnaires.
   - Frequency of missing or incomplete data points in datasets.

2. Stakeholder Participation:

   The level of stakeholder engagement in responding to questionnaires and providing requested data directly impacts the reliability of the system.

   A stakeholder reliability score is proposed to quantify participation:

$$\boldsymbol{R_s} = \frac{\boldsymbol{C}}{\boldsymbol{T}} \times \mathbf{100} \qquad \text{(Eq. 1)}$$

   where $\boldsymbol{R_s}$ is the reliability score for a stakeholder, $\boldsymbol{C}$ is the number of completed data points, and $\boldsymbol{T}$ is the total required data points.

3. System Uptime and Data Synchronization:

   Reliability is also defined by the technical stability of the T&T system. Metrics such as API uptime, response times, and synchronization efficiency are monitored to ensure that data flows seamlessly from WP5's ICT tool to WP3's analytics framework.

Proposed formula for overall reliability:

A unified reliability metric $\boldsymbol{R_o}$ can be calculated by combining multiple dimensions of reliability:

$$\boldsymbol{R_o} = \boldsymbol{w_1 R_s} + \boldsymbol{w_2 R_a} + \boldsymbol{w_3 R_t} \qquad \text{(Eq. 2)}$$

Where:

R_s: Stakeholder participation reliability.

R_a: Data availability reliability (percentage of complete datasets received).

R_t: Technical reliability (system uptime, latency metrics).

w_1, w_2, w_3: Weights assigned to each component based on their relative importance.

By factoring all aspects of reliability into one comprehensive score, stakeholders will gain a clear understanding of system performance.

Impact of non-participation on reliability

Non-participation or incomplete data submissions by stakeholders significantly reduce the T&T system's reliability. For instance:

- A stakeholder failing to complete the required questionnaires lowers the R_s score.
- Missing critical data, such as material flow or recycling rates, compromises R_a, reducing the overall reliability score.

By integrating this into the reporting framework, stakeholders can see how their participation directly affects the system's performance and transparency.

Reporting Mechanisms

1. Web-based metric display
2. Reliability metrics are displayed through the web application within the ICT tool, providing stakeholders with real-time insights.
3. Automated alerts:

   Notifications are triggered when reliability scores fall below acceptable thresholds, encouraging corrective actions.
4. Regular reports:

   Periodic reports summarize reliability metrics and highlight areas requiring improvement, ensuring accountability across the supply chain.

By quantifying and visualizing reliability through key metrics, stakeholders are empowered to maintain high levels of participation and data integrity, ultimately aligning with the project's objectives of sustainability and traceability.

2.3.2 Conformity reporting

Conformity reporting within the T&T system enables companies to track their own compliance progress over time rather than relying on external benchmarks. This self-assessment approach allows companies to calculate their conformity level, implement improvements, and reassess at regular intervals, fostering continuous improvement.

Framework for conformity reporting

The WP3 T&T system integrates data validation methods and historical tracking to help companies measure their alignment with environmental and regulatory goals. Unlike rigid compliance frameworks, this system supports self-evaluation, allowing companies to compare their own past performance rather than industry-wide benchmarks. This self-assessment approach allows companies to calculate their conformity level, implement improvements, and reassess at regular intervals, fostering continuous improvement.

Self-assessment and continuous improvement

1. Companies calculate an initial conformity score based on their sustainability and regulatory performance.
2. After implementing improvement measures, they recalculate the score at a later stage to track progress.
3. This method enables organizations to identify trends and evaluate the effectiveness of their sustainability initiatives over time.

Proposed conformity calculation approach

Instead of a fixed benchmarking system, conformity is tracked dynamically using a self-referenced scoring model. Companies assess their own past performance rather than comparing against external thresholds.

Example approach to Self-Assessment Scoring:

- Companies assign an initial conformity score at Year 0.
- After a set period (e.g., one year), they recalculate the score, factoring in implemented sustainability measures.
- Progress is measured by comparing the new score against the previous score to assess improvement.

Reporting mechanisms

- Self-assessment reports: Companies receive tailored reports comparing their current vs. past conformity levels.
- Progress tracking web app: Instead of displaying a pass/fail compliance metric, the web app shows conformity trends over time.

Key benefits of the self-assessment model

The self-assessment model offers several key benefits for stakeholders. It encourages gradual improvement by allowing companies to track their progress over time rather than following rigid compliance enforcement. This approach provides flexibility in monitoring sustainability goals without relying on external benchmarking, enabling organizations to focus on their own operational improvements. Additionally, it enhances transparency, helping companies understand how their actions impact overall sustainability performance and empowering them to make informed decisions for continuous development.

2.3.3 Overall performance reporting

The concept of overall performance represents a comprehensive evaluation of how well a product or process aligns with the project's sustainability, reliability, and conformity goals. It consolidates multiple performance indicators into a single, meaningful metric that stakeholders can use to assess and improve their operations. By taking advantage of advanced computational techniques, the overall performance metric provides an integrated view of circularity, reliability, and conformity, ensuring transparency, traceability, and alignment with regulatory standards.

Definition and purpose

The overall performance metric is calculated using the Customer Optimization Route and Evaluation (CORE) evaluation framework, which integrates data from:

- CIs: Evaluating the sustainability of materials and processes across the supply chain.
- Reliability metrics: Assessing the completeness and accuracy of data provided by stakeholders.
- Conformity metrics: Measuring compliance progress over time using self-referenced tracking.

This unified approach ensures that all critical dimensions of performance are captured, allowing stakeholders to make informed decisions and identify areas for improvement. The CORE structured evaluation framework, as a mathematical backbone, translates these complex inputs into a single satisfaction value, making it easier to interpret and act upon [16]. CORE serves as a structured methodology to bridge the gap between stakeholder expectations and system outputs by quantifying qualitative performance metrics. This model enables a systematic approach to measuring how well the implemented T&T framework aligns with the needs of key actors in the supply chain.

Methodology

The calculation of the overall performance metric involves the following steps:

1- Data Collection:
   - CIs, such as recycling rates and resource efficiency, are computed based on WP2's predefined methodologies (D2.3) and real-time/historical data shared via WP5's ICT tool.
   - Reliability scores are determined by analysing the completeness of stakeholder responses to questionnaires and the quality of data provided.
   - Conformity metrics are derived by assessing operational parameters over time, enabling companies to track their compliance progress through self-referenced evaluations.

2- Normalization:
   Each input is normalized to a scale (e.g., 0 to 1) to ensure compatibility and fair comparison across diverse metrics.

3- Applying of the CORE model:
   The CORE framework is applied to CIs to compute a unified circularity score.

4- CORE-Based Aggregation:
   The CORE model serves as the aggregation mechanism, combining individual satisfaction values from circularity, reliability, and conformity assessments into a single overall performance score.

5- Visualization:
   The calculated overall performance metric is returned and displayed on the ICT tool through interactive dashboards, providing stakeholders with a clear and actionable view of their performance.

The Role of the CORE formula

At the heart of the overall performance metric is the CORE formula, a sophisticated computational tool that ensures accuracy and adaptability. By dynamically adjusting to different boundary conditions and penalizing significant deviations, the CORE provides a robust and reliable framework for performance evaluation. Its flexibility allows for the integration of new indicators or adjustments to existing benchmarks, maintaining its relevance over time.

For example, in a practical scenario:

- A fishery product's CIs, such as recycling rates and energy efficiency, are calculated based on IoT and questionnaire data.
- The reliability of the data is assessed by evaluating the completeness of stakeholder questionnaires and the consistency of sensor inputs.
- Conformity is measured by comparing operational parameters, such as water usage, against EU regulatory standards.

The CORE formula synthesizes these inputs into a single satisfaction value, which is then displayed alongside explanatory visualizations [16].

Advantages of the approach

1. Holistic evaluation: The overall performance metric provides a multidimensional assessment framework that captures the interconnected aspects of sustainability, reliability, and regulatory compliance across bio-based supply chains. By aggregating data from diverse indicators—such as circularity metrics, conformity scores, and stakeholder reliability—it delivers a comprehensive evaluation of how well products and processes align with BioReCer's sustainability goals. This metric ensures a balanced approach, taking into account both the environmental performance of materials and processes and the operational and regulatory commitments of stakeholders. This integrated assessment offers a clear picture of supply chain dynamics, give an opportunity to stakeholders to identify and address inefficiencies while celebrating areas of success.
   Incorporating the CORE formula, the overall performance metric adjusts dynamically based on input parameters, ensuring relevance across varying supply chain contexts. For example, the weighting of CIs or reliability scores can shift to align with sector-specific priorities or regulatory changes. The system's adaptability allows for continuous improvement, driving better alignment with evolving sustainability standards. Furthermore, the holistic perspective provided by the metric encourages stakeholders to take collective responsibility for enhancing supply chain transparency and efficiency, strengthening trust and compliance throughout the T&T framework [17].

2. Actionable insights: A unified metric allows stakeholders to promptly pinpoint areas of strength and weakness within the supply chain, providing a clear overview of performance across various dimensions. By consolidating data from CIs, reliability measures, and conformity assessments, the metric offers a comprehensive perspective that simplifies complex evaluations. This clarity empowers stakeholders to prioritize targeted interventions, whether by addressing inefficiencies, optimizing resource usage, or ensuring regulatory compliance. Over time, such focused efforts contribute to a culture of continuous improvement, driving innovation and fostering resilience within the bio-based supply chain. Additionally, the unified metric serves as a benchmark for performance, enabling comparisons across time periods, products, or supply chain segments, thus encouraging accountability and transparency.
3. Alignment with BioReCer goals: The methodology ensures that all aspects of the supply chain align with the project's objectives of sustainability, circularity, traceability, and transparency.
4. Future-proof design: The system's adaptability allows for the integration of new indicators or changes to existing criteria, ensuring its continued relevance as stakeholder needs evolve.

The methodology for overall performance reporting can be expanded to include:

Predictive analytics: Using machine learning models to forecast future performance trends based on historical data.

Sector-Specific customization: Tailoring the weighting factors and boundary conditions of the CORE formula to reflect the unique needs of different bio-based sectors.

Real-Time feedback: Enhancing the ICT tool with real-time alerts and recommendations based on the overall performance metric.

By combining advanced computational techniques with a clear and actionable framework, the overall performance metric serves as a cornerstone of BioReCer's T&T system, driving sustainability, transparency, and continuous improvement across bio-based supply chains.

3. Technical specifications

The technical foundation of the BioReCer T&T system is designed to support its advanced functionalities, ensuring efficiency, scalability, and reliability. This section provides a comprehensive overview of the system's underlying architecture, data management strategies, and cloud-based integration. By detailing the technical specifications, the chapter highlights how the system meets the demands of real-time analytics, stakeholder collaboration, and sustainability reporting. Each section delves into the key components that enable seamless operation and adaptability to evolving requirements within bio-based supply chains.

3.1 Software architecture and design

This section describes the overall architecture of the BioReCer T&T system, focusing on its modular design, technical workflows, and integration strategies. It emphasizes how the software architecture supports real-time data processing, stakeholder interoperability, and scalability.

3.1.1 Overview of the architecture

The BioReCer T&T system is designed as a robust, modular, and scalable platform tailored to the unique demands of bio-based supply chains. Its architecture supports seamless data integration, advanced analytics, and dynamic reporting which ensures stakeholders have access to accurate and actionable insights. The system bridges multiple domains, integrating environmental, operational, and regulatory data to promote sustainability, transparency, and compliance.

At its core, the system relies on a centralized blockchain infrastructure that provides immutable data storage and secure traceability across the supply chain. This blockchain serves as the backbone for maintaining transparency and accountability, gives permission to stakeholders to verify data integrity and monitor supply chain activities in real time. Data flows into the system from WP5's ICT tool via APIs, which act as conduits for structured and validated data. This integration ensures that WP3 processes only high-quality, pre-validated inputs for real-time computations and reporting.

The modular design of the T&T system enhances its flexibility and adaptability. Key modules include:

- **Data collection and acquisition module**: Gathers and retrieves data from the ICT tool, IoT devices, and other external sources, ensuring seamless interoperability.

- **Processing engine**: Powered by Apache Spark, this module handles real-time data analytics, CI computations, and performance evaluations.
- **Storage layer**: Combines centralized blockchain and local storage components to ensure secure, efficient, and scalable data retention.
- **Reporting and visualization module**: Provides stakeholders with interactive dashboards and automated reports, accessible through WP5's ICT tool.

The software architecture also incorporates advanced security features, such as RBAC, encryption, and automated compliance checks. These measures protect sensitive data and ensure that the system aligns with regulatory frameworks like GDPR and ISO 27001.

The modular design, advanced analytics, and robust data management of the BioReCer T&T system make it a versatile solution for addressing the complexities of bio-based supply chains. Its architecture supports growth and adaptability, allowing for the seamless integration of new technologies, additional stakeholders, and larger data volumes without compromising performance or reliability.

3.1.2 Integration with external systems.

The T&T system is built to seamlessly interact with external systems, particularly WP5's ICT tool, which serves as the central repository for pre-validated datasets. This integration maintains the streamlined flow of information between various components of the BioReCer framework, supporting the calculation of CIs, compliance monitoring, and real-time decision-making.

APIs as communication bridges

The integration primarily relies on robust APIs to exchange data between WP3's T&T application and WP5's ICT tool. These APIs:

- Allow dynamic querying of datasets, guaranteeing that only relevant and updated data is retrieved for further analysis.
- Support the ingestion of real-time data streams from IoT devices and other data sources connected to the ICT tool, enabling timely updates in the T&T system.
- Facilitate bi-directional data exchange, where processed results, such as computed indicators or conformity metrics, are sent back to the ICT tool for stakeholder reporting.

Data validation mechanisms

To maintain data accuracy and consistency, the system incorporates validation checks during data exchange:

- Retrieved data undergoes pre-processing to standardize formats, removing any inconsistencies that might hinder further computations.
- Any anomalies detected in the data (e.g., missing fields or outliers) trigger automated alerts, prompting corrective actions before the data is utilized in calculations or reports.

Interoperability with stakeholder tools

The architecture supports compatibility with third-party systems used by stakeholders, such as certification platforms or compliance databases. For instance:

- Certification bodies can retrieve compliance metrics directly from the T&T system, reducing manual intervention.
- Manufacturers or suppliers can upload additional data to the ICT tool, which is then automatically integrated into the T&T system for performance evaluations.

Scalability for future integration

As the BioReCer project evolves, the system is equipped to handle additional integrations with minimal structural changes. For example:

- The modular API framework can incorporate new data sources or tools, such as advanced IoT sensors or machine learning models for predictive analytics.
- Stakeholder-specific modules can be added to enhance functionality, such as customized dashboards for individual partners.

This integration with external systems not only enhances the efficiency of the BioReCer framework but also provides a scalable and adaptable foundation for future expansions. It reinforces the project's objective to establish a unified and transparent supply chain management platform.

3.1.3 Data processing workflow

The data processing workflow in the BioReCer T&T system is meticulously designed to transform raw, diverse inputs into actionable insights for stakeholders. This workflow integrates real-time data retrieval, pre-processing, computation, and output generation, providing a systematic approach to managing bio-based supply chain information.

Data ingestion and pre-processing

The workflow begins with the retrieval of data from WP5's ICT tool via APIs. This step involves:

- Collecting raw data streams from IoT devices, questionnaires, and other input sources that are pre-validated by the ICT tool.
- Standardizing and normalizing data formats to maintain consistency across all datasets.
- Conducting preliminary checks to detect missing or inconsistent values, triggering automated alerts for data correction.

Computational core

The T&T system's computational core is designed to process large volumes of data efficiently:

- Modular computing framework: By adopting a modular structure, the system isolates specific computational tasks, such as CI calculations, conformity checks, and reliability assessments. This segmentation enhances accuracy and reduces computational overhead.
- Use of Apache Spark: Apache Spark provides a scalable and fault-tolerant environment for handling both real-time and historical datasets. Its Structured Streaming module dynamically processes incoming data, such as sustainability metrics, ensuring timely outputs for decision-making.

Indicator calculations

At the heart of the workflow lies the computation of various indicators, including circularity metrics, reliability, conformity, and overall performance:

- The formulas, extracted and interpreted from WP2's Excel file, are applied to calculate approximately 40 CIs. These formulas adapt to evolving requirements with minimal manual intervention.
- Composite indicators, such as the overall performance metric, integrate multiple dimensions of assessment, including environmental and operational factors.
-

Data Storage and Security

Processed data is partitioned for optimized storage and retrieval:

- Immutable records, such as compliance metrics, are stored on the blockchain, safeguarding their integrity.

- Frequently accessed datasets, such as real-time alerts and summaries, are cached locally to improve access speed and system responsiveness.

Output generation and reporting

The workflow concludes with the generation of reports and dashboards tailored for different stakeholders:

- Stakeholders receive personalized reports detailing their performance against predefined benchmarks.
- Interactive dashboards provide real-time visualization of key metrics, supporting immediate decision-making and long-term strategic planning.

This workflow, grounded in modularity and advanced computational techniques, represents a robust system for processing complex datasets. It serves as the backbone of the BioReCer T&T application, delivering insights that drive transparency, compliance, and sustainability within bio-based supply chains.

3.1.4 Security and compliance considerations

The BioReCer T&T system is built on a foundation of robust security measures and compliance mechanisms, safeguarding sensitive data and adhering to international standards. These considerations are critical for maintaining trust among stakeholders and ensuring the integrity and confidentiality of the bio-based supply chain's data.

Data security measures

1. **Encryption techniques**:
   - All data exchanges, including communication with WP5's ICT tool, are secured using TLS/SSL protocols. This prevents interception or tampering during transmission.
   - Advanced Encryption Standard (AES) is applied to locally stored data and blockchain records, protecting sensitive information from unauthorized access.
2. **Authentication and access control**:
   - A multi-layered authentication process, including OAuth protocols and multi-factor authentication (MFA), ensures that only authorized personnel can access the system.
   - RBAC restricts data access based on stakeholder roles. For instance:
     - Primary producers can view and update data relevant to raw material sourcing.

- Certification bodies can access compliance metrics without interfering with operational data.

3. **Anomaly Detection and Monitoring**:
   - Real-time monitoring algorithms track unusual patterns in data usage or access attempts, flagging anomalies for immediate review.
   - Automated alerts notify system administrators of potential security threats, enabling swift countermeasures.

Compliance framework

The T&T system is aligned with internationally recognized standards to meet both legal and ethical obligations:

1. Regulatory Adherence:
   - The system complies with GDPR for data privacy and protection, ensuring the secure handling of personal and sensitive information.
   - ISO 27001 guidelines are followed to establish a systematic approach to managing information security risks.
2. Dynamic Compliance Reporting:
   - Automated reports verify that data handling processes meet sustainability and ethical standards, such as those outlined in the EU Circular Economy Action Plan.
   - Compliance metrics, such as CIs, are dynamically calculated and made available for audits and certification schemes.
3. Collaboration with WP4:
   - Close collaboration with WP4 ensures alignment with ethical guidelines and addresses potential risks related to data misuse or inaccuracies.
   - WP4’s expertise is leveraged to continuously update the compliance framework based on evolving regulatory landscapes.

Blockchain Integration

The integration of blockchain technology further strengthens the security and transparency of the T&T system:

- Immutable records stored on the blockchain provide an auditable trail of transactions, enhancing accountability among stakeholders.
- Smart contracts automate compliance checks, reducing manual errors and streamlining regulatory reporting.

Further enhancements

As BioReCer evolves, additional measures will be explored to enhance security and compliance:

- Implementation of advanced threat detection systems using machine learning to identify sophisticated cyberattacks. The WP3 team possesses strong expertise in machine learning, which positions them uniquely to develop and deploy cutting-edge threat detection models. This forms a significant component of WP3's strategic plans for the future, ensuring the system can anticipate and mitigate complex cyber threats effectively.
- Expanding the compliance framework to incorporate new standards as they emerge, ensuring the system remains future-ready.

By embedding comprehensive security and compliance considerations into its design, the BioReCer T&T system upholds the highest standards of trust, transparency, and reliability, setting a benchmark for bio-based supply chain management.

3.2 Data handling and storage protocols

The T&T system requires a robust framework for managing, storing, and processing diverse datasets generated across various stages of the bio-based supply chain. Efficient data handling and storage are crucial not only for supporting real-time analytics but also for upholding transparency, security, and compliance. This subsection delves into the methodologies and protocols WP3 employs for data organization, storage infrastructure, and secure transmission, ensuring seamless integration with WP5's ICT tool while preparing for scalability and future growth.

3.2.1 Efficient data organization and management

The T&T system leverages advanced methodologies for data organization and management to efficiently process and analyse the diverse datasets inherent in bio-based supply chains. This approach prioritizes scalability, reliability, and precision, confirming that stakeholders can access actionable insights without encountering data inconsistencies or inefficiencies.

Dynamic data segmentation

The system's architecture adopts a hierarchical approach to data segmentation, structuring datasets into logical partitions based on predefined categories. These categories include:

- Supply chain stages: Data is categorized into distinct stages such as production, processing, and waste management. This segmentation enables granular tracking of supply chain activities and facilitates stage-specific analytics.

- Stakeholder roles: By isolating data for primary producers, intermediate processors, and waste managers, the system supports targeted reporting and minimizes redundancy.
- Geographic partitioning: Data is further segmented by region or jurisdiction, ensuring compliance with location-specific regulations and enabling regional performance comparisons. While geographic partitioning may seem less critical within the EU, it can still be relevant for the following reasons:
    1. Regional regulatory variations:
    While the EU has overarching regulations like GDPR and the Circular Economy Action Plan, individual member states may implement additional regional rules or interpretations. Geographic partitioning allows the system to accommodate these nuances without affecting datasets from other jurisdictions.
    2. Performance benchmarking:
    Segmenting data by region can support performance comparisons across member states, identifying best practices or areas needing improvement in specific regions.
    3. Future scalability:
    Geographic partitioning sets the groundwork for expanding the system beyond the EU or into specialized projects that may involve specific regions within the EU. For example, fisheries data might require partitioning by maritime zones.
    4. Stakeholder-Specific Needs:
    Some stakeholders may prefer regional data isolation to focus on local operations or regulatory submissions.

This dynamic segmentation allows the system to process queries and computations efficiently, reducing processing time and computational load. For example, a query to analyse recycling rates in a specific region can be executed within a localized data partition without affecting the entire dataset.

RBAC

As mentioned earlier, RBAC ensures that stakeholders can only access datasets pertinent to their responsibilities. This approach incorporates:

- Granular permissions: Access rights are assigned at a fine-grained level, enabling users to interact with specific subsets of data without exposing unrelated or sensitive information. D3.3 will discuss these levels in detail.

- Dynamic access allocation: Permissions are dynamically assigned or revoked based on stakeholder roles and changing project requirements, ensuring adaptability and data confidentiality.

For instance, certification bodies are granted access to compliance metrics and audit trails but are restricted from modifying data inputs provided by manufacturers or waste managers. This segregation of access rights mitigates potential risks of data tampering or misinterpretation.

Real-time data preprocessing

The T&T system integrates a robust preprocessing layer to handle data retrieved from WP5's ICT tool. Key functionalities of this layer include:

- Standardization: All incoming data is converted into a unified format compatible with the T&T system's analytical modules. This harmonization reduces inconsistencies arising from diverse data sources.
- Error detection: Anomalies, such as missing fields or out-of-range values, are flagged during preprocessing. Automated correction protocols or alerts to relevant stakeholders address these discrepancies before the data enters the computational pipeline.
- Prioritization of time-sensitive data: Critical datasets, such as real-time environmental sensor readings, are prioritized in the preprocessing queue to support dynamic analytics and decision-making.

For example, preprocessing modules identify and correct discrepancies in moisture content data reported by IoT sensors during material transit. By ensuring that corrected data enters the system promptly, the T&T application delivers reliable sustainability metrics for stakeholders.

Scientific contributions

The methodologies employed in the BioReCer T&T system represent a significant advancement in the field of bio-based supply chain management. The combination of dynamic segmentation, RBAC, and real-time preprocessing provides a scalable framework capable of handling complex datasets. Furthermore, these strategies establish a robust foundation for advanced analytics, such as predictive modelling and machine learning, which can be integrated in future iterations of the system.

3.2.2 Storage infrastructure

The storage infrastructure of the BioReCer T&T system is designed to accommodate the unique demands of bio-based supply chains, balancing the need for high-speed data access with robust long-term storage solutions. By integrating advanced storage technologies, the system achieves secure and efficient data retention, supporting dynamic analytics and compliance requirements.

Hybrid storage architecture

The T&T system employs a hybrid storage model that combines the advantages of centralized blockchain technology with localized storage solutions:

1- Centralized blockchain:
- Serves as the primary repository for critical data, such as T&T data, compliance metrics, CIs, and audit trails.
- Immutable data storage on the blockchain ensures transparency and traceability, enabling stakeholders and auditors to verify historical records.
- While centralized, this blockchain is hosted on a local server, tailored to the specific needs of BioReCer, providing a streamlined and secure approach to data management.

2- Local storage for real-time processing:
- Frequently accessed datasets, such as real-time sustainability metrics or alerts, are stored locally to reduce latency and enhance system responsiveness.
- This storage layer complements the blockchain, optimizing performance for time-sensitive computations while retaining the ability to archive critical information on the blockchain for long-term use.

Scalability and resource optimization

The storage infrastructure is designed to handle the growing complexity and volume of bio-based supply chain data:

1. Elastic storage capabilities:

   Cloud-ready configurations allow the system to dynamically scale storage capacity based on demand, ensuring uninterrupted operations during peak data loads or stakeholder expansion.
2. Partitioning and indexing:

Data is partitioned into logical segments based on categories such as supply chain stage, stakeholder role, or time period. This approach minimizes storage redundancy and improves query performance.

3. Energy-efficient solutions:

    The infrastructure integrates green data centre technologies, including liquid cooling systems, to minimize energy consumption. These innovations align with BioReCer's sustainability objectives by reducing the environmental footprint of data storage.

Data security in storage

1. Encryption:

    All stored data, whether on the blockchain or local servers, is encrypted using Advanced Encryption Standard (AES). This safeguards sensitive information against unauthorized access.

2. Access Controls:

    RBAC mechanisms are implemented to restrict data retrieval or modification to authorized users, promoting confidentiality and data protection.

3. Backup and Recovery:

    The system incorporates automated backup protocols, enabling quick recovery in case of data loss or corruption.

Compliance with standards

The storage infrastructure adheres to international standards for data management and security:

1. ISO 27001:

    Certifies a structured approach to risk management and information security, covering all aspects of data storage and handling.

2. GDPR Compliance:

    The system complies with GDPR regulations, providing mechanisms for data access requests, deletion, and anonymization as needed.

Future-ready design

The storage infrastructure is built to adapt to emerging challenges and technologies:

- Integration with advanced analytics:

    The architecture supports future integration with machine learning models and predictive analytics tools, allowing scalability for evolving project needs.

- Exploration of distributed storage:

  Distributed ledger technology (DLT) could be explored as a future enhancement to decentralize storage, improving fault tolerance and data redundancy. This approach would involve close collaboration with WP5 to align with their existing ICT tool architecture and ensure seamless integration with BioReCer's broader framework.

The storage infrastructure of the BioReCer T&T system represents a balanced and forward-looking approach to data retention, combining high-speed accessibility with secure and sustainable long-term storage solutions. This robust framework underpins the system's ability to meet the operational, analytical, and compliance needs of bio-based supply chains.

3.3 Scalability and optimization

The current system is engineered to handle the dynamic and growing demands of bio-based supply chains. By incorporating scalable architecture and advanced optimization techniques, the system maintains high performance, even as data volumes, user interactions, and stakeholder complexity increase. This section outlines the strategies employed to maintain operational efficiency and adaptability while preparing the system for future expansions.

Scalable system architecture

The T&T system adopts a modular and distributed design that supports seamless scaling:

1. Horizontal scalability:

   The architecture allows for the addition of computational and storage resources without disrupting ongoing operations. This approach is particularly beneficial for handling increasing volumes of data from IoT devices, stakeholder inputs, and CI computations.

2. Cloud-ready configurations:

   While primarily centralized, the system integrates cloud-ready components to dynamically allocate resources during peak loads or project expansions. Cloud-based scaling ensures that the system can accommodate additional stakeholders or regions without overburdening local infrastructure.

Optimization techniques

To maintain performance and efficiency, the system employs a combination of advanced techniques:

1. Partitioning and indexing:

   Data is partitioned into logical segments based on supply chain stages, stakeholder roles, or geographic regions, allowing targeted access and reducing computational overhead. Indexed queries accelerate retrieval times for commonly used datasets, such as recycling rates or conformity metrics.

2. Caching mechanisms:

   Frequently accessed datasets, including real-time analytics and compliance summaries, are cached locally. This reduces latency and enhances the responsiveness of the system for stakeholders.

3. Streamlined data processing:

   Apache Spark's distributed processing capabilities are leveraged to execute real-time analytics and batch processing efficiently. The structured streaming module confirms low-latency computations while handling large-scale data streams.

Future scalability considerations

The T&T system is crafted to support long-term growth and adaptability, addressing the increasing demands of bio-based supply chains while maintaining operational efficiency.

- Dynamic system evolution:

  The infrastructure allows for seamless scalability, accommodating the growing complexity of supply chains and the integration of new stakeholders or data streams. Future extensions are supported through modular design, which minimizes disruptions during upgrades or expansions.

- Optimized resource utilization:

  The system employs advanced strategies for load balancing and resource allocation, ensuring stable performance under varying operational conditions. By utilising partitioning and intelligent data workflows, it efficiently handles the influx of data without overburdening storage or computational resources.

This scalability-focused approach strengthens the framework, enabling it to respond to the demands of expanding supply chains and emerging technological advancements, without repeating existing methodologies from earlier sections.

3.4 Cloud-based infrastructure overview

The BioReCer T&T system's cloud-based infrastructure represents a pivotal element in managing bio-based supply chains effectively. As supply chains grow more intricate and the demand for real-time traceability increases, using cloud technologies provides the system with the scalability, adaptability, and robustness necessary for dynamic data management. This section goes into the features, advantages, and potential growth of the cloud-based architecture, emphasizing its alignment with regulatory frameworks and operational objectives.

Architectural principles of the cloud-based system

The design of the T&T cloud-based infrastructure is driven by three primary principles: scalability, integration, and security. These principles underpin its ability to handle vast datasets while maintaining high performance and accessibility across diverse stakeholders.

1. **Scalability and elastic resource allocation**
   - The cloud platform dynamically adjusts resources to meet varying workloads, optimizing performance during high-demand scenarios, such as regulatory audits or mass data ingestion events.
   - Elastic resource provisioning reduces operational overhead while maintaining cost-efficiency, particularly valuable for periodic high-computation tasks, such as CI analysis or sustainability performance benchmarking.

2. **Centralized yet flexible data management**
   - The cloud serves as a unified repository for all data, facilitating the seamless integration of inputs from IoT devices, manual questionnaires, and WP5's ICT tool.
   - Advanced indexing and data partitioning enable efficient querying and processing without overloading computational resources. For example, specific datasets required for evaluating energy usage or recycling rates can be accessed without disrupting other operations.

3. **Robust API Framework for Integration**
   - A sophisticated API-driven architecture bridges the T&T system with WP5's ICT tool, developing interoperability.

- APIs facilitate real-time data exchange, ensuring that processed circularity metrics, compliance indicators, and reliability reports are readily available for stakeholders.
- The modular API framework also allows for easy incorporation of future technologies or datasets, supporting the evolving needs of the BioReCer ecosystem.

Advantages of a cloud-based infrastructure

The adoption of a cloud-based system offers several distinct benefits, positioning the T&T application as a forward-thinking solution within bio-based industries.

1. Dynamic collaboration and accessibility:
   Stakeholders, regardless of geographical location, gain access to the system's data and analytics dashboards, promoting transparent and informed decision-making. Collaborative tools integrated into the cloud platform facilitate multi-stakeholder engagement, enabling certification bodies, producers, and policy makers to work cohesively on shared objectives.
2. Operational reliability and high availability:
   Cloud-based redundancies minimize the risk of data loss or downtime, ensuring continuous functionality even in the face of unexpected events such as server failures. High availability is critical for ongoing monitoring of supply chains, particularly in scenarios requiring real-time interventions, such as detecting anomalies in sustainability metrics.
3. Enhanced security and compliance measures:
   Security protocols embedded within the cloud infrastructure, such as AES encryption and advanced access controls, safeguard sensitive data from unauthorized access or cyberattacks. The system adheres to EU-specific regulatory requirements, aligning with standards such as GDPR, thereby promoting ethical and legal accountability across supply chain operations.
4. Cost efficiency and future proofing:
   By using a pay-as-you-go model for computational and storage needs, the cloud infrastructure optimizes resource utilization while minimizing unnecessary expenditures. The architecture is built to integrate seamlessly with emerging technologies, such as machine learning models and predictive analytics, without the need for major system overhauls.

Key applications within BioReCer

1. Real-time analytics and CIs:
   The cloud infrastructure facilitates the computation of over 40 CIs, integrating real-time data streams from IoT sensors with historical datasets stored on the blockchain. These metrics are visualized through intuitive dashboards, enabling stakeholders to monitor compliance and performance dynamically.
2. Stakeholder-specific data access:
   RBAC mechanisms allow stakeholders to retrieve only the data pertinent to their operations. For example, manufacturers can access material recycling rates, while certifiers focus on compliance summaries.
3. Regulatory Audits and Reporting:
   The cloud's centralized repository simplifies the generation of audit-ready reports, which can be dynamically updated to reflect the latest data inputs. These reports serve as evidence of adherence to EU sustainability standards, such as ISO 14001.

Integration with ICT tool

The BioReCer T&T cloud infrastructure is intricately linked with WP5's ICT tool, which serves as the primary gateway for data exchange. This integration facilitates:

- Streamlined data ingestion.
  Pre-validated data from the ICT tool is ingested into the cloud platform, eliminating redundancies and reducing preprocessing overhead.
- Feedback loops for improved data quality
  Results generated by the T&T system, such as reliability scores or conformity reports, are fed back into the ICT tool. This creates a feedback loop that enhances data accuracy over time, benefiting all stakeholders in the supply chain.

Future directions for cloud-based operations

1. Integration of machine learning and predictive analytics:
   - Advanced analytics tools, powered by machine learning, will be incorporated into the cloud framework to predict supply chain risks, such as material shortages or compliance violations.
   - Predictive insights derived from cloud-stored data will empower stakeholders to take pre-emptive actions, enhancing overall supply chain resilience.
2. Exploration of multi-cloud architectures:

- To mitigate risks associated with vendor lock-in, the BioReCer T&T system may adopt a multi-cloud strategy, utilising the strengths of various cloud providers.
- Multi-cloud architectures will also improve fault tolerance and optimize resource allocation for specific workloads, such as large-scale data validation tasks.

The cloud-based infrastructure of the T&T system is a cornerstone of its capability to manage complex bio-based supply chains. Its dynamic, secure, and scalable design addresses current project needs while providing a robust platform for future expansion. By integrating with WP5's ICT tool and supporting advanced analytical capabilities, this infrastructure lays the groundwork for achieving the BioReCer project's ambitious goals in sustainability, traceability, and compliance.

4. Hyperspectral imaging (HSI) integration

HSI represents a significant technological advancement for the BioReCer project, specifically in enhancing traceability and valorization of biological byproducts. The following subsections provide detailed insights into HSI's underlying principles, current applications, and its tailored integration within the BioReCer framework.

4.1 Introduction to HSI technology

HSI is an advanced technology that integrates imaging and spectroscopy to provide detailed information about the spectral properties of materials in each scene. Unlike traditional imaging systems, which typically capture only three bands of colour (red, green, and blue), hyperspectral systems acquire data across hundreds of contiguous spectral bands. This capability allows for the detailed characterization of objects and materials based on their unique spectral fingerprints, enabling precise identification and analysis that would otherwise be unattainable with conventional imaging methods.

The foundation of hyperspectral technology lies in the principle that every material interacts with light differently across various wavelengths. This interaction generates a spectral signature—a unique pattern of reflectance, absorption, or emission that acts as a fingerprint for the material. Hyperspectral sensors capture these signatures by recording the reflected or emitted light in narrow, contiguous spectral bands, spanning a broad range of wavelengths, from the visible (400–700 nm) to the near-infrared (700–1000 nm), and even extending into the short-wave infrared (SWIR) or thermal infrared regions, depending on the system's design [18].

The resulting data is typically represented as a three-dimensional dataset known as a "hypercube." In this hypercube, two dimensions correspond to the spatial coordinates of the scene, while the third dimension contains the spectral information for each pixel. This combination of spatial and spectral data enables a deeper understanding of the physical and chemical properties of the scene, making HSI a powerful tool for a wide range of applications.

Applications of HSI

HSI has found extensive use in diverse fields due to its unparalleled ability to discriminate materials and detect subtle differences in composition. Some key application areas include:

- Agriculture: Monitoring crop health, detecting pests, assessing soil conditions, and optimizing water and fertilizer usage.

- Environmental Monitoring: Identifying pollutants, mapping vegetation, and assessing water quality.
- Defence and Security: Detecting camouflaged objects, monitoring borders, and analysing hazardous materials.
- Medical Diagnostics: Identifying tissue abnormalities, such as cancerous lesions, based on their spectral characteristics.
- Food Quality Control: Inspecting food products for contaminants, spoilage, or adulteration.
- Mineral Exploration: Mapping mineral compositions and identifying valuable resources.

The adoption of hyperspectral technology continues to grow as sensor technology improves, costs decrease, and data processing capabilities advance. Modern systems are increasingly compact and efficient, enabling deployment in satellites, aircraft, drones, and even handheld devices, making the technology more accessible across industries [19].

Challenges and future directions

Despite its advantages, HSI faces challenges that limit its broader adoption. The sheer volume of data generated by hyperspectral sensors requires significant storage, computational resources, and advanced algorithms for processing and analysis. Additionally, the complexity of interpreting hyperspectral data often demands specialized expertise, making it less accessible to non-experts.

Future developments in hyperspectral technology are likely to focus on improving sensor resolution, reducing costs, and enhancing data processing techniques through machine learning and artificial intelligence. These advancements will further expand the potential applications of HSI, unlocking new possibilities in both scientific research and commercial industries.

4.2 Role of HSI in the BioReCer framework

In the context of biological byproduct valorization, HSI could play a crucial role in assessing the potential of raw materials for generating high-value biological byproducts. In our case of study, algae byproducts hold significant promise due to their diverse biochemical composition, including sugars, proteins, fats, and moisture, which can be exploited for applications in industries such as food, cosmetics, biofuels, and pharmaceuticals. HSI facilitates the non-destructive, rapid, and precise evaluation of these parameters, making it a potentially valuable tool in this domain.

Material characterization and quality assessment

HSI enables the detailed characterization of algae by analysing their spectral signatures, which correspond to their physical and chemical properties. This capability allows researchers to:

- Quantify Biochemical Components: HSI can estimate the content of critical components such as sugars, proteins, fats, and moisture by correlating spectral data with reference laboratory measurements. This information helps determine the suitability of algae for specific applications.
- Assess Homogeneity and Quality: By mapping the spatial distribution of biochemical parameters within the sample, HSI identifies variations in composition, enabling the selection of high-quality biomass for further processing.
- Detecting Contaminants: HSI can identify impurities or undesired materials in algae samples, ensuring the purity of raw materials used in subsequent production processes.

Process optimization and industrial applications

The integration of HSI into the valorization process enhances efficiency and product quality by:

- Real-Time Monitoring, since Hyperspectral systems can be implemented in-line during algae processing to monitor changes in biochemical composition, optimizing extraction methods and ensuring consistency.
- Reducing Waste by accurately evaluating the potential of raw algae. HSI technology could minimize the use of unsuitable materials, thereby reducing waste and improving sustainability.
- Enabling Targeted Applications with precise knowledge of biochemical content. In this way, algae byproducts can be tailored to meet the specific needs of industries such as nutraceuticals, cosmetics, or bioenergy.

Future perspectives

As HSI technology continues to advance, its application in biological byproduct valorization will become increasingly sophisticated. Emerging trends include the use of machine learning algorithms to enhance spectral data interpretation and the development of portable HSI devices for field-based assessments. These innovations will further streamline the evaluation and utilization of algae and other biological materials, unlocking new opportunities for sustainable resource management and industrial innovation.

4.3 Data collection and processing with HSI

As outlined in task T3.2 (Design, build and implementation of the T&T software and hardware apparatus) of the proposal, an HSI detection system has been developed to perform an initial assessment of the potential of algal biomass for exploitation. Obtaining reliable models for the characterization of some of the main physico-chemical parameters of the primary biomass resources could be highly valuable in predicting the yield and quality of by-products derived from its processing.

To this end, samples of some of the main algae species harvested for processing along the Galician coast have been collected. The spectral characteristics of each sample were extracted using a specially designed setup, and a training set was configured with the analytical values obtained in the laboratory to develop a predictive model for each selected parameter.

Setup design for the application

ANFACO-CECOPESCA has proven experience in the study and characterization of food matrices of diverse nature with HSI technology. In their facilities they count with two SPECIM pushbroom sensors covering the visible and near infrared spectrum range: the VIS-NIR FX10 model (from 400nm to 900nm) and the NIR FX17 model (from 900nm to 1700nm). These devices are designed for in-line mounting and require a setup to coordinate the movement of the samples on the horizontal axis to compose the hypercubes, which are the files containing the spectral information of each pixel for each spectral band.

Although usually these sensors are used in reflectance or transreflectance configurations, where what is measured is the emitted light that reaches the sample and returns to the sensor, for this application it has been considered that a transmittance configuration, where the light is emitted from below and, after passing completely through the sample, reaches the sensor (Figure 9). This configuration could present a number of advantages:

- As the light passes completely through the sample, this type of set-up provides direct information about the internal composition. This is critical for the determination of parameters such as moisture, fat, protein or salt, which affect the spectral absorption of light in specific bands.
- The interaction of light is not so much dependent on the surface, but on the internal absorption of the chemical compounds. This reduces scattering caused by rough or uneven surfaces, a common problem in dry algae. This results in a better signal-to-noise ratio and mitigates the need to apply advanced corrections such as MSC

or SNV to remove the effects of light scattering and surface variations, simplifying data processing.

For this reason, an industrial machinery company has been subcontracted, that designed and manufactured a customized setup that allows to present the dry algae samples in movement under the capture line of the cameras while they are illuminated from below.

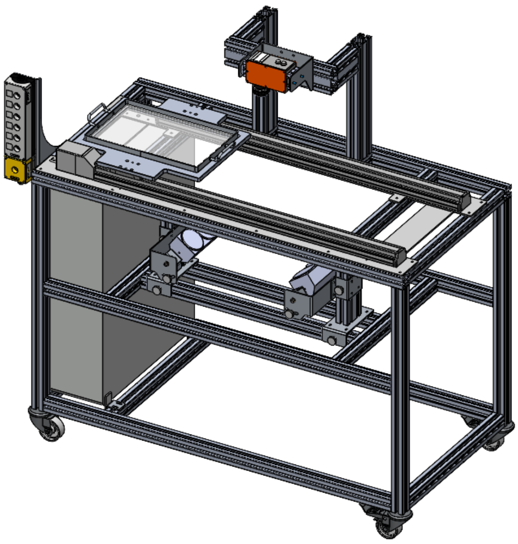

Figure 9: CAD design for transmittance setup

Sampling and data acquisition

The algae species of greatest interest for this study have been identified as those most abundant along the coasts of Galicia, the region where this case study is conducted, and which are harvested for industrial purposes. The selected species are as follows:

- *Himanthalia elongata*
- *Ulva lactuta*
- *Laminaria ochroleuca*
- *Undaria pinnatifida*
- *Codium tomentosum*
- *Porphyra umbilicalis*

Initially, field trips were organized to collect fresh samples from the intertidal zones of various beaches along the Galician coast (Figure 10). However, this approach presented significant challenges. Due to the high water content of the biomass (approximately 90%), obtaining around 100 grams of material required for subsequent analysis necessitated collecting nearly one kilogram of fresh algae. Furthermore, the collected samples required additional steps, including washing, drying, and processing, making this method both labour-intensive and time-consuming (Figure 11). Moreover, the variability required in the samples for the different species could not always be guaranteed.

Figure 10: Fresh algae harvesting on the beach.

To overcome these challenges, the decision was made to collaborate with local suppliers offering crushed, dried seaweed as part of their product catalogue.
Samples from different batches and suppliers were labelled and analysed using a HSI system to generate spectral hypercubes for each sample in both the VIS-NIR and NIR ranges. Figure 12 shows the process of preparing algae samples for laboratory analysis, an essential step to validate and calibrate hyperspectral imaging results for biochemical characterization.

Figure 11: Algae spectral data acquisition with transmittance setup

As dependent variables, key nutritional parameters of significant interest were selected, based on existing literature and the expertise of ANFACO, which has demonstrated success in developing predictive models. These parameters include:

- Moisture
- Ashes
- Proteins
- Fat

- Carbohydrates
- Sodium
- Salt

Once the spectral hypercubes for each sample were obtained, the samples were sent to the laboratory for chemical analysis to determine their nutritional parameters. These analytical values, combined with the mean spectral data from each hypercube, were used to create training sets for developing the predictive models (

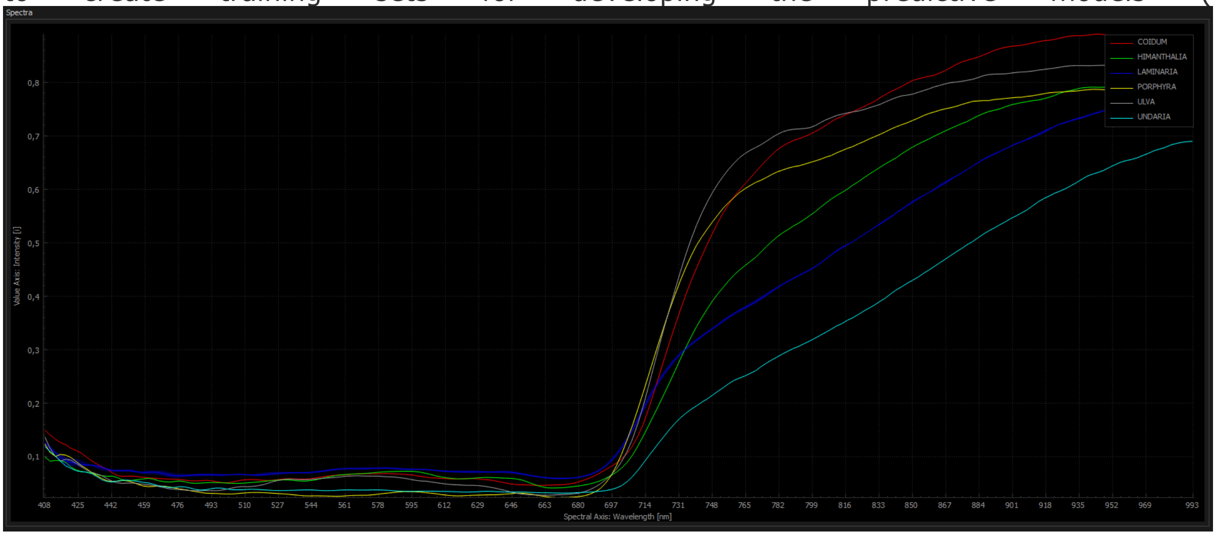

Figure 13).

Figure 12: Sample preparation for laboratory analysis

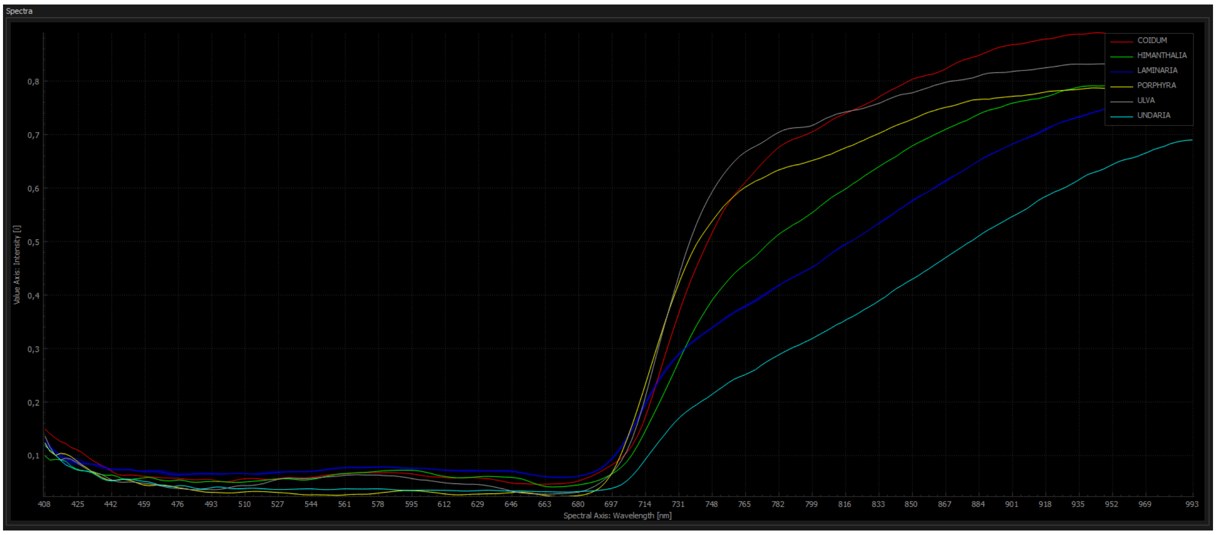

Figure 13: Mean spectrum of the samples for each algae species

4.4 Benefits of HSI for biological feedstock traceability

one of the main advantages of HSI technology is its ability to provide detailed, non-destructive, real-time information on the physical and chemical composition of biological materials, ensuring that the specific requirements for obtaining high quality by-products are met at source.

As the data model has been proposed in the BioReCer framework under NGSI-LD standard, all the information related to the results obtained by the predictive models of the specific characteristics of each biomass could be associated as attributes to any product at any point in the chain, so that the users of the tool would have this additional information to evaluate the potential of each product according the standards set by the industry or their own standard.

This is also an important point for producers, as it would help them meet the requirements for certain certifications schemes. For all these reasons, it can be stated that this extra layer of information provided by this type of online quality control system at source would ultimately help all stakeholders by providing transparency on biomass characteristics that would aid the decision-making process from the point of view of efficiency and sustainability.

5. Implementation and performance

The implementation and performance of the T&T system demonstrate its capacity to handle complex bio-based supply chain requirements effectively. This chapter details the methodologies employed during system deployment, the key performance metrics evaluated, and the results of initial validation phases. By showcasing both the operational strengths and areas for future optimization, this section highlights the practical impact of the T&T framework in achieving BioReCer's sustainability and traceability goals.

5.1 Deployment methodology

The deployment of the T&T system follows a structured and iterative methodology, emphasizing real-world applicability, stakeholder engagement, and seamless integration with existing tools. This approach justifies that the system effectively addresses the complexities of bio-based supply chains while remaining adaptable to evolving requirements.

Phased implementation strategy

The deployment process is divided into distinct phases to manage complexity and reduce risks:

1. **Preparation and integration**:
   - Establishing connections with WP5's ICT tool using secure APIs to retrieve pre-validated data.
   - Configuring data pipelines for ingesting real-time inputs from IoT devices, questionnaires, and HSI systems.
   - Initial validation of the centralized blockchain infrastructure to ensure accurate and immutable data storage.

2. **Pilot testing**:
   - Conducting pilot deployments across select supply chain scenarios, such as fisheries or agricultural feedstock processing.
   - Testing critical features, including CI computations, real-time analytics, and compliance reporting.
   - Gathering feedback from stakeholders to refine workflows and address usability challenges.

3. **Full-scale rollout**:
   - Expanding system deployment to all participating stakeholders, incorporating the lessons learned from pilot testing.
   - Configuring stakeholder-specific access controls, dashboards, and reporting mechanisms.
   - Establishing real-time monitoring capabilities for continuous performance evaluation.

Collaborative development with stakeholders

The deployment process actively involves stakeholders from various supply chain stages:

1. Producers: Validate raw material data accuracy and initial circularity metrics.
2. Processors and manufacturers: Test intermediate computations and conformity assessments.
3. Certifiers and policy makers: Confirm seamless access to compliance data and audit trails.

This collaborative approach enhances the system's relevance and promotes stakeholder buy-in and long-term adoption.

Technical tools and resources

1. **Deployment framework**:
   - Apache Spark serves as the backbone for real-time analytics and data processing, ensuring scalability and fault tolerance.
   - Dockerized environments streamline system configuration and deployment, allowing rapid replication of the T&T framework across different sites [20]. This topic is discussed in the next subsection.

2. **Performance testing tools**:
   - Load testing tools simulate high data volumes to evaluate system reliability and responsiveness under stress.
   - Automated error detection tools identify bottlenecks or potential failures during deployment.

Challenges and mitigation strategies

The deployment process can encounter several challenges, which are addressed proactively:

- **Data integration complexities**:

  To address data integration complexities, we employ preprocessing and validation pipelines to standardize inputs and resolve inconsistencies. These pipelines act as an intermediary layer, ensuring that all incoming data conforms to predefined formats and quality standards before entering the analytical and storage systems. By detecting and addressing issues such as missing values, duplicate entries, or format mismatches, these pipelines enhance data accuracy and reliability. This standardization not only facilitates seamless integration with the T&T system but also improves the overall efficiency of downstream processes.

- **Stakeholder engagement**:

  To enhance stakeholder engagement, regular feedback loops and targeted training sessions will be implemented (Here, we need to contribute with our other partners). Feedback loops provide stakeholders with opportunities to share their experiences and challenges while interacting with the system, allowing for iterative improvements tailored to their needs. Training sessions can assist to familiarize users with the system's features, therefore, they can navigate and utilize its functionalities effectively. This combination of continuous feedback and education fosters user confidence, promotes acceptance, and encourages active participation across all levels of the bio-based supply chain.

- **Scalability constraints**:

  The adoption of a modular architecture and cloud-ready configurations facilitates seamless system scaling to accommodate growing demands. The modular design allows individual components of the T&T system, such as data processing or storage modules, to be expanded or upgraded independently without affecting the overall framework. Meanwhile, cloud-ready configurations enable the system to dynamically allocate additional computational and storage resources during peak workloads. This dual approach ensures that the T&T system remains adaptable to increasing data volumes, stakeholder interactions, and evolving project requirements, all while maintaining optimal performance.

Dockerized Environments: Streamlining configuration and deployment

Dockerized environments refer to the use of Docker, a containerization platform, to package software applications along with their dependencies into lightweight, portable containers. These containers provide a consistent environment for running applications, regardless of the underlying infrastructure. Dockerization simplifies the deployment process by encapsulating the T&T application and its dependencies into reusable and easily deployable containers [21].

Important features of Dockerized environments

1. **Portability across systems**:
   Docker containers operate the same way across different operating systems and hardware configurations. This consistency eliminates compatibility issues, making it easier to replicate the T&T system across multiple deployment sites or stakeholder environments. For example, a Dockerized T&T framework can be deployed on local servers, cloud platforms, or even individual stakeholder environments without requiring extensive reconfiguration.
2. **Simplified deployment and rollback**:
   Docker allows the entire T&T application, including its software dependencies, libraries, and runtime environment, to be packaged into a single container image. This simplifies the setup process, reducing deployment times and minimizing the risk of configuration errors. Additionally, if an update or deployment encounters issues, Docker enables seamless rollback to a previous version of the container, ensuring minimal disruption to operations.

Benefits for the BioReCer T&T System

1. **Centralized deployment and rapid replication**
   Docker allows the T&T system to be quickly deployed within centralized environments, such as the server infrastructure supporting WP5's ICT tool. By packaging the system and its dependencies into a single container, the deployment process becomes highly streamlined. Updates or additional computational nodes can be implemented by simply running the Docker container, reducing setup complexity and ensuring consistent performance across deployments.
2. **Version control and iterative development**
   Dockerized environments support versioned containers, enabling the T&T development team to test, refine, and deploy new versions of the system without disrupting live operations. This capability is crucial for ongoing improvements, such as incorporating updated circularity formulas or refining performance metrics. Containers can also be rolled back to previous versions in the event of an issue, maintaining system stability.
3. **Resource efficiency and lightweight operations**

Docker containers share the host operating system's kernel, making them more efficient than traditional virtual machines. This lightweight design minimizes resource usage, allowing the T&T system to perform optimally on limited hardware resources. Although primarily hosted in centralized environments, this efficiency ensures that the system can scale flexibly as data volumes or computational demands increase.

4. **Seamless collaboration**
   Stakeholders, including WP5 teams and external partners, can deploy the same application version effortlessly, maintaining system consistency across the network.
5. **Dynamic scaling**
   Containers can be scaled horizontally to accommodate additional workloads, such as increased data streams from IoT devices or expanded stakeholder participation.
6. **Enhanced security**
   Isolated environments prevent containers from interfering with each other or the host system, reducing the risk of security breaches. Each container operates independently, with its own set of resources and dependencies, ensuring that activities within one container cannot interfere with others or compromise the host system. This isolation significantly reduces the attack surface, as even if one container is compromised, the breach is contained and cannot spread to other containers or the underlying infrastructure.
   Additionally, containers enforce strict boundaries using kernel-level features such as namespaces and control groups, which segregate process access and resource usage. These mechanisms prevent unauthorized access to critical system components and sensitive data stored on the host. Combined with RBAC and secure configuration practices, container isolation offers a multi-layered security approach that is well-suited for handling sensitive supply chain data in the system.

By employing Dockerized environments, the project achieves a streamlined, scalable, and resilient deployment process, ensuring that the framework can adapt to diverse operational requirements across bio-based supply chains.

5.2 Performance metrics and benchmarking

Evaluating the performance of a system is essential to demonstrate its operational efficiency, reliability, and ability to handle complex bio-based supply chain requirements. Key performance metrics were carefully defined, tested, and validated during implementation to ensure the system meets its objectives of scalability, transparency, and real-time functionality.

Key performance indicators

Data processing speed:

The T&T system is designed to handle incoming data streams with high efficiency and in real time. Data processing speed is a critical performance metric, as it determines how quickly the system can compute key outputs. This metric reflects the system's ability to deliver actionable insights promptly, enabling stakeholders to make informed decisions without delays.

The average processing time (T_{avg}) is calculated using the following formula:

$$T_{avg} = \frac{\sum_{i=1}^{n} T_i}{n} \quad \text{(Eq. 3)}$$

Where:

T_i: The time taken to process a single data input (e.g., a dataset from an IoT sensor or a stakeholder's submission).

n: The total number of data inputs processed during the evaluation period.

To measure system efficiency under varying workloads, a weighted variant of the formula can be applied:

$$T_{avg}^{weighted} = \frac{\sum_{i=1}^{n} (T_i \times W_i)}{\sum_{i=1}^{n} W_i} \quad \text{(Eq. 4)}$$

Where:

W_i: A weight assigned to each input based on its complexity or size (e.g., larger datasets or more complex computations have higher weights).

If larger datasets are weighted higher, the weighted average formula would be applied to account for the processing load associated with more complex inputs.

By regularly monitoring T_{avg}, the system's performance can be benchmarked against predefined thresholds, ensuring its capability to process real-time data without bottlenecks. Anomalies, such as significant increases in T_{avg}, could indicate issues in data flow or computational inefficiencies, prompting immediate intervention.

This quantitative approach establishes a clear and measurable foundation for evaluating and optimizing the T&T system's processing efficiency, facilitating its continued alignment with BioReCer's operational goals.

Query response time:

Query response time is a performance metric for evaluating the efficiency of the T&T system. It measures how quickly the system retrieves and processes data requests, indicating that stakeholders and the ICT tool receive timely access to key datasets. These datasets include sustainability metrics, compliance summaries, and CIs essential for decision-making and reporting.

A lower query response time indicates a more efficient system, enabling real-time interactions and seamless workflows across bio-based supply chains. This metric is particularly important during high-demand scenarios, such as regulatory audits or stakeholder consultations, where rapid access to accurate data is essential.

The average query response time (Q_{avg}) is calculated based on the following formula:

$$Q_{avg} = \frac{\sum_{i=1}^{n} Q_i}{n} \qquad \text{(Eq. 5)}$$

Where:

T_i: The time taken to process a single query (measured from the moment the query is received to when the response is delivered).

n: Total number of queries processed during the evaluation period.

If weighted by the importance or complexity of queries, the formula can be extended:

$$Q_{avg}^{weighted} = \frac{\sum_{i=1}^{n} (Q_i \times W_i)}{\sum_{i=1}^{n} W_i} \qquad \text{(Eq. 6)}$$

Where:

W_i: Weight assigned to each query based on its complexity, such as retrieving historical data vs. real-time metrics.

Optimization insights

By monitoring Q_{avg}, system administrators can identify bottlenecks, such as high latency during peak loads or inefficient query handling.

Optimizations may include:

1. Caching frequently accessed data to reduce query processing time.
2. Partitioning and indexing datasets to enable targeted retrieval.
3. Scaling computational resources during high-demand periods.

These insights guarantee that query response time remains within acceptable thresholds, supporting the system's goal of delivering actionable insights in real time.

System uptime and reliability:

System uptime and reliability are fundamental metrics for assessing the system's performance and operational dependability. These metrics reflect the system's ability to remain accessible and functional, minimizing disruptions that could impact stakeholders or supply chain workflows. High system uptime ensures that stakeholders can consistently access real-time data, compliance metrics, and CIs, which are important for decision-making and reporting.

Uptime is typically expressed as the percentage of time the system remains operational over a specific period, such as a month or year. Reliability, on the other hand, extends beyond uptime by also considering the system's ability to recover quickly from failures or handle unexpected disruptions effectively.

To measure system uptime and reliability:

1. **Monitoring operational hours**:
   The system's status is continuously monitored using automated tools that log whether it is functional or experiencing downtime. Logs include timestamps for when the system becomes unavailable and when it resumes operation.
2. **Calculating total availability**:
   The total operational time is calculated as the difference between the total time in a given period (e.g., a month) and the cumulative downtime during that period. For instance:
   If a month has 43,200 minutes and the system experienced 90 minutes of downtime, the uptime percentage would be:

$$Uptime\% = \frac{43{,}200-90}{43{,}200} \times 100 = 99.79\% \qquad \text{(Eq. 7)}$$

3. **Assessing recovery time**:
   Reliability also accounts for how quickly the system can recover from failures. Mean time to recovery (MTTR) measures the average time taken to restore functionality after an issue is detected. Short MTTR values indicate a resilient system.
4. **Tracking failure rates**: **Measuring MTBF**
   The Mean Time Between Failures (MTBF) is a reliability metric that quantifies the average operational time of a system between consecutive failures. It provides

insights into the stability and robustness of the system, with higher MTBF values indicating fewer interruptions and better reliability.

To calculate and track MTBF, we begin by clearly defining what constitutes a failure in the system, such as downtime, API failures, or blockchain unavailability, and measure the operational time between these events. Then we need to collect data by logging the start and end times of operational periods alongside timestamps for each failure.

$$MTBF = \frac{Total\ Operational\ Time}{Number\ of\ Failures} \qquad \text{(Eq. 8)}$$

where total operational time is the cumulative duration of system functionality, and the number of failures represents all recorded disruptions during the evaluation period. The data collection was automated. The calculation uses a monitoring tool, which can track system status, flag downtime events, and periodically compute MTBF. This streamlined process provides insights into system reliability, supporting ongoing efforts to minimize failures and improve overall stability.

5. **Automated alerts and redundancy**:

Automated monitoring tools are configured to alert administrators when downtime occurs, enabling immediate corrective actions. Additionally, redundant infrastructure components (e.g., backup servers, replicated data storage) are employed to reduce the impact of hardware or software failures.

Insights and optimization strategies

- **Real-time monitoring**:

  Advanced monitoring tool, Prometheus, integrated into the T&T system, provides real-time insights into uptime and performance trends. This tool helps detect potential issues proactively, such as increased latency or server load, before they lead to failures.

  Prometheus is an open-source monitoring system designed for high-dimensional data collection and analysis. It collects metrics from the T&T system's components (e.g., API response times, database performance) and generates alerts when anomalies are detected [22]. For instance, it monitors the load on Apache Spark during CI computations and confirms the system doesn't exceed resource limits.

- **Redundancy planning**:

Redundant servers and failover mechanisms ensure uninterrupted availability by automatically redirecting workloads during unexpected downtime.

- **Scheduled maintenance**:
  Regular maintenance windows are planned to minimize unplanned downtime while allowing system updates or upgrades.
- **Stakeholder assurance**:
  High uptime percentages demonstrate the system's reliability, building trust with stakeholders who depend on it for continuous access to critical data.

By maintaining robust uptime and reliability metrics, the T&T system establishes itself as a dependable platform for managing bio-based supply chains. These metrics ensure that stakeholders can rely on the system for uninterrupted access to insights and operations, even under demanding conditions.

Scalability and load handling:

Scalability and load handling are critical metrics to evaluate the system's capacity to manage increasing demands without compromising performance. These metrics assess how well the system maintains consistent processing speeds and responsiveness as the volume of data or the number of user interactions grows. In a complex bio-based supply chain, where data from IoT sensors, questionnaires, and compliance tools flow into the system simultaneously, the ability to handle such loads effectively is paramount.

To measure scalability, the system is tested under various simulated workloads, including increasing API calls, data streams, and concurrent user requests. Key indicators include the number of simultaneous API calls or data streams processed without noticeable degradation in processing speed or query response times. For instance, during stress testing, the system is subjected to high volumes of CI computations and stakeholder requests to evaluate its operational limits.

Advanced tools like Apache Spark's distributed processing capabilities and Dockerized environments play a crucial role in maintaining scalability. These technologies allow the system to dynamically allocate resources, such as computing power and memory, to accommodate surges in demand. Additionally, caching mechanisms and efficient indexing strategies ensure that frequently accessed datasets are retrieved rapidly, minimizing latency during peak loads.

By tracking scalability and load handling metrics, the T&T system demonstrates its robustness and preparedness to support expanding datasets, increased user engagement, and future integration of additional stakeholders and tools. Continuous monitoring and optimization further strengthen the system's ability to adapt to evolving project requirements.

Data accuracy and consistency

These metrics validate the reliability of the system's computations, which form the basis for decision-making and regulatory compliance in bio-based supply chains.

Data accuracy refers to the system's ability to produce outputs that match expected results, as defined by WP2's circularity formulas or other project-specific benchmarks. For example, the recycling rate of a specific product batch must be computed precisely, based on standardized inputs and calculation methodologies. Consistency, on the other hand, evaluates whether data outputs remain stable and uniform across repeated computations or under varying operational conditions, ensuring stakeholders receive dependable insights.

The primary metric for assessing data accuracy and consistency is the percentage of validated outputs that pass automated checks during computation. These checks include:

- Validation against predefined thresholds: For instance, verifying that computed indicators do not exceed acceptable limits set by sustainability regulations.
- Cross-referencing with historical data: Comparing current circularity metrics with previous results to identify discrepancies or anomalies.

The system employs preprocessing pipelines and automated validation mechanisms to detect and resolve errors before outputs are generated. These mechanisms identify issues such as incomplete datasets, outliers, or discrepancies in input values, flagging them for correction. Additionally, the centralized blockchain infrastructure provides a transparent and immutable record of computations, enabling stakeholders to trace and verify data integrity.

By maintaining high data accuracy and consistency, the T&T system builds stakeholder trust and ensures alignment with project goals, regulatory requirements, and operational standards. These metrics also serve as a foundation for continuous improvement, as anomalies or validation failures highlight areas for refinement in data handling and computational workflows.

5.3 Results from initial validation and testing

Robust testing and validation processes assess the system's performance, reliability, and scalability. These processes demonstrate that the system can handle real-world operational demands while delivering accurate and actionable outputs. A combination of stress testing, load testing, and targeted validation was employed to simulate diverse scenarios and identify areas for improvement. After each test, the key outcomes are reported.

5.3.1 Stress testing:

Stress testing evaluates the system's reliability and stability under extreme conditions. The T&T system was subjected to high data volumes and simultaneous queries to replicate peak load scenarios. This involved:

- Data ingestion simulation: Multiple IoT devices and data streams were simulated to test the system's ability to process real-time inputs without bottlenecks.
- Concurrent stakeholder access: Scenarios involving simultaneous API requests from the ICT tool, representing aggregated stakeholder queries, were modelled to assess query response times and processing efficiency. This approach facilitates that the system's interactions with the ICT tool effectively handle the high demand originating from multiple stakeholders accessing data indirectly.

Key Outcomes:

- The system maintained operational stability during stress tests, processing up to 150% of expected peak workloads without significant degradation in performance.
- Areas requiring optimization, such as API latency during high-load periods, were identified and addressed through improved caching and resource allocation.

5.3.2 Load testing:

Using Load testing, BioReCer measured and evaluated the scalability and robustness of the T&T system. It involved gradually increasing the volume of data inputs and the number of user requests to observe how the system performed under typical and elevated operational loads. This testing method showed that the system could dynamically adapt to growing demands without compromising performance or reliability.

Simulating incremental data streams

During load testing, simulated IoT data streams were introduced in increments to assess the system's ability to process real-time inputs while maintaining consistent performance. These data streams included:

- Environmental metrics such as temperature and humidity readings from IoT sensors.
- Real-time data collected from APIs connected to WP5's ICT tool.

The processing capabilities of Apache Spark's Structured Streaming module were tested, evaluating its ability to handle simultaneous data streams without latency. Key focus areas that we included:

- Processing speed: Confirming that the system computes metrics, such as CIs, within acceptable timeframes as data streams increase.
- System throughput: Measuring the number of data packets processed per second under varying loads.

Scaling user requests

Although stakeholder interactions occur indirectly through WP5's ICT tool, simulated API calls were used to replicate aggregated user requests. These requests mimicked real-world scenarios where multiple stakeholders query the system simultaneously for circularity metrics, compliance data, or performance summaries. Key elements tested included:

- Concurrent API requests: Simulating a growing number of simultaneous API queries to assess the system's response times.
- Peak load scenarios: Evaluating the system during high-demand periods.

Identifying bottlenecks

Load testing provided valuable insights into potential bottlenecks, such as:

1. Latency during peak loads:
   - Observed slight increases in query response times when data volumes exceeded 120% of normal load.
   - Optimization through database indexing and caching strategies reduced latency to acceptable levels.
2. Resource utilization:
   High CPU and memory usage were identified during complex circularity computations. This was mitigated by distributing workloads across additional nodes in Apache Spark's framework.

Key outcomes and improvements

1. Scalable data handling:
   The system demonstrated the ability to handle up to 150% of expected workloads without significant performance degradation. This result demonstrates readiness for future expansions.
2. Optimized query response:

Improvements in data partitioning and API throughput resulted in consistent response times even during peak loads.

3. Proactive resource allocation:
   The system dynamically allocated additional computational resources during high-demand scenarios, preventing service interruptions.

These tests not only highlighted areas for optimization but also reinforced the system's capability to adapt to growing datasets and user interactions efficiently.

5.3.3 Functional Testing

Functional testing focused on verifying that all components of the T&T system, including data ingestion, computation, and reporting, performed as intended. This included:

- Validation of computation (i.e., CIs): Outputs generated by the T&T system were cross-checked against WP2's predefined formulas to confirm accuracy.
- Real-time analytics: The system's ability to compute sustainability metrics and generate compliance summaries in real time was thoroughly tested.

Key Outcomes:

- All outputs matched expected results, with no discrepancies.
- Real-time analytics features performed seamlessly, with average query response times of less than 2 seconds.

5.3.4 End-to-End system validation

End-to-end testing shows that all components of the T&T system worked cohesively, from data ingestion to output visualization. This included:

- Simulated data flows: Test datasets were processed through the system, replicating real-world scenarios such as sustainability assessments and regulatory audits.
- Feedback loops: Results were compared against manual calculations and stakeholder expectations to refine workflows and improve system accuracy.

Key Outcomes:

- The system successfully processed complex datasets involving multiple stakeholders, demonstrating its readiness for real-world deployment.
- Feedback from pilot tests highlighted enhancements to improve usability and streamline data workflows.

5.3.5 Automated error detection

Automated error detection mechanisms were integrated into the testing framework to identify and address potential issues during validation:

- Data anomaly detection: Algorithms flagged missing, inconsistent, or outlier data points (Intentionally created to check the system) for correction during preprocessing.
- Performance alerts: Automated monitoring tools provided real-time notifications for issues such as increased API latency or resource exhaustion.

Key Outcomes:

- Automated tools significantly reduced manual intervention, improving the efficiency and reliability of the validation process.
- Proactive error detection minimized disruptions during system testing and deployment.

Through comprehensive testing and validation, WP3 T&T system demonstrated its robustness, scalability, and accuracy under diverse conditions. These processes verified the system's current capabilities and provided insights for ongoing optimization and future development.

6. Conclusion

Deliverable D3.2 represented a significant milestone, transitioning theoretical insights from D3.1 into actionable applications within the T&T system. The efforts outlined in this document highlight strategic planning, innovative technology deployment, and extensive collaboration among consortium partners. The cloud-based software application developed in this deliverable exemplifies the commitment to sustainability, transparency, and accountability within bio-based supply chains.

Built on a centralized blockchain infrastructure, the T&T system introduced robust mechanisms for data immutability, real-time monitoring, and seamless integration with WP5's ICT tools. Advanced technologies such as HSI, IoT-enabled data capture, and API-driven interoperability addressed the multifaceted challenges of modern bio-based sectors, including fisheries, forestry, and etc. This innovative approach provided stakeholders with access to accurate, validated, and actionable data, enhancing decision-making processes and compliance with sustainability and regulatory standards.

The implementation of CIs, derived from WP2's methodologies, equipped stakeholders with quantifiable metrics to evaluate environmental performance. This integration bridged the gap between theoretical assessments and practical applications, positioning the T&T system as a replicable model for advancing circular economy principles across diverse bio-based industries.

The T&T system's design prioritized scalability and adaptability, maintaining the framework accommodates future expansions in stakeholder participation, technological advancements, and data volumes. This future-ready architecture aligns with BioReCer's vision to establish a resilient and sustainable bio-based economy that addresses evolving global needs.

The collaborative spirit within the BioReCer consortium has been pivotal in achieving the goals of D3.2. From integrating sector-specific tools to developing user-friendly interfaces and training initiatives, this deliverable reflects a collective effort to empower stakeholders with practical solutions that drive meaningful change.

7. List of Abbreviations

AES	Advanced Encryption Standard
API	Application Programming Interface
BA	Batch
BF	Biological Feedstock
BioReCer	Biological Resources Certification Schemes
CE	Circular Economy
CI	Circularity Indicator
COR	Conformity
CORE	Customer Optimization Route and Evaluation
CPU	Central Processing Unit
DLT	Distributed Ledger Technology
EC	European Commission
EU	European Union
GDPR	General Data Protection Regulation
HSI	Hyperspectral Imaging
ICT	Information and Communication Technology
ID	Identification
ISO	International Organization for Standardisation
ISO/IEC	International Organization for Standardization/the International Electrotechnical Commission
JS	JavaScript
KPIs	Key Performance Indicators
MFA	Multi-Factor Authentication
MTBF	Mean Time Between Failures
MTTR	Mean time to recovery
NIR	Near Infrared
NGSI-LD	Next Generation Service Interfaces-Linked Data
OAuth	Open Authorization
RBAC	Role-Based Access Control
RFID	Radio Frequency Identification

SHA	Secure Hash Algorithm
SQL	Structured Query Language
T&T	Track and Traceability
TLS/SSL	Transport Layer Security and Secure Sockets Layer

8. References

[1] H. Y. Ting and A. F. Atanda, "Enhancing Supply Chain Traceability through Blockchain and IoT Integration: A Comprehensive Review," *Green Intelligent Systems and Applications,* vol. 4, no. 1, pp. 11-28, 2024.

[2] A. Schlemitz and V. Mezhuyev, "Approaches for data collection and process standardization in smart manufacturing: Systematic literature review," *Journal of Industrial Information Integration,* vol. 38, p. 100578, 2024.

[3] D. Song, K. {. Silva}, M. D. Brooks and M. Kamruzzaman, "Biomass prediction based on hyperspectral images of the Arabidopsis canopy," *Computers and Electronics in Agriculture,* vol. 210, p. 107939, 2023.

[4] N. TL, V. C, K. A, J. CB, K. A, K. SB, A. A and Q. MT, "Advances in real time smart monitoring of environmental parameters using IoT and sensors," *Advances in real time smart monitoring of environmental parameters using IoT and sensors. Heliyon,* vol. 10, no. 7, 2024 .

[5] B. Unhelkar, S. Joshi, M. Sharma, S. Prakash, A. K. Mani and M. Prasad, "Enhancing supply chain performance using RFID technology and decision support systems in the industry 4.0–A systematic literature review," *International Journal of Information Management Data Insights,* vol. 2, no. 2, p. 100084, 2022.

[6] K. Einola and M. Alvesson, "Behind the Numbers: Questioning Questionnaires," *Journal of Management Inquiry,* vol. 30, no. 1, 2020.

[7] D. L. Vannette, "Questionnaire Design: Theory and Best Practices - MAXIMIZING THE RELIABILITY AND VALIDITY OF SURVEY DATA COLLECTION," 15 September 2014. [Online]. Available: https://docslib.org/doc/6785571/questionnaire-design-theory-and-best-practices. [Accessed 20 Febrary 2025].

[8] M. Koot, M. R. Mes and M. E. Iacob, "A systematic literature review of supply chain decision making supported by the Internet of Things and Big Data Analytics," *Computers & Industrial Engineering,* vol. 154, p. 107076, 2021.

[9] H. Aguinis, N. S. Hill and J. R. Bailey, "Best Practices in Data Collection and Preparation: Recommendations for Reviewers, Editors, and Authors," *Feature Topic on Reviewer Resources,* vol. 24, no. 4, p. 678–693, 2021.

[10] S. Sadowski and P. Spachos, "Wireless Technologies for Agricultural Monitoring using Internet of Things Devices with Energy Harvesting Capabilities," *arXiv:2005.02477,* 2020.

[11] J. Sahoo and K. Barrett, "Internet of Things (IoT) Application Model for Smart Farming," *arXiv:2101.03722,* 2021.

[12] A. L. Selvakumar and C. S. Ganadhas, "The Evaluation Report of SHA-256 Crypt Analysis Hash Function," in *International Conference on Communication Software and Networks, ICCSN*, China, 2009.

[13] S. N. Khan, F. Loukil, C. Ghedira-Guegan, E. Benkhelifa and A. Bani-Hani, "Blockchain smart contracts: Applications, challenges, and future trends," *Peer-to-Peer Networking and Applications,* vol. 14, no. 5, pp. 2901-2925, 2021.

[14] N. Tantalaki, S. Souravlas and M. Roumeliotis, "A review on Big Data real-time stream processing and its scheduling techniques," *International Journal of Parallel Emergent and Distributed Systems ,* vol. 35, no. 5, 2019.

[15] K. S. Prasad, "Real-time Data Streaming using Apache Spark on Fully Configured Hadoop Cluster," *Journal of mechanics of continua and mathematical sciences,* vol. 13, no. 5, 2018.

[16] A. Mousav, P. Adl, R. T. Rakowski, A. Gunasekaran and N. Mirnezami, " Customer optimization route and evaluation (CORE) for product design," *International Journal of Computer Integrated,* vol. 14, no. 2, pp. 236-243, 2001.

[17] A. Gupta and A. Khanna, "A holistic approach to sustainable manufacturing: Rework, green technology, and carbon policies," *Expert Systems with Applications,* vol. 244, p. 122943, 2024.

[18] A. Bhargava, A. Sachdeva, K. Sharma, M. H. Alsharif, P. Uthansakul and M. Uthansakul, "Hyperspectral imaging and its applications: A review," *Heliyon,* vol. 10, no. 12, 2024.

[19] A. Zahra, R. Qureshi, M. Sajjad, F. Sadak, M. Nawaz, H. A. Khan and M. Uzair, "Current advances in imaging spectroscopy and its state-of-the-art applications," *Expert Systems with Applications,* vol. 238, p. 122172, 2024.

[20] M. Waseem and A. Ahmad, "Containerization in Multi-Cloud Environment: Roles, Strategies, Challenges, and Solutions for Effective Implementation," *arXiv:2403.12980v2,* 2025.

[21] B. Kim, S. Lee, Y. Lee, Y. Park and J. Jeong, "Design and Implementation of Cloud Docker Application Architecture Based on Machine Learning in Container Management for Smart Manufacturing," *Appl. Sci.,* vol. 12, p. 6737, 2022.