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)

BioReCer platform reference architecture

Deliverable reference number
D5.1

Work package number
WP5

Due date of deliverable
31.08.23

Actual submission date
28.06.2024

Authors

Name	Organisation	Email
Romain Magnani	EGM	romain.magnani@egm.io
Benoit Orihuela	EGM	benoit.orihuela@egm.io
Marianne Marot	EGM	marianne.marot@egm.io
Franck Le Gall	EGM	franck.le-gall@egm.io
Pablo Martinez Pavon	ANFACO	pmartinez@anfaco.es
Miquel Tome	CETAQUA	miquel.tome@cetaqua.com
Alireza Mousavi	BRUNEL	Alireza.Mousavi@brunel.ac.uk
Sebelan Danishvar	BRUNEL	Sebelan.Danishvar@brunel.ac.uk

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

Funded by the European Union

Type

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

Dissemination Level

X	**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

Lead beneficiary

EGM

Contributing beneficiaries

ANFACO, BRUNEL, CETAQUA

Internal reviewer(s)

Name	Organization	Email
Monica Pandey	MEO	pandey@meo-carbon.com
Lucia Gonzalez	CETAQUA	lucia-jimena.gonzalez@cetaqua.com

Change log

Date	Issue/Version	Reason for change
16/08/2023	0.1	Incomplete first draft
18/08/2023	0.2	Reviews and comments
24/08/2023	0.3	Formatting & Quality checks
07/09/2023	1.0	Finalising section 2
19/06/2024	2.0	Reviewed version after PR1 meeting

Report

Action Title	Biological Resources Certifications Schemes
Action Acronym	BioReCer
Action Number	101060684
Deliverable Identifier	D5.1
Deliverable Title	BioReCer Platform Reference Architecture
Document Status	Final Version
Version	2.0
Authors	WP5 Partners
Lead Beneficiary	EGM
Deliverable Type	Other
Dissemination Level	public
Format	Word, .doc
Due Month	12
Date	28.06.2024
DOI	
Keywords	Architecture, NGSI-LD, data model, requirements, BIT, Platform

Document History

Version	Description	Date
0.1	Incomplete first draft	11/08/2023
0.2	Incomplete first draft - reviews	18/08/2023
0.3	Formatting and quality check	24/08/2023
1.0	Finalising section 2	07/09/2023
2.0	Two revisions were made following the comments from external evaluators during the project evaluation meeting: i) a terminological inconsistency regarding the BioReCer platform (see section 5.6 in page 61); and ii) the switch of our cloud deployments from Amazon Web Services to OVH Cloud (see section 6 framed introduction in page 64).	19/06/2024

Table of contents

List of figures

Publishable executive summary

The current D5.1 deliverable describes the ongoing work on designing the BioReCer ICT Tool (BIT) reference architecture. The BIT is the technical foundation of the BioReCer project which aims at enhancing the environmental performance, traceability, and social acceptance of bioproducts.

This document shows how the platform reference architecture relies on open standards, carefully selected for their relevance regarding context-rich information systems within the Internet of Things, performance, and interoperability, and how it allows collaboration among stakeholders and ease decision-making.

The report also outlines the efforts to design a relevant data model for the platform, based on Key Performance Indicators (KPIs) and requirements highlighted by other work package research. Besides, this deliverable raises security concerns regarding authorizations, certifications, data breach, data veracity and trust, and it proposes a way to tackle these issues.

Multiple components, tools and user interfaces are presented in this document which are expected to make it possible to achieve the project goal by empowering stakeholders, such as data visualization tools, data storage components, monitoring applications, and marketplace dedicated to bioproducts.

Eventually, the D5.1 deliverable explains how the BIT platform is expected to be deployed, with scalability and performance as a priority for the whole BioReCer ecosystem. It displays the cloud native approach chosen to overcome the challenge of being able to smoothly run all these layers which compose the BioReCer reference architecture.

1 Introduction

With a core mission to enhance the environmental performance, traceability, and social acceptance of bioproducts, BioReCer encompasses a comprehensive framework meant to redefine the way biological feedstocks are utilised. At the heart of this endeavour lies the BioReCer ICT Tool, a dynamic ecosystem that gathers cutting-edge technologies to bring forth a new era of bio-based industry practices. This deliverable is a first milestone on how to tackle the technical challenge behind accomplishing such an ambitious project.

Therefore, in the overall architecture description part, we focus on how the ETSI NGSI-LD specifications lay the groundwork for the entire architectural edifice. We describe the NGSI-LD information model, with its foundational concepts of Entity, Relationship, and Property. We also detail its API, exposed by the Stellio context broker, and how it allows us to provide real-time access to contextual information, fostering seamless collaboration among stakeholders and data sources.

The BioReCer ICT tool (BIT), is a digital cornerstone that extends the reach and scope of the project. Through its integration, the platform not only bridges the physical gaps but also empowers stakeholders with real-time access to context-rich information. This interplay between the physical and digital worlds creates a resource for decision-making and sustainable practices.

As we must first establish trust and security, we dive into concepts such as W3C Verifiable Credentials, Decentralized Identifiers (DID) and Distributed Ledger Technology (DLT). We then start thinking about how these connect with ISCC certificates, machine learning, artificial intelligence, and monetization.

In this deliverable, we delve into the granular intricacies of the BioReCer technical platform. From the conceptual foundations rooted in ETSI NGSI-LD specifications to the data layer that acts as the bedrock of insights, we describe each component's role and purpose. We will go through identified ontologies and data models, foreseen data sources, technical components, user interfaces and deployment strategies, that compose BioReCer's ICT tool, and that will enable traceability, and acceptance of bioproducts.

2 Requirements for BIT

Regarding requirements for BIT, two main subjects are considered at this stage: Data acquisition and Stakeholder needs. Knowing that data acquisition depends on the stakeholder needs, we first need to identify stakeholders' and any potential users' needs, in order to acknowledge which data of which type are necessary for the BIT to run properly and achieve its goal.

2.1 Stakeholders' needs

Relying on WP2 and WP4 groundworks on Material Flow Analysis, a preliminary list of key indicators and Focus Group discussions (see D4.1 and D2.1), four typologies of BIT users are considered as of today (M12). Although those will probably evolve and change during the project, below is a first attempt at describing them and their objectives:

2.1.1 Biomass Producer

The first user's typology identified regroups any individual or employee representing a biomass producer company, specialised in bio-based sector processes revolving around harvesting and manufacturing. As such, they may be also representing owners of productions systems, with specific biomass transformation processes which generate particular products.

Therefore, the BIT platform must help the Biomass Producer:

- To access key indicators on his own company performances and allow him to effectively manage and leverage his data. In doing so, the BIT must have a key role in decision-making regarding the reduction of energy consumption of production systems, increase of renewable energy contribution in the transformation process and its efficiency, improve visibility on generated product and byproduct quality, contaminants presence, composition…
- By improving bio-based and circular economy acceptability, the BIT platform must have positive side effects on the industry. Improvement of citizens' perception of bio-based goods may lead to a structural shift and trigger an increase in jobs demand related to bio-based industry and revenues for firms involved in the bio-based sector.
- By facilitating synergies between stakeholders, with the transporter and the certifier, but also with the end consumer.

2.1.2 Biomass Transporter

As a waste management company representative, the Biomass Transporter user shares the same kind of objectives as the Biomass Producer. He is specialised in bio-based sector processes revolving around transportation and logistics. As such, the company may be owner of a fleet of several specific transport vehicles, designed to move particular biomasses, and owner of different storage modes.

Therefore, the BIT platform must help the Biomass Transporter:

- To access key indicators on his own company performances and allow him to effectively manage and leverage his data. In doing so, the BIT must be a cornerstone of decision making regarding: (i) the reduction in energy consumption of the transporters fleets and the storage modes, by tracking their respective carbon footprint; (ii) the increase of renewable energy contribution in the transportation process and on the storage mode, and their respective efficiency; (iii) the improvement of visibility on the volumes carried, monitoring of chemical usages, failures, and maintenance…
- By improving bio-based and circular economy acceptability, the BIT platform must have positive side effects on the industry. Improvement of citizens' perception of bio-based goods, thanks to initiative like the BIT platform, may lead to a structural shift and trigger an increase in jobs demand related to the bio-based industry and revenues for firms involved in the bio-based sector.
- By facilitating synergies between stakeholders, biomass producers and certifiers, but also with the end consumers.
- To reduce the amount of raw materials imported.

2.1.3 Certifier

The Certifier users' typology regroups any individual or employee representing a certification scheme company. He is specialised in sustainability and as such, may develop sustainable solutions thanks to material certifications schemes and supply chain due diligence, carbon footprint improvement guidelines, …

Therefore, the BIT platform must help the Certifier to:

- Improve and facilitate material certification processes
- Complement existing certification schemes for bio-circular products
- Automate or assist in building the conformity assessment of the generated byproduct by providing visibility on Key Performance Indicators from different stakeholders and actors of the platform

2.1.4 Consumer

Finally, the Consumer user typology encapsulates any other kind of users among the BIT platform visitors. It represents any user that tends towards a positive impact on the environment. He considers it to be a very important matter and is curious about the processes behind bioeconomy. He would need to cherry pick information to make an informed decision before buying, or could simply need to educate himself through exploratory features that the BIT could propose regarding bioeconomy processes and data. The BIT must help the Consumer with his concerns regarding:

- The reduction of natural resources depletion
- Having a general positive impact on the environment
- Education about bioeconomy as well as circular economy
- Having a responsible consumer behaviour thanks to smart, data-based, decision-making

Based on those different user's needs and objectives, we now can better appreciate how to approach data acquisition to satisfy them as much as possible, as well as to drive our data models and overall technical architecture, as we will see in later sections.

2.2 Data acquisition

The data acquisition is the second identified main subject to drive the BioReCer ICT Tool requirements. As of today, it is possible to collect most of the required data through the BioReCer Web Application (BWA), described in greater detail in the Components description section.

The purpose of BWA is to provide an easy-to-use interface for acquiring data directly from stakeholders through electronic interfaces, processes, conducting the necessary computations (e.g. Circularity Indexing and Environmental Impact factor), assembling the information on companies, their products and their role in the supply chain, reporting on certifications associated with the company/product and communicating this information in an informative way to all stakeholders/users.

In order to achieve this objective, we classify the data acquisition and perform a fusion process into four stages:

2.2.1 Company Registration

The company registration stage is conducted via an online form available on the world wide web written in combination of HTML, JavaScript, Ajax and associated web technologies allowing all standards known to all web browsers to seamlessly communicate with the user. This application will also be built under the necessary dynamic standards to be compatible

with mobile phones so that the information can be collected via such media. Information such as geographical location that can be automatically collected are also embedded in the application to reduce data entry requirements.

Upon registration, the company needs to be validated. This process is partially automated (i.e., online validation of the data entry format) and partially by validated by auditors who will be ratifying the company with regards to industry standard procedures (e.g., legal, financial…).

Through an API, this information can also be communicated to other recipients of the data within the consortium (e.g., EGM, MECS…). Upon validation, the company will be allocated a Unique Identification Number and a Digital Identity, as is part of the overall Digital Passport and the first item of the T&T data framework which will be linked to the corresponding products and the supply chain. Next step is product registration.

2.2.2 Product registration

Following the validation of the company, the latter will be able to register every product within their company portfolio which will include:

- General information: name, unit, images (if applicable), presentation, price, mode of delivery, amount, …
- Detailed technical specification associated with the product that enables stakeholders to determine the ingredients, composition, physical-chemical properties, …

This information is also part of T&T data that will be added (stage 1) for the creation of a unique Digital Passport of the product associated with the company. Through an API, this information will be accessible to all partners requiring the information for evaluation and validation.

2.2.3 Life Cycle Assessment and Environmental Impact Assessment

The BWA will be providing a user interface and the required data acquisition interface with the user in the form of questionnaire consisting of a number of independent and interlinked sections that collects the information vis-a-vis production process, ingredients, recyclability, reusability, energy consumption per unit of product, emissions, methods of processes, handling, storage, hazard profile, certificates … to create the necessary information where the data will be fused and calculations will be made (heuristic and inferential) to calculate and infer carbon footprint (KPI 1), circularity index (KPI 2), environmental conformity (KPI 3 = KPI 1 & KPI 2), and level of certainty on the information (based on how complete and verifiable the information gathered/provided is) (KPI 4).

Upon completion of this stage, a full Digital Identity of the product will be generated as the information from Stages 1-3 will be assembled and linked to a unique identification number presented as a QR/Barcode.

2.2.4 Supply Chain Creation with Concatenation

This stage is about Logical/Computational Assembly of the information: BWA has the capability for the user to define and create a full supply chain (e.g., supplier of raw materials, production/manufacturing processes (plant), transport, and any other members of supply chain) including the information from stages 1-3, thus creating a full trackable and traceable story of the product with KPIs 1-4 information of the full supply chain. The final Digital Passport will enable stakeholders to extract all the necessary details of the product and its total sustainability value pertinent to all 4 KPIs.

These four stages prepare the presentation of the information to the internal and external stakeholders at the required level and format (Customisation): The information of the final results for a product, can be displayed in a clear, informative and engaging interface as vignette of required information and dials, showing environmental, circularity indicators and level of certainty of information (i.e., complete and verifiable) (AI driven).

The above approach can be considered a human instigated entry of data. In our existing case studies, we will explore the current capabilities of automatic data acquisition (e.g., plant control systems) of companies, provided they exist, and we will build the APIs to acquire the information necessary and directly populate the relevant fields of the database (e.g., energy consumption, emission sensors, process steps, distance travelled using transport system GPS…).

The BIT platform may be able to automate the data acquisition and data processing (mentioned in Task 5.1) in its entirety, then the manual data entry of BWA will become redundant and only used where such data acquisition capabilities are not possible.

Both main topics regarding requirements, Stakeholders' needs and Data acquisition, are reported in our BioReCer Data models designs, which are presented later in section 4. The next part delves into the platform's overall technical architecture.

3 Architecture overall description and relevant standards

The BioReCer ICT Tool (BIT) is data driven. In this section, we describe how we intend to be driven by on-going open standards such as NGSI-LD, which clarifies how we bring interoperability and seamless collaboration among stakeholders and data sources to the project. We then delve into other standards such as W3C Verifiable Credentials and Distributed ledger technology to show how we intend to enforce trust and security regarding the platform's data streams. And in doing so, consolidating the BIT requirements described in the previous section.

3.1 ETSI NGSI-LD specification

3.1.1 ETSI

The European Telecommunications Standards Institute (ETSI) is a renowned independent, not-for-profit standards organisation that develops globally applicable standards for information and communication technologies (ICT).

As a recognized European Standards Organization (ESO), ETSI operates in collaboration with various stakeholders, including industry, academia, and regulatory bodies, to establish a comprehensive framework of standards that enables seamless interoperability and facilitates technological advancements.

In the context of proliferation of diverse data sources, platforms, and applications in smart cities which often results in data silos and fragmented systems that hinder efficient utilisation of data assets, ETSI's NGSI-LD standard (Next Generation Service Interface for Linked Data) is being developed in order to tackle this issue. Indeed, by creating the NGSI-LD standard, ETSI aims to establish a common data model and API specifications, enabling seamless exchange and integration of data across different applications and platforms. This standardisation fosters interoperability, scalability, and openness, allowing stakeholders to effectively manage and leverage their data.

3.1.2 Introduction to NGSI-LD

The basic concepts of the NGSI-LD information model are Entity, Relationship and Property. The data model is derived from property graphs, with formal semantics defined on the basis of RDF and the Semantic Web framework. It can be serialised in the JSON-LD format. Each entity and relationship must be assigned a unique IRI (Internationalised Resource Identifier), making the corresponding data exportable to the "Data in the Web". The suffix -LD refers precisely to this affiliation to the universe of Linked Data.

- **An NGSI-LD entity** is the informational representation of something (a referent) that is supposed to exist in the real world, outside of the NGSI-LD platform. This referent does not need to be something strictly physical (it can be a legal or administrative entity) or stand-alone (it can be a system-level distributed construct). Any instance of such an entity is expected to be uniquely identified by a URI and characterised by a reference to one or more NGSI-LD entity types. In property graph language, it is a node.
- **An NGSI-LD Property** is an instance that associates a characteristic, an NGSI-LD value with an NGSI-LD entity, an NGSI-LD relationship, or another NGSI-LD property.
- **An NGSI-LD Relationship** is a directed link between a subject (starting point), which can be an NGSI-LD Entity, an NGSI-LD Property, or other NGSI-LD Relationship, and an object (endpoint), which is an NGSI-LD entity.
- **An NGSI-LD value** is a JSON value (that is, a string, a number, true or false, an object, or an array), or a value of type JSON-LD (that is, say a string as a lexical form of the value with a type, defined by an XSD base type or more generally an IRI), or a JSON-LD structured value (i.e. a set, a list, a labelled string in language).

3.1.3 The NGSI-LD API

The NGSI-LD Context Information Management API enables users to provide, consume, and subscribe the contextual information in multiple scenarios and involving multiple stakeholders. It allows near real-time access to information from many different sources (not just IoT data sources), called "context sources", as well as the publication of this information via interoperable data publication platforms.

It provides advanced geo-temporal queries and includes subscription mechanisms, so that content consumers are notified when content matching certain constraints becomes available.

The API is designed to be architecture-independent (central, distributed, federated, or combinations thereof), so that the applications that produce and consume information do not have to be tailored to the specifics of the system that distribute/broker context information for them.

API operations include:

- Information operations, concerning Provision (creation of NGSI-LD entities and update of their attributes), Consumption (query of NGSI-LD entities) and Subscription (subscription to specific information, with specified constraints, in order to be notified when corresponding entities appear, carrying the specified information).

- Context sources operations, concerned by Recording (making a new source of context information available in the global distributed system, by recording it) and Discovery (querying the system on the recorded context sources, that offer information of a specified type).

3.1.4 Benefits of NGSI-LD

The benefits of the NGSI-LD standard for BioReCer are:

- The information model and its associated API allow the "data control" of the platform.
- The information model and its API natively manage (contextualization):
    - The temporal characteristics (time-series) of the data with an API providing aggregation functionalities.
    - Spatial features, compatible with the OGC GeoJSON standard, simplify data exchange with geographic information systems (GIS) and facilitate cartographic representations.
    - Relations between data and datasets
- The API not only provides advanced querying possibilities but also notifications on spatio-temporal alerts (e.g.: generate an automatic notification when the average flow rate observed over a week in a given territory is greater than a given threshold).
- The existence of a FIWARE community allows the reuse of open-source bricks as well as data models.
- The graph approach of the data model makes it easy to extend it to adapt to new use cases.

3.2 W3C Verifiable Credentials and related specifications

A lot of effort is currently ongoing for a new generation of specifications aiming at improving the way the credentials are managed.

One of the main points is that Web 2.0 has seen platform services emerge, with the consequence that our online identity and personal data is centrally managed by a few big platforms. With the uprising of the Web 3.0, there appears a need for new privacy-preserving and decentralised services that give users back the control on their credentials.

Figure 1. EBSI Verifiable Credentials Explained[1]

To support this evolution, a new trust model is needed where the central role of the platform is replaced by a decentralised model where the information can be trusted because it can be verifiable by everyone. This new trust model is one of the main objectives of the Verifiable Credentials Data Model specification[2] from the W3C. Indeed, it allows to express credentials on the Web in a way that is cryptographically secure, privacy respecting and machine verifiable. As clearly explained in the Verifiable Credentials Data Model specification (see W3 specification section 1.1), in the physical world, a credentials might consist of information related to:

- Identifying the subject of the credential (for example, a photo, name, or identification number)
- The issuing authority (for example, a city government, national agency, or certification body)
- The type of credential (for example, a Dutch passport, an American driving licence, or a health insurance card)
- Specific attributes or properties being asserted by the issuing authority about the subject (for example, nationality, the classes of vehicle entitled to drive, or date of birth)
- How the credential was derived
- Constraints on the credential (for example, validity period, or terms of use).

A verifiable credential can represent all of the same information that a physical credential represents. The addition of technologies, such as digital signatures, makes verifiable credentials more tamper-evident and more trustworthy than their physical counterparts. A verifiable credential can be used to create verifiable presentations. They are derived from

[1] https://ec.europa.eu/digital-building-blocks/wikis/download/attachments/597952490/Chapter%200%20-%20Verifiable%20Credentials%20An%20introduction.pdf?version=1&modificationDate=1676459051355&api=v2

[2] https://www.w3.org/TR/vc-data-model-2.0/

verifiable credentials and can contain one or more claims (or a selective disclosure of a claim) for one or more verifiable credentials.


```
{
  "@context": [
            "https://www.w3.org/ns/credentials/v2",
            "https://www.w3.org/ns/credentials/examples/v2",
            "https://w3id.org/security/suites/ed25519-2020/v1"
  ],
  "id": "http://university.example/credentials/3732",
  "type": [
            "VerifiableCredential",
            "ExampleDegreeCredential"
  ],
  "issuer": {
            "id": "did:example:76e12ec712ebc6f1c221ebfeb1f",
            "name": "Example University"
  },
  "validFrom": "2010-01-01T19:23:24Z",
  "credentialSubject": {
            "id": "did:example:ebfeb1f712ebc6f1c276e12ec21",
            "degree": {
            "type": "ExampleBachelorDegree",
            "name": "Bachelor of Science and Arts"
            }
  },
  "proof": {
            "type": "Ed25519Signature2020",
            "created": "2023-08-09T23:57:37Z",
            "verificationMethod": "did:example:76e12ec712ebc6f1c221ebfeb1f#key-1",
            "proofPurpose": "assertionMethod",
            "proofValue": "z4uEQ5dfx7LeynNC8vgULxGfVB3gz5ouTAKWzzA3Hmf8HqngJ788gEzZ
xKyCx2qo2iZViAggxjqoSAFLH9RK1GFeN"
  }
}
```


Figure 2. Example of a verifiable credential

In the operational and daily usages of verifiable credentials, the main implied roles are depicted in the figure below:

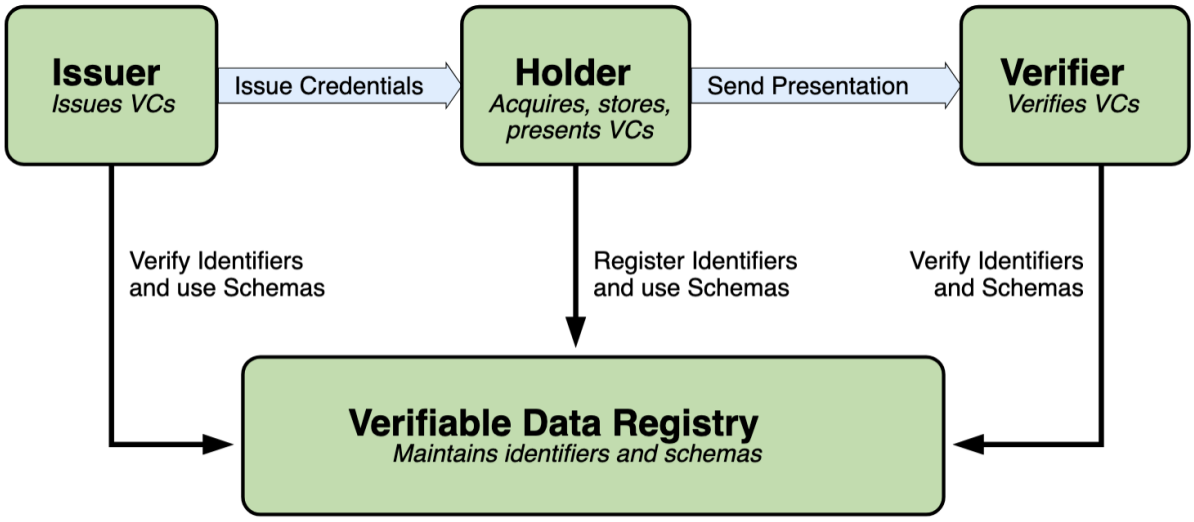


Figure 3. Main roles implied in the lifecycle of a Verifiable Credential

In more details, the roles are the following:

- Issuer: issues verifiable credentials containing claims about one or more subjects and transmits them to a holder. Typical issuers are governments, corporations, …
- Holder: owns one or more verifiable credentials and generates verifiable presentations from them. Typical holders are customers, clients, employees, …
- Subject: an entity for which claims are made. Often, the holder of a verifiable credential is also the subject, but it may not be the case (a parent, acting like a holder, may hold the verifiable credential of its child, here being the subject).
- Verifier: receives one or more verifiable credentials, optionally packaged inside a verifiable presentation. Typical verifiers are employers, websites, …
- Verifiable Data Registry: a system that can mediate the creation and verification of identifiers, keys, and other relevant data that may be required to use verifiable credentials. Typical verifiable data registries are government ID databases, distributed ledgers, ...

Applied to the context of the BioReCer project, verifiable credentials are a great opportunity to bring more confidence and trust to the products exchanged inside the BIT (verifying the origin of a - organic - product is one the common use cases for the verifiable credentials), more precisely by leveraging the certifications schemes that are part of the project.

It is also worth to note that, to support this new paradigm of verifiable credentials, the W3C published other accompanying specifications and notes that will be leveraged inside the BIT platform to provide the more trustworthy, securing and privacy preserving platform possible:

- Decentralized Identifiers (DIDs) v1.0[3]: a DID is a portable URL-based identifier associated with an entity, designed to be decoupled from any central registry or identity provider. It is then resolvable into a DID document that contains cryptographic materials, verification methods, … It is used in verifiable credentials and associated with subjects. This specification elaborates and details the DID concept and will be used to guide us for a good design and implementation of DIDs inside the BIT platform.
- Verifiable Credentials Use Cases[4]: provides guidance and examples on how to make the best use of the verifiable credentials. It will help us ensure the BIT platform is perfectly aligned with the state of the art in security and digital identity.

Concerning the concrete implementation of the mechanisms and behaviours specified in the Verifiable Credentials specification, it is supported by the implementation of some OpenID Connect specifications, and more precisely:

- Self-Issued OpenID Provider v2 (OpenID Connect SIOP v2)[5]
- OpenID for Verifiable Presentations - draft 18[6]
- OpenID Connect for Verifiable Credentials Issuance[7]

In the scope of the BioReCer project, the objective is definitely not to implement these different OpenID specifications, but to leverage existing implementations (ongoing or mature) in order to focus more on the business added value using these secure and decentralised services.

3.3 DLT, MLOps, monetization

3.3.1 Distributed ledger technology

Distributed ledger technology (DLT) refers to a database of which there are multiple identical copies distributed among several participants and which are updated in a synchronised manner by consensus of the parties. Although the best-known application of this technology relates to crypto-assets (particularly the Bitcoin), in recent years several initiatives have proliferated in the financial sector, particularly in fields involving complex processes and numerous actors (e.g., securities trading and post-trading or foreign trade finance). In comparison with crypto-assets, these initiatives show significant differences in terms of the complexity of the consensus mechanism or the characteristics of participants, which make them easier to implement. Thus, distributed ledgers are starting to be used

[3] https://www.w3.org/TR/did-core/
[4] https://www.w3.org/TR/vc-use-cases/
[5] https://openid.net/specs/openid-connect-self-issued-v2-1_0.html
[6] https://openid.net/specs/openid-4-verifiable-presentations-1_0.html
[7] https://openid.net/specs/openid-4-verifiable-credential-issuance-1_0.html

as a tool that may contribute to reduce costs and increase the traceability, transparency and, in certain circumstances, speed of these processes.

A distributed ledger is basically a decentralised, single database that is managed by several participants. It is a database of which there are multiple identical copies distributed among several participants and which are updated in a synchronised manner. An important difference between a distributed ledger and a "traditional" distributed database lies in the updating procedure: while in a traditional distributed database participants trust each other and cooperate to maintain data consistency, in a distributed ledger the parties do not trust each other completely (or there are conflicting interests) and, accordingly, a mechanism needs to be implemented to collectively verify ledgers before they are shared. In other words, the updates are not performed by a central authority, but by consensus among the parties, in accordance with a set of rules or procedures accepted by all.

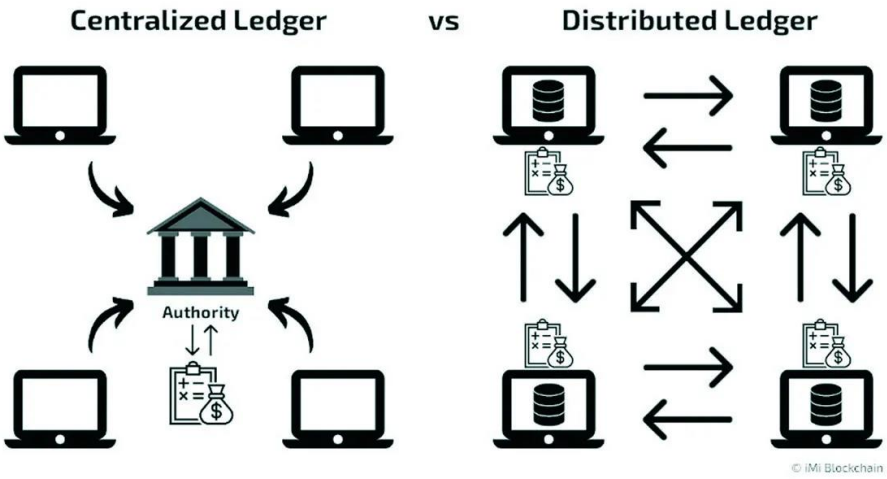


Figure 4. Distributed Ledger Technology introduction[8] (1)

3.3.1.1 How does DLT work?

DLT is basically the result of combining three already existing technologies:

- Peer-to-peer (P2P) networks: In these models, each network participant (node) acts simultaneously as client and server, contributing and consuming resources. This technology became popular in 1999 with the launch of Napster, a software which basically allowed its users to share music.

[8] Romero Ugarte, J. L. (2018). Distributed ledger technology (DLT): introduction. *Banco de Espana Article*, *19*, 18.

- Cryptography: Specifically, asymmetric cryptography,6 which allows for the secure exchange of information between two parties. It is used to authenticate the sender, to ensure the integrity of the message and, by means of encryption, to prevent third parties from accessing the information in the event they manage to intercept it.

- Consensus algorithms: They allow several participants, who may not know or trust each other, to reach an agreement to add new entries to the ledger. There are different methods to reach such consensus, i.e., to ensure that the ledgers of all the network participants are identical and that there is no fraud or duplication of information. The most popular one is "proof of work" (PoW), colloquially known as the mining process, which involves solving complex computational problems for the validation and creation of each new block in the chain. This mechanism, generalised by Bitcoin, entails consuming a large amount of energy in validating transactions (some estimates point to annual consumption of 71.12 TWh, similar to that of Chile, although there are no accurate data in this respect) and involves lengthy processing, which has led to the search for other more efficient mechanisms (e.g. "Proof of stake").

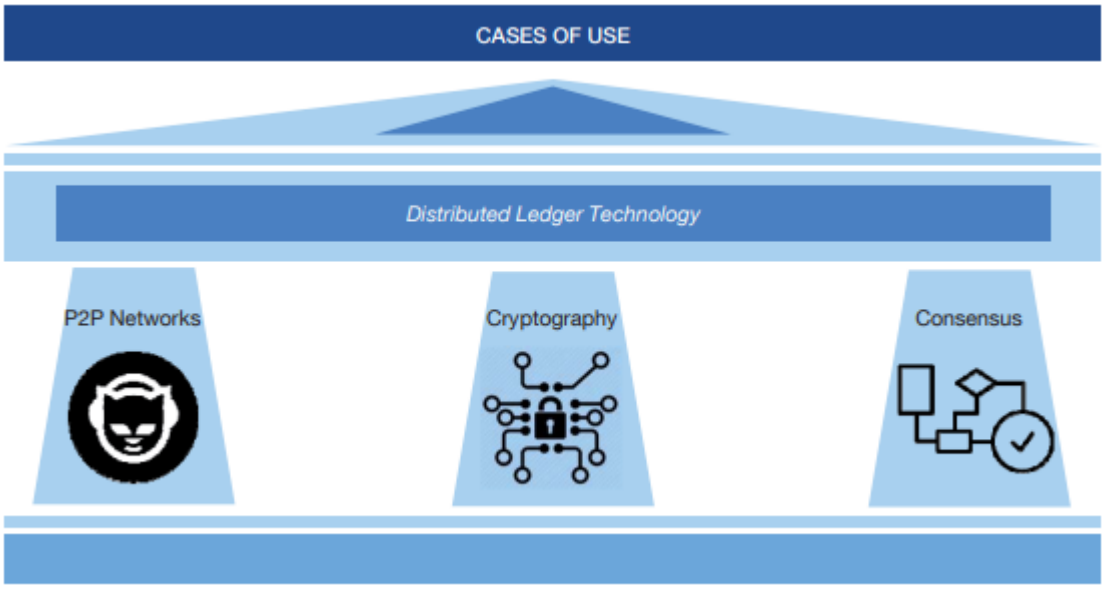


Figure 5. Distributed Ledger Technology introduction[9] (2)

The storage, maintenance and updating of ledgers in a distributed ledger is the core of the technology. The responsibility for updating the ledgers is distributed among the nodes,

[9] Romero Ugarte, J. L. (2018). Distributed ledger technology (DLT): introduction. *Banco de Espana Article*, *19*, 18.

who may be located in different environments, institutions or jurisdictions. Accordingly, changes in the ledger are often not updated simultaneously at all the nodes and there may be long wait times until all the versions are in sync.

3.3.1.2 What is it for?

DLT technology has potential in many different areas, although especially in those in which there are many actors and a lack of trust between the parties. Currently, the number of projects is proliferating, some of which are merely trials, although this does not necessarily mean that this technology is optimal for the processes involved.

The main stimulus to DLT technology has come from its applications as a medium for the exchange of crypto-assets, but its transforming potential is greater. In recent years, both the financial industry and certain authorities have launched many projects to experiment with this technology in different areas.

3.3.2 Monetization of intellectual property

In an increasingly global, connected, and digital world, the management, protection, enforcement and monetization of intellectual property has never been so challenging and critical at the same time. Challenging because intellectual property, especially in a digital form, can be easily copied, deployed, stolen or misappropriated. Critical because nowadays intellectual property is everywhere, it is present in all areas of economic activities, it enables companies to create competitive moat and its attached monetary value is material. Knowledge should also be openly available; therefore, intellectual property law is basically about finding the right balance between authors interests (protection, enforcement, monetization) and users' interests (usage, access). Blockchain and its underlying technology of distributed ledgers has the potential to disrupt the way intellectual properties are managed, protected and monetized. Where the access and distribution of content has been revolutionised by the internet, distributed ledger technology might offer an alternative path helping intellectual property law enter the digital age and address its original intent which is at its core to protect and reward creators.

3.3.3 What is MLOps?

MLOps is the short term for machine learning operations, and it represents a set of practices that aim to simplify workflow processes and automate machine learning and deep learning deployments. It accomplishes the deployment and maintenance of models reliably and efficiently for production, at a large scale.
MLOps is slowly evolving into an independent approach to the machine learning lifecycle that includes all steps – from data gathering to governance and monitoring. It will become

a standard as artificial intelligence is moving towards becoming part of everyday business, rather than just an innovative activity.

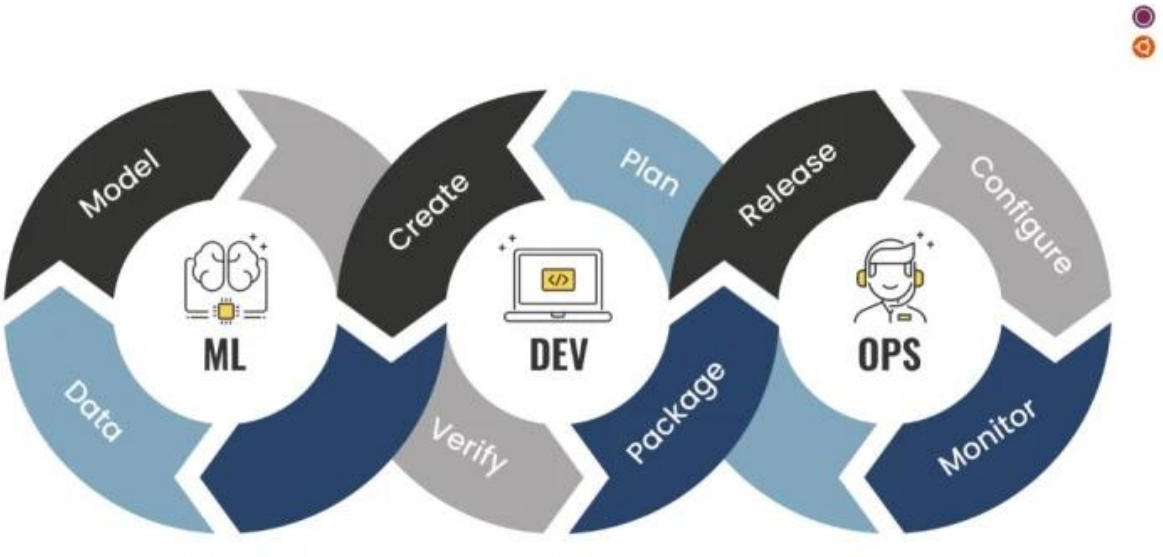


Figure 6. What is MLOps[10]

Machine learning development pipelines can see changes at three levels: data, machine learning model and code. When it comes to MLOps principles, they are designed to impact the ML-based software on one of these three levels.

The MLOps principles focus on:

- Versioning: It treats ML scripts, models and datasets as essential parts of DevOps processes. It tracks data and model versioning, using system controls and alerting changes.
- Testing: It needs to be performed at all levels of machine learning systems, having a different scope when ensuring performance and expected outcomes.
- Automation: The level of automation determines the level of maturity of the ML initiative. The objective of any MLOps team is to automate the deployment of ML models.
- Reproducibility: Having reproducible and identical results in a machine learning workflow, given the same input, is a key MLOps principle.
- Deployment: Model deployment should be done based on experiment tracking, which includes feature stores, containerisation of the ML stack and the ability to run on-prem, on the cloud or at the edge.

[10] https://canonical.com/blog/what-is-mlops

- Monitoring: Ensuring that ML models perform as expected, once deployed, is essential. Monitoring covers changes around dependencies, data, source systems and upgrades.

MLOps or machine learning operations is, in fact, a set of practices that aim to simplify workflow processes and automate machine learning and deep learning deployments. It accomplishes the deployment and maintenance of models reliably and efficiently for production, at a large scale. MLOPs play a crucial role in aligning business demands and regulatory requirements. Its benefits include:

- Increased productivity
- Reproducibility
- Cost reduction
- Monitorability

4 The data layer

4.1 Foreseen data sources

For the development of the architecture, different data sources were considered that will be relevant for the development of the BioReCer ICT Tool (BIT).

4.1.1 Circ4Bio Supplier Shop

BRUNEL University, one of the consortium partners, has developed in a previous LIFE Project**,** **CIRCforBIO,** a web tool that can be useful in order to enable biomass producers to manually include their products on the BioReCer ecosystem.

Figure 7. CIRCforBIO main page with different user types[11]

More detailed information about this tool can be found in the section 5.6 later in this document.

The platform is very useful as it allows users to buy and sell products and services, it adapts to different user profiles, and it can tackle the traceability through the different stages of a flow chain including the transport between companies.

Its functioning resembles some e-shops, such as Amazon and eBay. The user must insert in a form different information about each product in order to incorporate it to the market or as part of an established flow chain.

[11] https://circ4bio.com

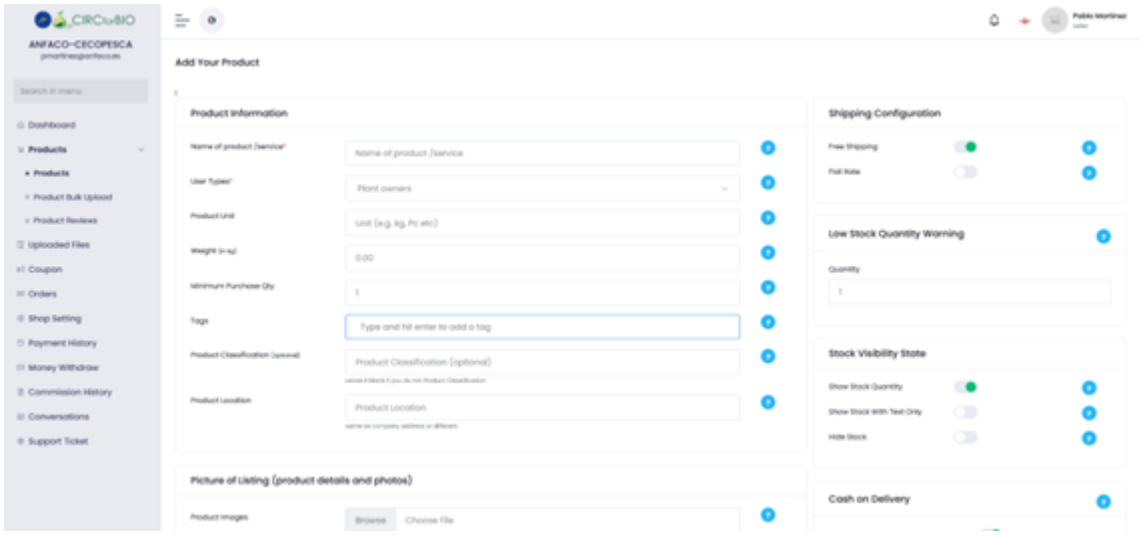


Figure 8. Form to add products to the biomass e-shop[12]

The platform will be modified to meet BIT requirements, as more info would be needed to permit a proper traceability and sustainability assessment of the products. Some of this extra information should be mandatory, but others will remain as optional as some companies may not have all the information to answer the input forms or want to maintain some data as private. This can affect the significance of the results of the biomass Life Cycle assessment (LCA), but it has been planned this way to maximize BIT user engagement and minimize that users withdraw from the tool.

At this point of the project, it is possible that the stakeholders currently involved in the project are not going to represent all the different actors that compose the case studies flow chains. Thus, the platform should give the option, besides selling its transformed product, to establish what biomasses or by-products companies are willing to buy and the maximum price by measurement unit (cubic metres, tonnes, etc.). This will facilitate completing the flow chains.

[12] https://circ4bio.com

4.1.2 International Sustainability and Carbon Certification (ISCC)

As one of the main objectives of the project is to complement the current biological resources certification schemes, the outcomes of the project will be integrated in one of the renowned voluntary certification schemes ISCC which stands for International Sustainability and Carbon Certification.

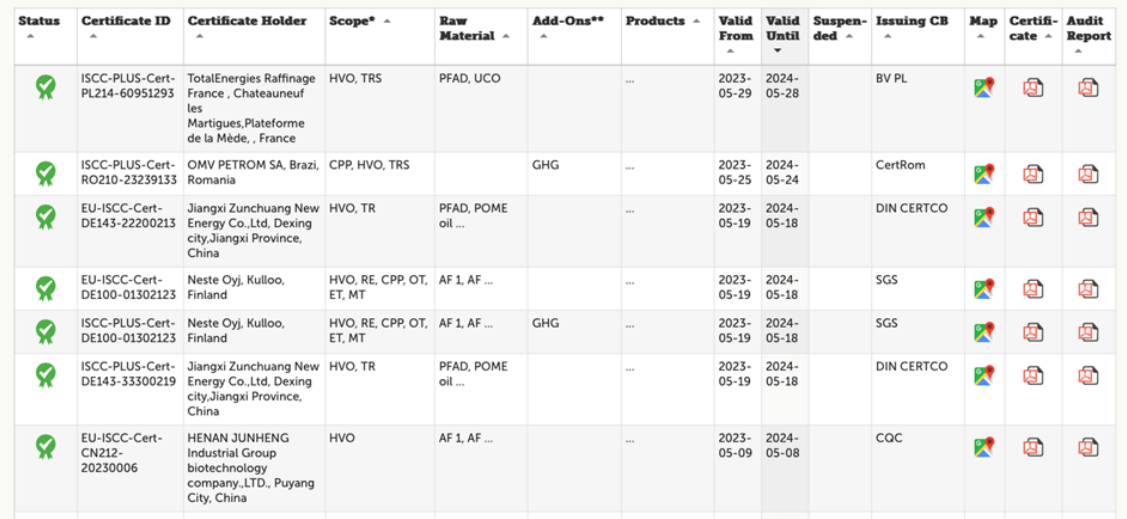
ISCC is an independent multi-stakeholder initiative and a leading certification system supporting sustainable supply chains. ISCC operates in diverse sectors and has different scopes but particularly, ISCC PLUS certification is applicable for the bioeconomy and circular economy for food, feed, chemicals, plastics, packaging, textiles and renewable feedstock derived from a process using renewable energy sources. Thus, based on the biomasses proposed in the project case studies, the ISCC PLUS will be the one that the entities using the platform could implement.

Regarding this certification, two matters of interest should be considered:

1. Integration of Biomass Certificates

In the ISCC website[13] there is a Certificate Database that gathers all the certificates (valid, expired and withdrawn) issued by the certification body. Contacts with ISCC have been established, through MEO Carbon, one of the project consortium members, in order to evaluate the possibility to automatically check if a specific biomass (waste or residue) from a specific producer is already ISCC certified. The proposal was to connect the API of ISCC and ICT tool so the information on the producer, who is already ISCC certificate holder, can be automatically generated in the ICT tool. However, this idea is not applicable because the companies only agree to the ISCC's Terms of Use. The sharing of their information from the ISCC database is not possible due to data privacy.

[13] https://www.iscc-system.org/certification/certificate-database/all-certificates/



Status	Certificate ID	Certificate Holder	Scope*	Raw Material	Add-Ons**	Products	Valid From	Valid Until	Suspended	Issuing CB	Map	Certificate	Audit Report
	ISCC-PLUS-Cert-PL214-60951293	TotalEnergies Raffinage France , Chateauneuf les Martigues,Plateforme de la Mède, , France	HVO, TRS	PFAD, UCO		...	2023-05-29	2024-05-28		BV PL			
	ISCC-PLUS-Cert-RO210-23239133	OMV PETROM SA, Brazi, Romania	CPP, HVO, TRS		GHG	...	2023-05-25	2024-05-24		CertRom			
	EU-ISCC-Cert-DE143-22200213	Jiangxi Zunchuang New Energy Co.,Ltd, Dexing city,Jiangxi Province, China	HVO, TR	PFAD, POME oil ...		...	2023-05-19	2024-05-18		DIN CERTCO			
	EU-ISCC-Cert-DE100-01302123	Neste Oyj, Kulloo, Finland	HVO, RE, CPP, OT, ET, MT	AF 1, AF ...		...	2023-05-19	2024-05-18		SGS			
	ISCC-PLUS-Cert-DE100-01302123	Neste Oyj, Kulloo, Finland	HVO, RE, CPP, OT, ET, MT	AF 1, AF ...	GHG	...	2023-05-19	2024-05-18		SGS			
	ISCC-PLUS-Cert-DE143-33300219	Jiangxi Zunchuang New Energy Co.,Ltd, Dexing city,Jiangxi Province, China	HVO, TR	PFAD, POME oil ...		...	2023-05-19	2024-05-18		DIN CERTCO			
	EU-ISCC-Cert-CN212-20230006	HENAN JUNHENG Industrial Group biotechnology company.,LTD., Puyang City, China	HVO	AF 1, AF ...		...	2023-05-09	2024-05-08		CQC			

Figure 9. Screenshot of the list of all ISCC certificates in the "Certificate Database"

Thus, two options have been considered to automatically include this certificate on the BioReCer ICT Tool (BIT):

- Development of a scraping tool that extracts info from the web and the PDF documents that contain the certificate details.
- The biomass producer would have to manually upload its certificate to the BIT with all the information from the product.

CB Logo

ISCC CERTIFIED PLUS

ISCC PLUS Certificate

Certificate Number: ISCC-PLUS-Cert-CCXXX-XXXXXXXXXX

Name of CB
Harid Street 123, 22222, London
certifies that
ABC Farming Products
Bach Str. 342, 12345, Cologne
complies with the requirements of the certification system
ISCC PLUS
(International Sustainability and Carbon Certification)

Place of the audit
(if different from the legal address of the system user as stated above; only applicable for traders and traders with storage):
Bach Str. 342, 12345, Cologne

This certificate is valid from DD.MM.YYYY to DD.MM.YYYY.

The site of the system user is certified as:
First Gathering Point
Processing Unit

The scope of the certificate includes the following chain of custody options:
(not applicable for paper traders)
Mass Balance

DD.MM.YYYY, City		
Place and date of issue		Stamp, Signature of issuing party

The issuing Certification Body is responsible for the accuracy of this document.
Version / Date: 1 (no adjustments) / 14.08.2023

Page 1 of 2

Figure 10. ISCC PLUS Certificate Template

The first option will be tested in the following months to evaluate its performance and the correctness of the obtained data. In the meanwhile, users can work manually.

2. List of eligible materials for ISCC PLUS Certification

ISCC certifies a wide range of sustainable feedstocks including biomass, renewables and circular materials. In the ISCC website, there is a document with a list of the raw materials that can be eligible to be certified under different scopes of ISCC. Particularly, the material list specific to ISCC PLUS is relevant to BioReCer and can help to automatically identify if raw material introduced on the BIT would be certifiable under ISCC PLUS. Moreover, other materials not considered yet on the list may be added if it is necessary in consultation with ISCC. The list is also updated periodically.

List of material eligible for ISCC PLUS certification

(17 July 2023)

Table 1: Raw material		
Declaration of material on ISCC PLUS certificate	Additional information	Can be classified as waste/residue under ISCC PLUS
Almond		No
Apples		No
Basil		No
Berries (*specification*)	The type of berries should be specified in brackets (e.g. Berries (bilberry), Berries (cranberry), Berries (elderberry), Berries (strawberry))	No
Biobased plastic waste		Yes
Calamus palm (Rattan)		No
Cassava		**No**
Celler glass	Waste from the production of glass fibre	Yes
Chickpeas		No
Contaminated paper and cardboard		Yes
CO2	As specified in the ISCC PLUS system document (includes post-industrial, atmospheric and biogenic CO2)	Yes

Figure 11. Screenshot of List of eligible material for ISCC PLUS as of 17th July 2023

In this case, Meo Carbon coordinated with ISCC and has provided an Excel file with the material list table in order to easily access the data. This information will be integrated and updated regularly in BIT.

During the development of the project, other certification systems might be approached and integrated within the ICT tool.

4.1.3 Automatic data acquisition from users

At this stage of the project, the WP4 partners, who are setting up the BRSP (BioResources Stakeholders Platform); and WP6 case study partners are still contacting potential companies that fit the case studies flow chains and that are interested in being project stakeholders and future users of the BIT tool.

The situation of these companies towards data acquisition and exploitation can be very different due to its company profile, sector, volume of business, digitalization degree, etc. Ergo, one company can be using a commercial ERP/MES system, others have a tailormade local DB based system, work with Excel files or even they may be writing down the data on a notebook or not collecting data at all.

So, in the meanwhile, while the potential stakeholders are approached and involved on the project, WP5 partners are working on the development of a form in order to gather information about what kind of data companies are gathering, how they are acquiring it and where it is stored.

This information is important in order to help us with the identification, design and development of the interfaces that would be needed to automatically acquire data from users and also to plan its integration within the architecture of the BioReCer ICT Tool.

This job is aligned with T5.5. BioReCer DSS interfaces implementation and validation, starting on M12.

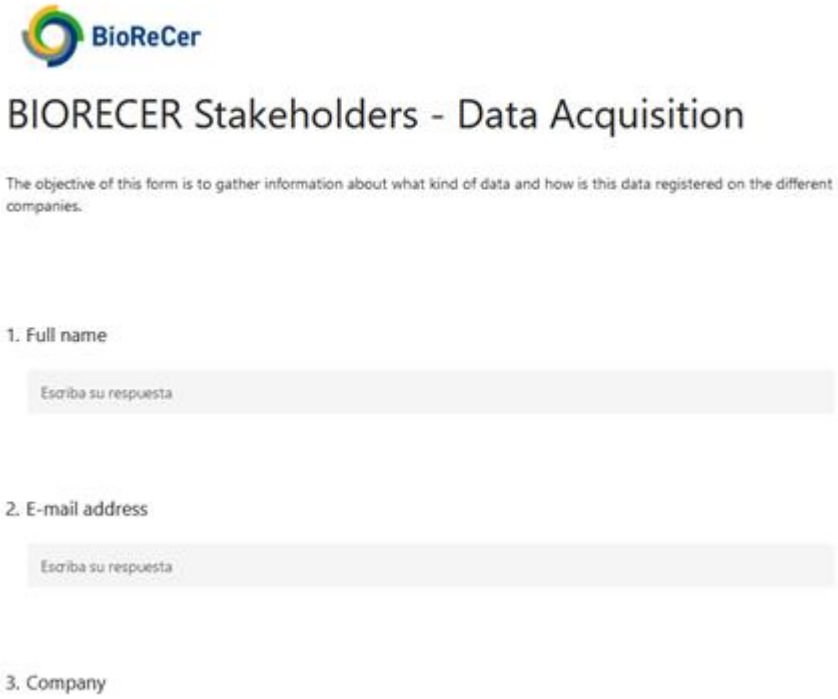


Figure 12. Form for acquiring data from stakeholders (source: ANFACO)

The form might be translated to the local language of each case study to ease its comprehension by the companies involved in the project.

4.1.4 Online available open-source data

On the internet, there are many datasets and repositories which are open source and could be interesting for the development of the BioReCer Project. At this point, the most interesting open data source will be Open Street Map (OSM) or similar, which will provide the maps that are going to be used by a GIS to represent the project logistics flows and the traceability of products.

Other sources like OpenWeatherMap could provide weather data if in some cases meteorological conditions can affect the case studies chains.

Data available through repositories like World Bank Open Data, Google Public Data Explorer or the European Data Portal can also be interesting sources to acquire significant information for the project.

Furthermore, specific data repositories related to circular economy and sustainability can be evaluated. Some of these data sources have been already identified by our colleagues from WP2 and are listed on deliverable D2.1.

4.2 BioReCer Data models

4.2.1 UNECE Supply Chain Reference Data Model (SCDRM)

This Data Model was proposed on the GA as one of the models that could be taken as the basis for the development of the BioReCer context data models. Nevertheless, once that has been properly analysed it has been identified that it is focused on the payments and regulations related with logistic flows, so it would not engage properly with our necessities regarding traceability, MFA (Material Flow Analysis) and LCA (Life Cycle Assessment).

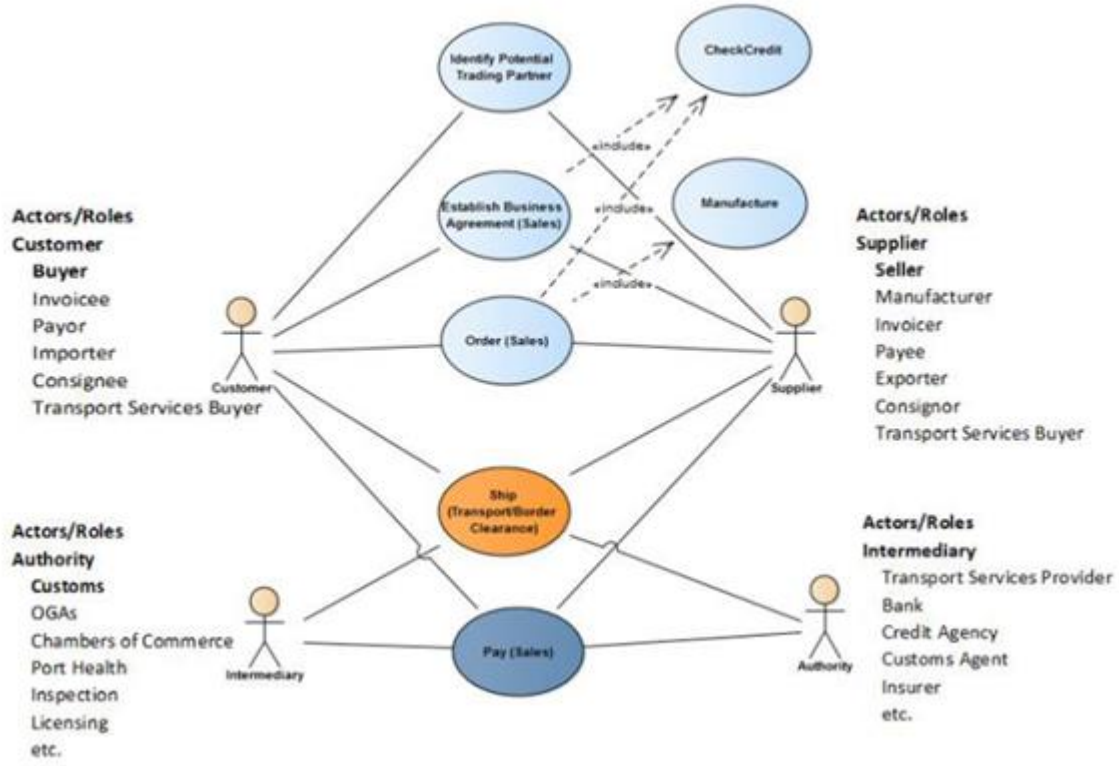


Figure 13. Caption from the SCRDM Business Requirement Specification Document[14]

As stated in the document Supply Chain Reference Data Model BUSINESS REQUIREMENT SPECIFICATION, "The scope of the SCRDM can be expressed as a subset of the scope of the ISCRM covering cataloguing through payment. Included are the delivery of goods and associated import/export regulatory declarations" and ISCRM covers processes in four business areas: Commercial. Logistics, Regulatory and Financial.

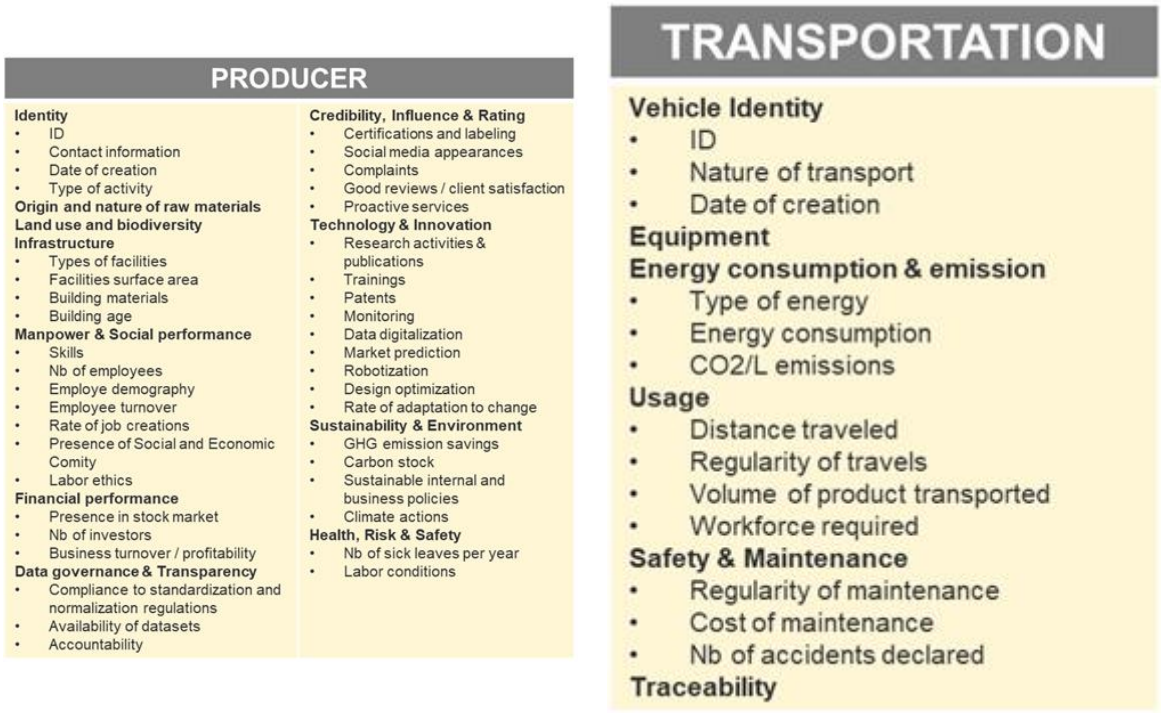
[14] https://unece.org/trade/uncefact/rdm

4.2.2 Identified ontologies and models

Instead of designing data models from zero, WP5 partners are going to use schemes and ontologies that are already published and used in other applications.

Based on the objective of the project, the partners involved on WP5 have searched for what available ontologies can be useful to compose the BioReCer context data models.

First, it has been identified a model which meets the requirements to describe a common production flow chain. This model proposes as entities: producer, transportation, processing unit and product.



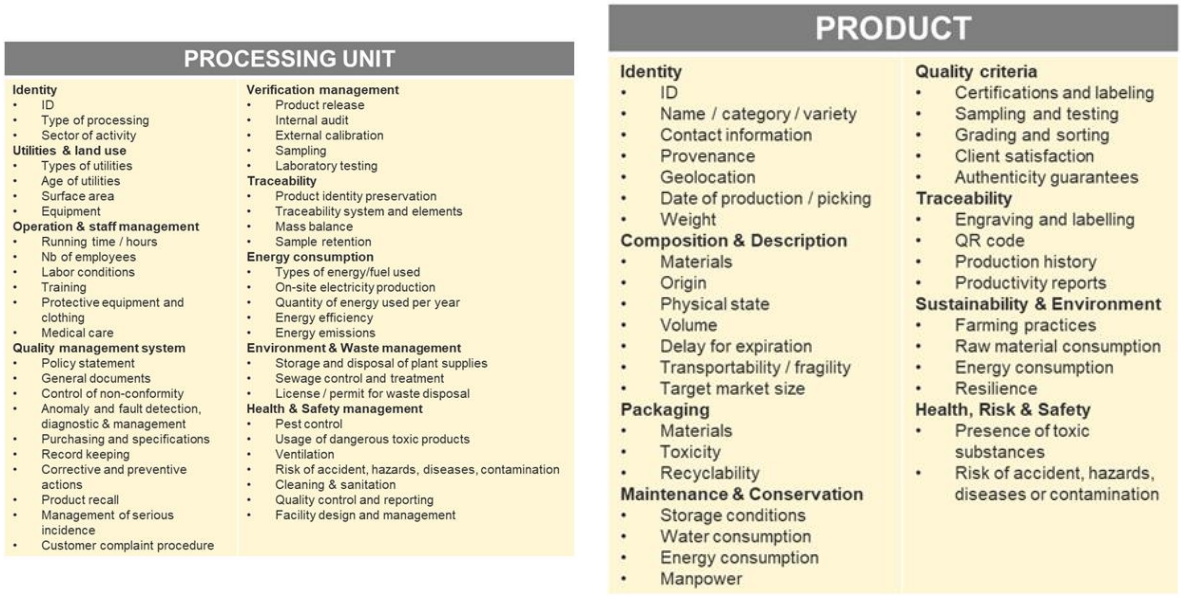


Figure 14. Production context data model proposal

Even though it covers all the necessities of an industrial process, perhaps it is not enough to characterize some of the project necessities, such as MFA, LCA or Certification.

Thus, among the different identified models, it is interesting the publication "LCA-oriented semantic representation for the product life cycle" that provides a structure to describe the complete Life Cycle Assessment of a product.

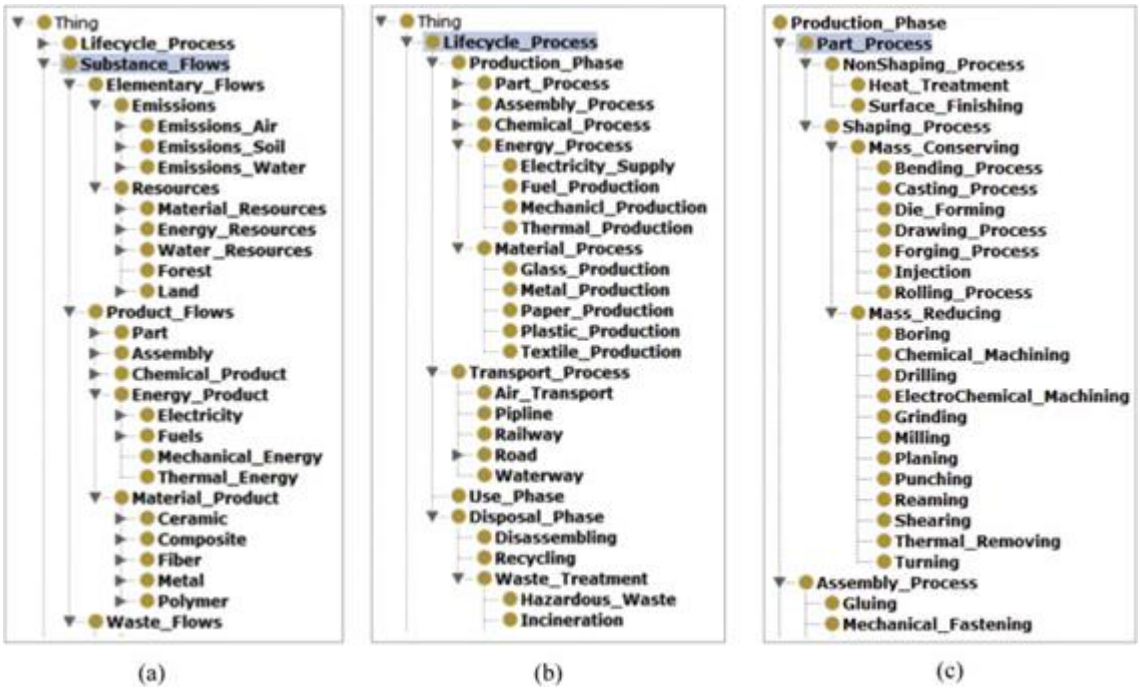
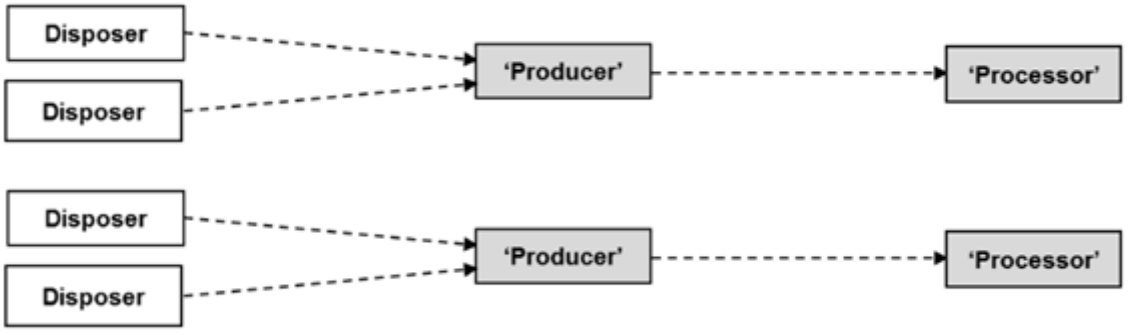


Figure 15. LCA-oriented semantic representation for the product life cycle[15]

Also, Better Biomass certification schema points out that sometimes there is an intermediate actor between the generators of biomass and the ones transforming it, the collector, which accumulates different flows of biomass coming originating on several factories.



[15] Yingzhong Zhang, Xiaofang Luo, Jennifer J. Buis, John W. Sutherland, LCA-oriented semantic representation for the product life cycle, Journal of Cleaner Production, Volume 86, 2015, Pages 146-162

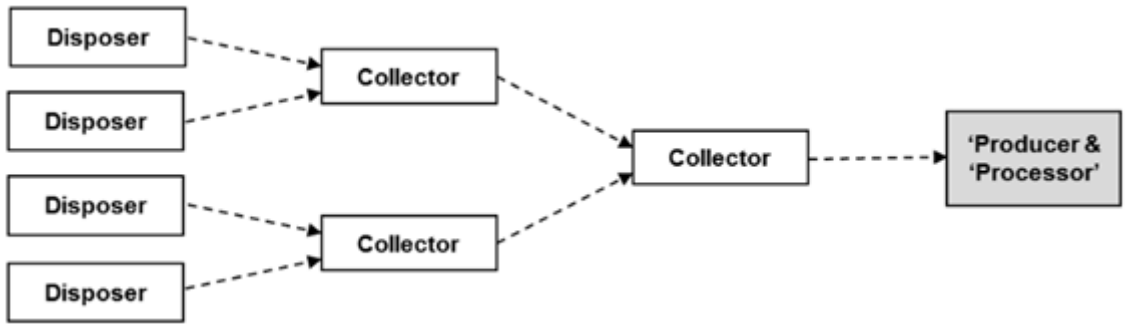


Figure 16. Types of residual flows

Finally, in the repository of Smart Data Models, some models have been identified that would be interesting for the specific case studies related to wastewater treatment. The available agriculture and forestry models seem not to have included any waste and residues related fields.

4.2.3 Info provided from others Work Packages

Currently, the main contribution to this section comes from the WP2 partners as they have defined a preliminary list of circularity indicators for the biological feedstocks of each case study.

As a summary, based on this document, it has been identified two key issues:

- Granularity level: MFA and LCA require data with a higher detail level than in a normal production process. This was already mentioned before.
- Quality parameters: Each product requires the registration of its respective specific/significant parameters that will allow it to evaluate product quality.

Moreover, other complementary parameters (e.g., as GHG emissions and renewable energy contribution) might be considered if they are available. If not, they would be estimated through calculations and bibliography data.

4.2.4 Proposed model and case studies application

With all this available information, WP5 partners have started to design a draft of the BioReCer context data model. Leaded by EGM, it is composed by the following modules:

- Harvesting and manufacturing, related mostly with the industrial processes. It defines the following entities (and its subsequent relations and properties): biomass, transformation, production system, manufacturer, and product.

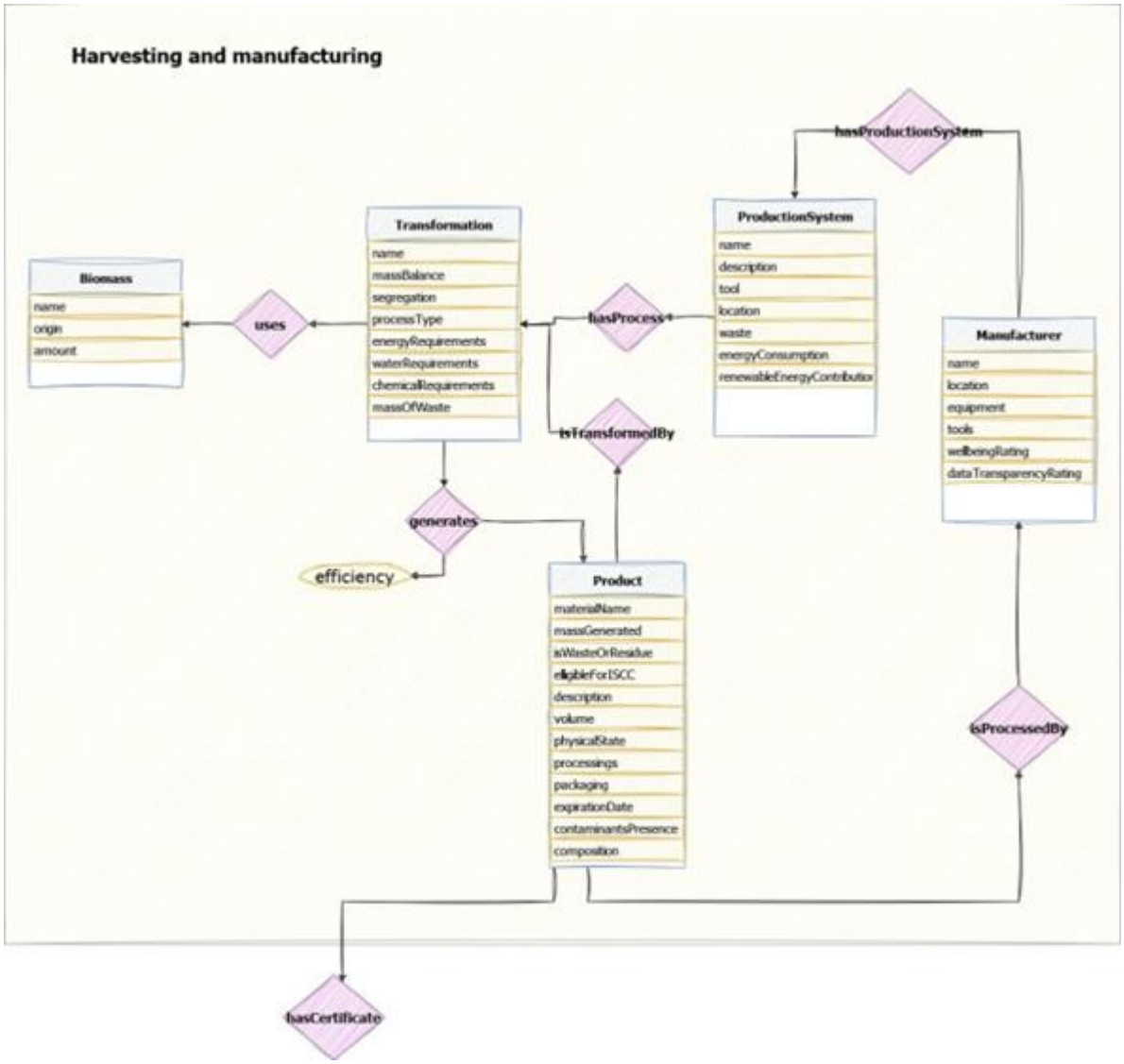


Figure 17. Harvesting and manufacturing entities and their relations and properties (source: EGM)

- Transportation and logistics, composed of storage, distributor, collector and transporters as entities.

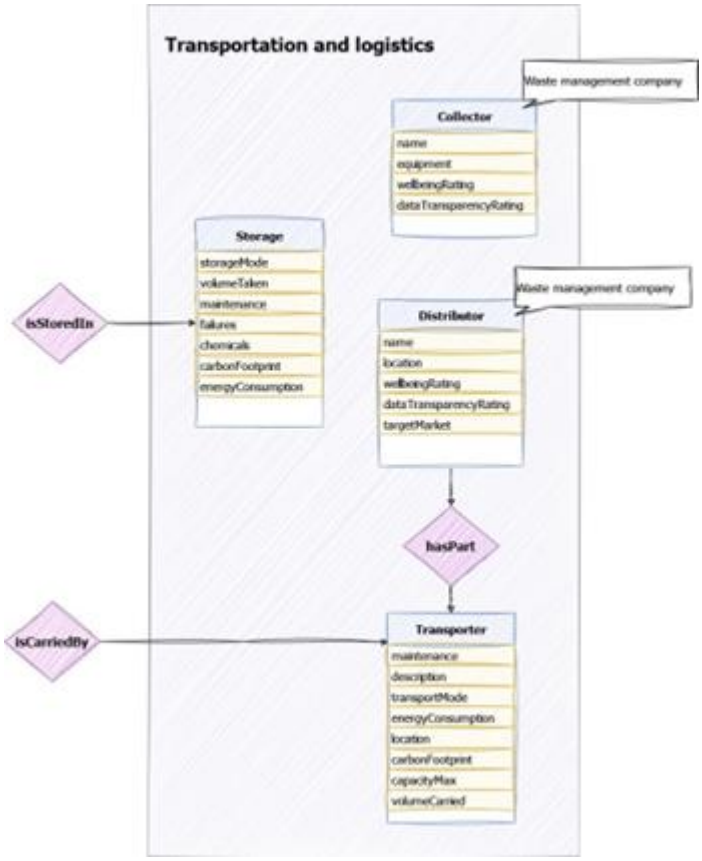


Figure 18. Transportation and logistics entities and their relations and properties (source: EGM)

- User profiles, that identifies the different kinds of users that BIT would have. In this case, the entities considered are consumer, biomass producer, biomass transporter and certifier.

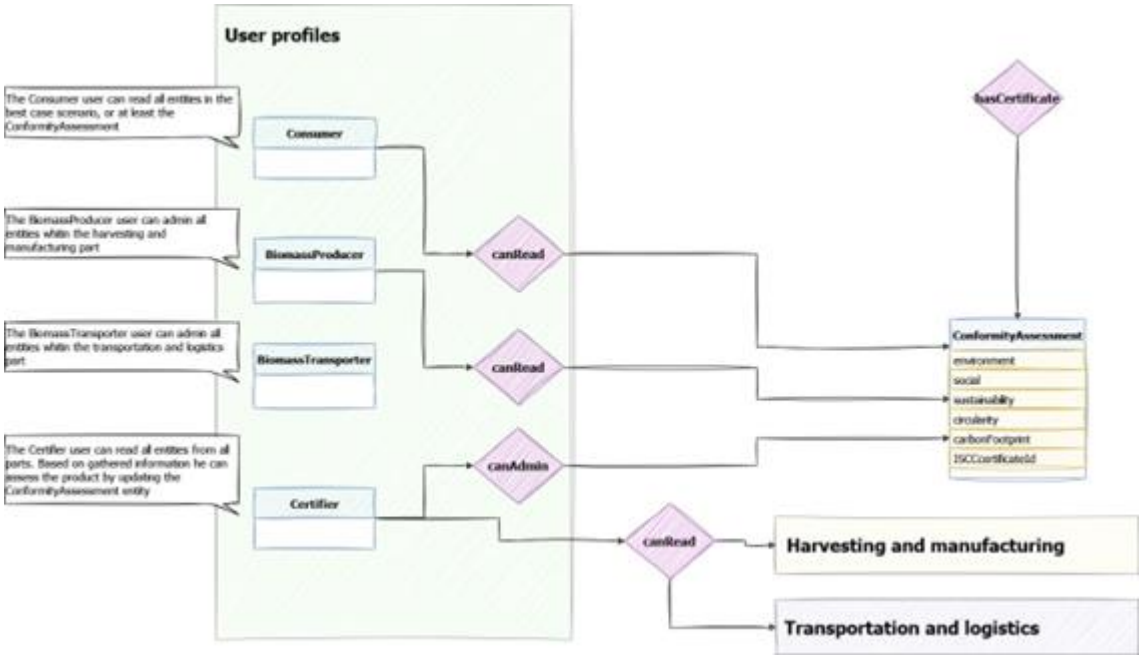


Figure 19. User profiles entities and their relations and properties (source: EGM)

At the current stage of the project, this is a test model that is going to undergo transformations and evolve to engage with all BioReCer project necessities.

ANFACO as WP6 leader (case studies) and participant of WP5 is working on a draft of the application of this model to its case study. Once finished, this draft will act as a guide of how to build the data context model and which data will be needed from each case study of the other WP6 partners.

This partner is in charge of the fishery case study and plans to work in three different flow chains: collagen extraction from fish discards, omega3 extraction from fish discards and proteins and fibre from discarded macroalgae.

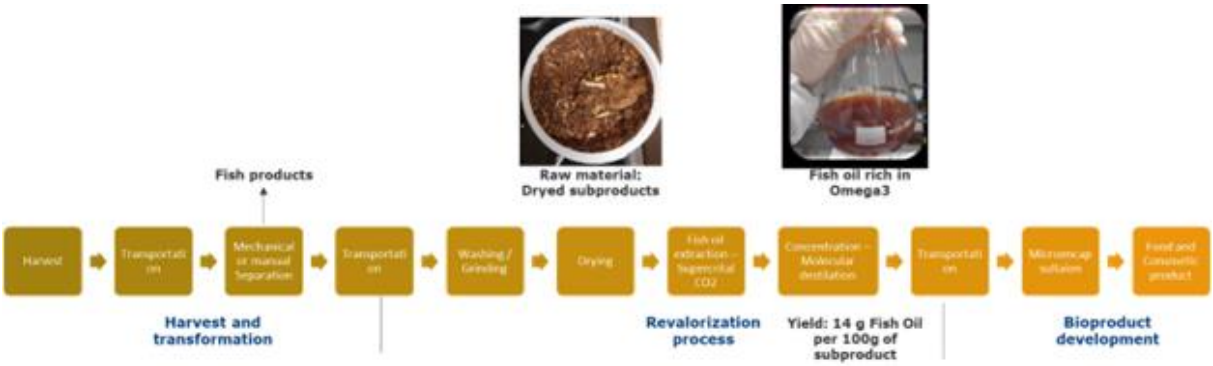


Figure 20. Biomass valorisation flow chain for Case Study 1, from fish discards to omega3 oils (source: ANFACO)

ANFACO has selected the second flow in order to prepare the data model draft. Thus, it has advanced on detailing the different flow subproducts and the subsequent mass balance. Also, it has prepared a table with mock-up data to simulate the different entities that would compose the chain.

Harvester		Transporter		Biomass producer		Valorizer		Bioproduct developer		Product	
ID	[redacted]	Distance	TBD	ID	[redacted]	ID	[redacted]	ID	[redacted]	Type of product	Cream - Cosmetics
Production	300tone/d	Volume of biomass	TBD	Type of processing	Cutting	Type of processing	Grinding	Type of processing	Degumming	Composition	80% fish oil concentrate
Certifications	TBD	Type of vehicule	TBD		Cooking		Drying		Neutralization	Certification	TBD
GHG emmis./ tonne	TBD	Fuel	TBD		Cleaning		Supercritic CO2		Deodorization	Labeling	TBD
Energy/tonne	TBD	Collection efficiency	TBD	Equipment	TBD	Equipment	TBD		Concentration	Packaging	TBD
Traceability	TBD	Nº travels	TBD	Energy/tonne	TBD	Energy/tonne	TBD		Microencapsulation	Expiration	TBD
				GHG emmis./ tonn	TBD	GHG emmis./ tonn	TBD	Equipment	TBD	Traceability	TBD
				Traceability	TBD	Traceability	TBD	Energy/tonne	TBD	Other components	TBD
				Certification	TBD	Certification	TBD	GHG emmis./ ton	TBD		
				Quality control	TBD	Quality control	TBD	Composition	TBD		
				Cleaning products	TBD	Chemicals used	TBD	Certification	TBD		
				Water use	TBD	Water use	TBD	Labeling	TBD		
				Mass balance	Heads -12%	Extraction Efficien	Mean: 81%	Traceability	TBD		
					Tails - 2%	Quality grade	Omega3: 20%	Efficiency	72,10%		
					Viscera - 3.3%		PUFA: 23%	Water use	TBD		
					ROE - 1.7%		EPA+DHA: 17%	Chemicals used	TBD		
					Skin - 9%		Acidity: 8%				
					Bones - 10%						
					Red Muscle - 8%						
					Tuna product - 56%						
				Efficiency	95%						

Figure 21. Mass Balance and draft characterization of entities composing the flow chain (source: ANFACO)

Currently, with all this information the diagram flow of this fishery case study is being built. This will help to identify problems and inefficiencies when it comes to implement the data context model on the real case and test the model performance.

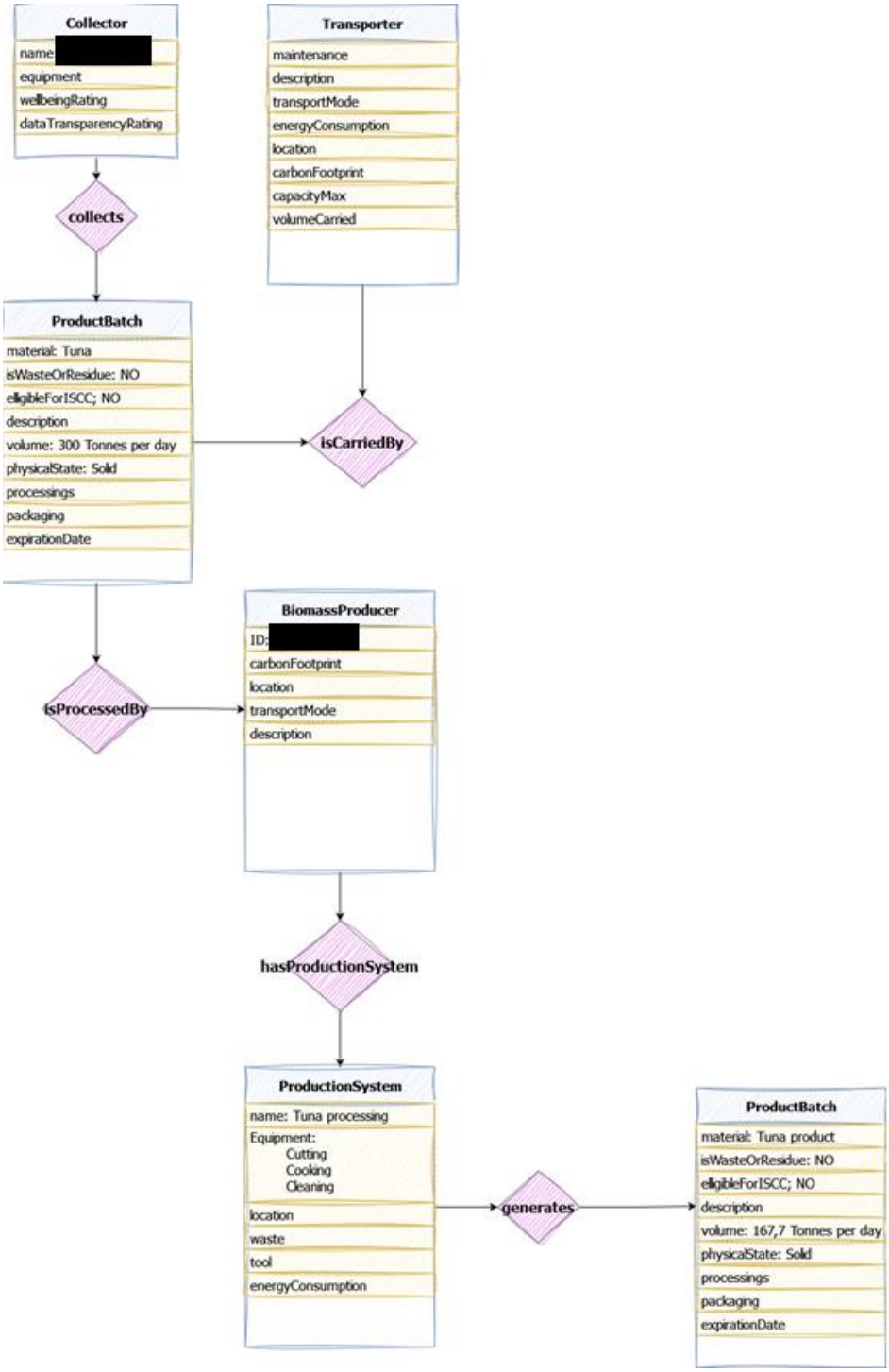


Figure 22. Data model flow diagram draft for CS1 flow chain (source: ANFACO)

Partners are working with mock-up data, as contacts with stakeholders are still being established. Also, this first approach to the actual data model is interesting as it allows us to check how it adjusts to different chain topologies: one company performing all the

valorisation process in one place, one company working between different locations, different companies representing each stage of the flow chain, etc.

5 Components description

5.1 Trust and security

The trust and security features of the BIT platform will be provided by the open source Keycloak[16] solution, developed and maintained by RedHat. It is a mature service, very feature rich, and supported by a very active community.

Keycloak implements full support for OpenID Connect and OAuth2 standards. It offers a complete administration interface allowing it to be configured very finely and to be easily administered on a daily basis, and in particular:

- Graphic customization of all pages presented to users.
- Configuration of all security rules: expiration of inactive sessions, validity period of OAuth2 tokens, verification of emails when creating accounts, etc.
- Activation of additional security functions, such as two-factor authentication (2FA) or one-time password (OTP).
- Definition of rules for creating user accounts, managing passwords (minimum length, etc.), etc.
- Organisation of users into groups.
- Traceability of access to applications, active sessions, etc.
- Manual revocation of tokens in the event of an identified risk (theft or loss of equipment, etc.).
- Possibility for users to modify their personal data.

All the components of the platform will be integrated with this service and use the credentials it provides. Then, depending on the service provided by each module and the associated constraints and specificities, each module can offer advanced rights management features.

It is also within this service that the user groups are defined, as well as the repository of roles known by each module. It is then up to each module to apply its own security policy according to the roles assigned to the user currently connected.

As introduced in the section 3, it is planned to integrate the secure and privacy preserving features backed by the Verifiable Credentials Data Model specification. To support an implementation of such a feature, some extensions to the OpenID Connect specification have been identified and introduced in section 3. There have been some preliminary discussions in the Keycloak community to add an official support for these OpenID Connect extensions but things have not progressed in the last months. Fortunately, the FIWARE

[16] https://www.keycloak.org/

community has started to work on the development of a Keycloak plugin[17] aimed at adding support for Self-Issued OpenID Provider v2 / OpenID for Verifiable Presentations clients and for the issuance of Verifiable Credentials through the OpenID for Verifiable Credentials Issuance specification. As part of the deployment of the BIT platform, it is planned to use and contribute to the further development of this plugin.

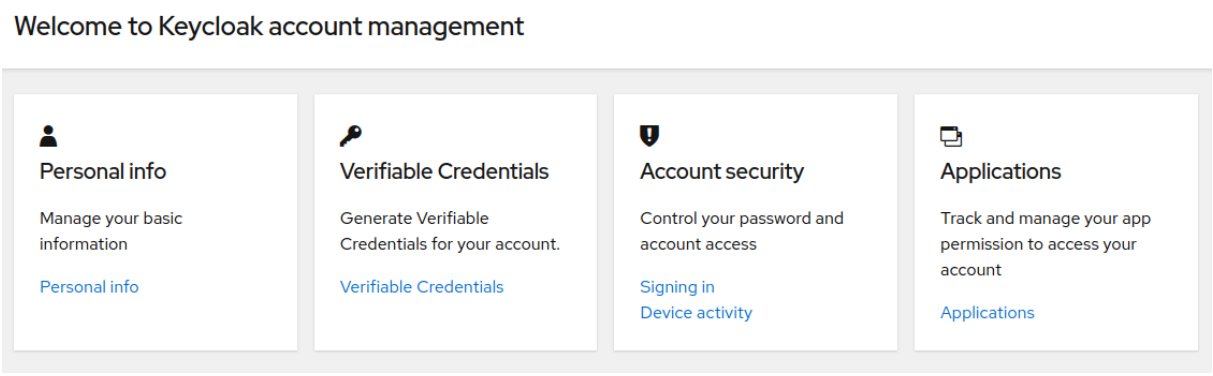


Figure 23. Keycloak user account enriched with the ability to generate verifiable credentials

[17] https://github.com/FIWARE/keycloak-vc-issuer

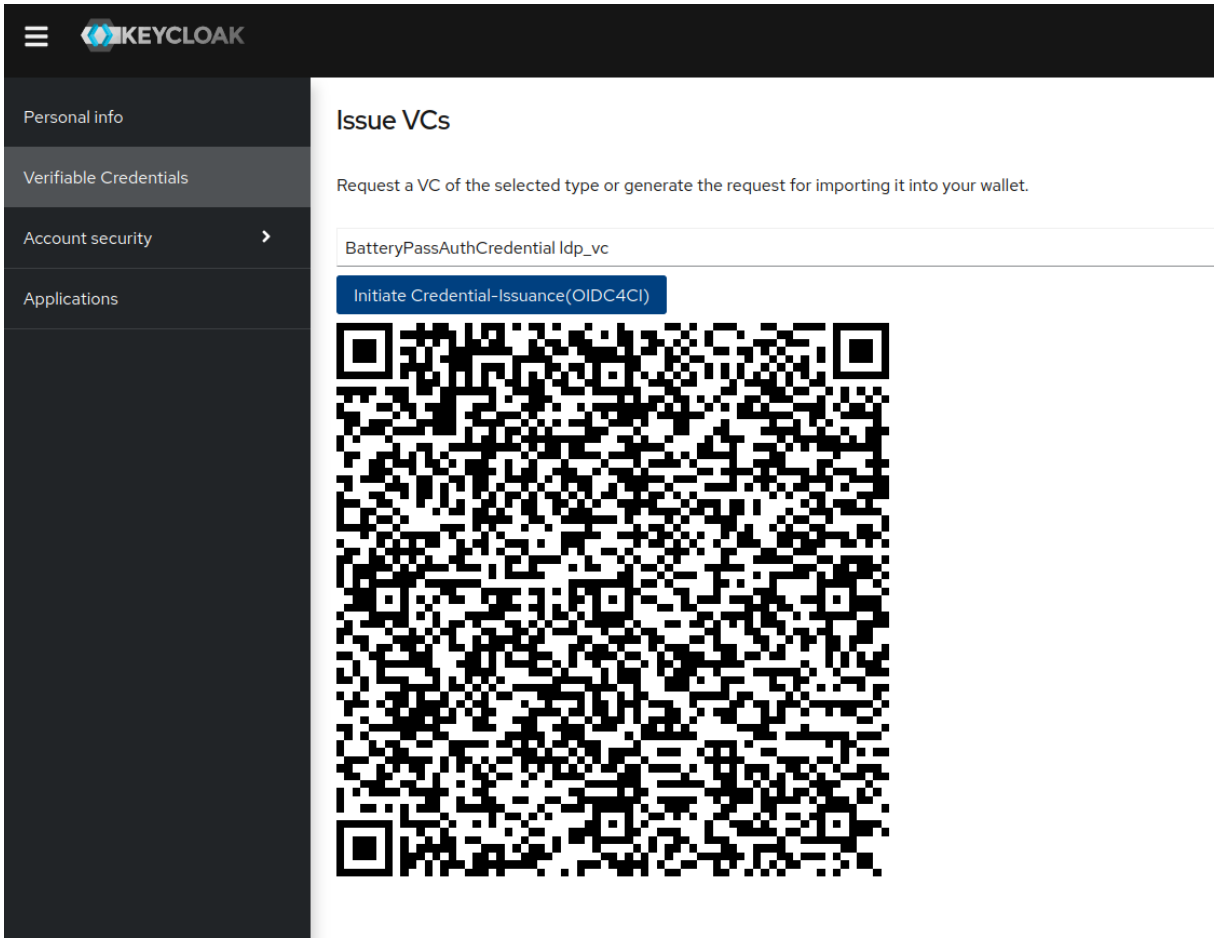


Figure 24. Generation of a QR code to request for a verifiable credential

5.2 Data Ingestion

This component is essential for transforming the incoming data to a NGSI-LD based model. The data ingestion step is characterised by the variety of data and the different mechanisms for collecting them.

At the south bound of the BIT platform, data is received from a large variety of data sources (sensors, external APIs, Open Data portals, …). In order to be able to later process it in an unified and standardised way, it needs to be first validated, optionally pre-processed, then transformed into an NGSI-LD compliant payload and following a defined data model.

To perform this task, the platform will integrate the NiFi tool from the Apache foundation. Designed with resiliency, scalability and security in mind, it is able to handle very large volumes of data, to distribute the processing of data flows among a cluster of instances, and to take care of back-pressure to ensure all the data flows are only receiving what they can deal with.

NiFi also natively provides a rich library of components aimed at dealing with the processing of data. Ranging from connectors to retrieve data from external sources (HTTP, MQTT, FTP, flat files, …) to connectors to filter, validate, identify, enrich and transform data, it offers all the components that will allow the platform to quickly integrate any new data source.

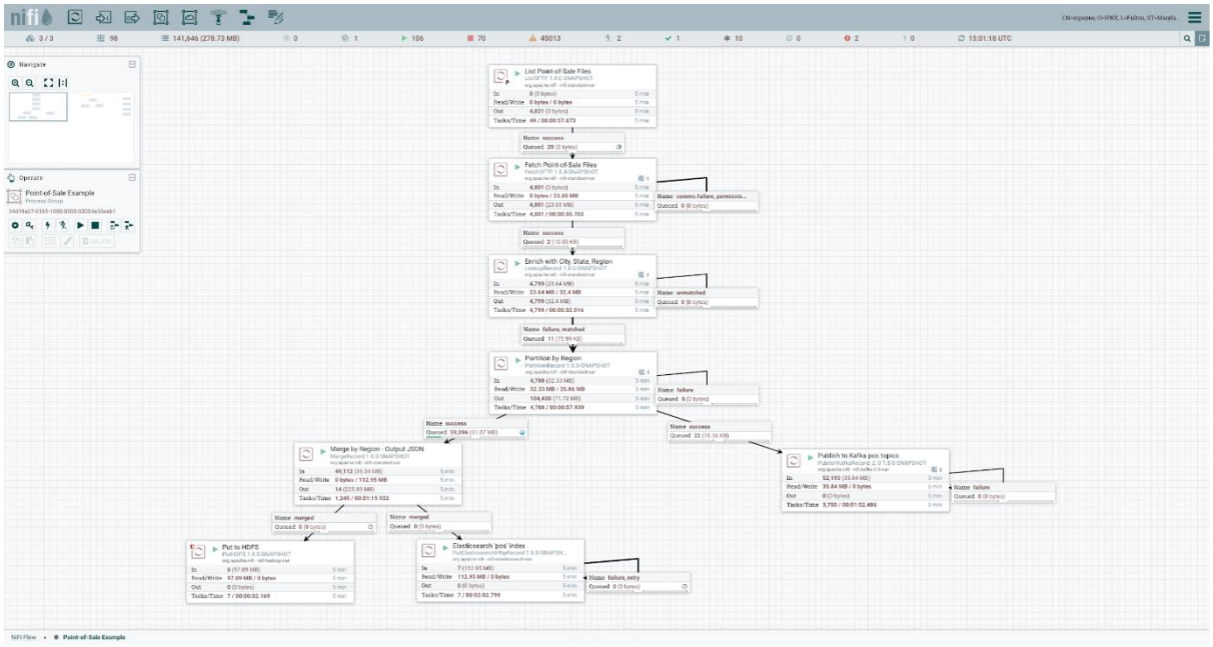


Figure 25. Example of a full data ingestion flow in Apache NiFi

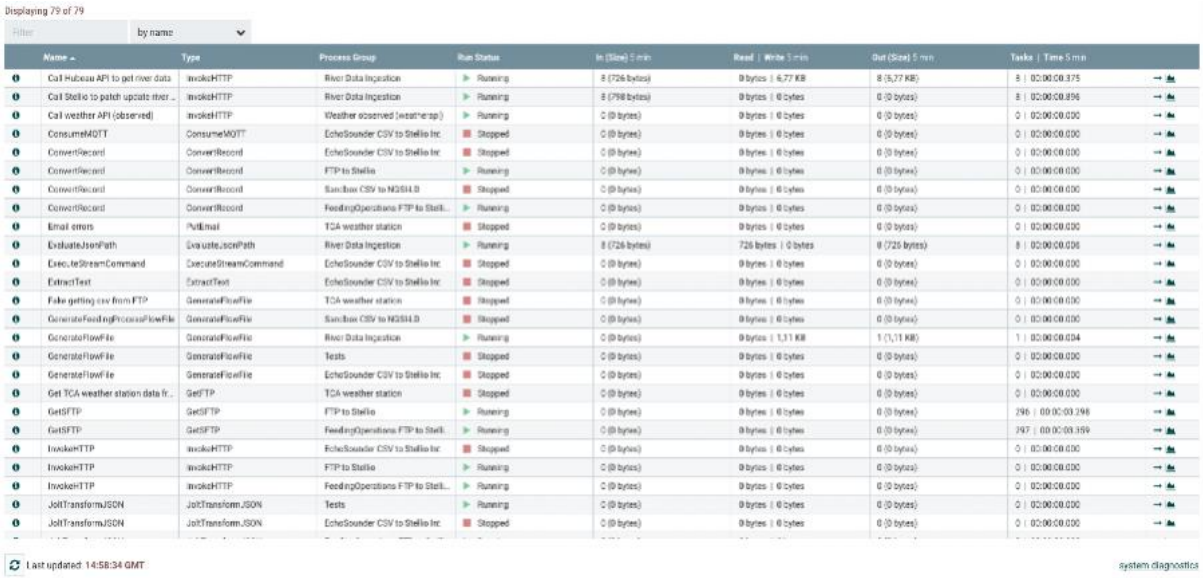
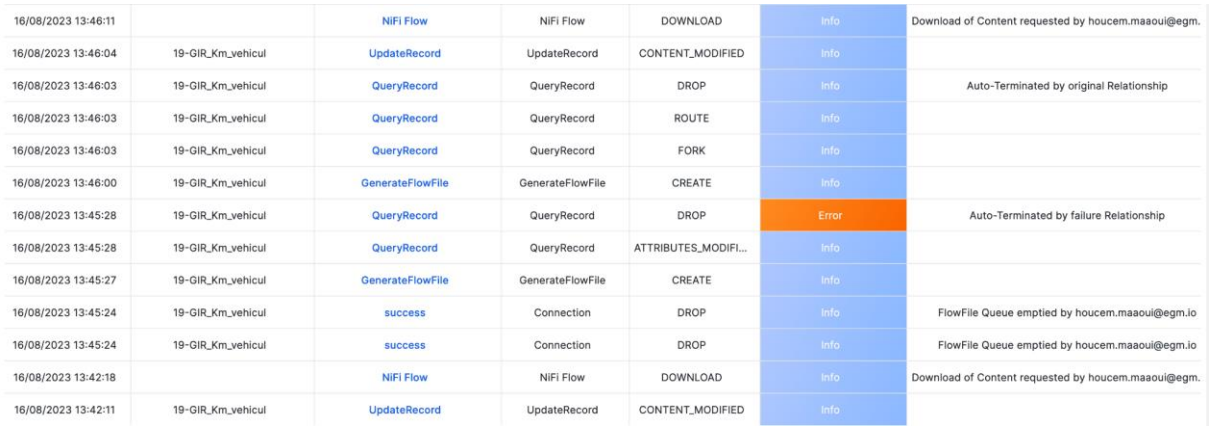


Figure 26. Global view of all flows deployed in a NiFi instance



16/08/2023 13:46:11		NiFi Flow	NiFi Flow	DOWNLOAD	Info	Download of Content requested by houcem.maaoui@egm.
16/08/2023 13:46:04	19-GIR_Km_vehicul	UpdateRecord	UpdateRecord	CONTENT_MODIFIED	Info	
16/08/2023 13:46:03	19-GIR_Km_vehicul	QueryRecord	QueryRecord	DROP	Info	Auto-Terminated by original Relationship
16/08/2023 13:46:03	19-GIR_Km_vehicul	QueryRecord	QueryRecord	ROUTE	Info	
16/08/2023 13:46:03	19-GIR_Km_vehicul	QueryRecord	QueryRecord	FORK	Info	
16/08/2023 13:46:00	19-GIR_Km_vehicul	GenerateFlowFile	GenerateFlowFile	CREATE	Info	
16/08/2023 13:45:28	19-GIR_Km_vehicul	QueryRecord	QueryRecord	DROP	Error	Auto-Terminated by failure Relationship
16/08/2023 13:45:28	19-GIR_Km_vehicul	QueryRecord	QueryRecord	ATTRIBUTES_MODIFI...	Info	
16/08/2023 13:45:27	19-GIR_Km_vehicul	GenerateFlowFile	GenerateFlowFile	CREATE	Info	
16/08/2023 13:45:24	19-GIR_Km_vehicul	success	Connection	DROP	Info	FlowFile Queue emptied by houcem.maaoui@egm.io
16/08/2023 13:45:24	19-GIR_Km_vehicul	success	Connection	DROP	Info	FlowFile Queue emptied by houcem.maaoui@egm.io
16/08/2023 13:42:18		NiFi Flow	NiFi Flow	DOWNLOAD	Info	Download of Content requested by houcem.maaoui@egm.
16/08/2023 13:42:11	19-GIR_Km_vehicul	UpdateRecord	UpdateRecord	CONTENT_MODIFIED	Info	

Figure 27. Example of the monitoring view of data flowing into the data ingestion module

Finally, it is designed around the no-code concept and thus provides an UI where all the data flows can be designed and monitored. This will greatly help in the adoption of this component by all the technical users of the platform.

5.3 Data storage

Stellio is a context broker, open-source implementation of the NGSI-LD specification, released last year from the FIWARE Foundation incubator and now validated as a mature component.

In accordance with the NGSI-LD standard, it implements and exposes the three main families of APIs: the management of the information context, the temporal and geospatial search and the management of subscriptions and notifications.

Beyond the implementation of the NGSI-LD standard, it is fully part of a "FIWARE architecture", both through the management of structural and contextual data according to the common NGSI-LD data model, and through integration with the other bricks of the FIWARE ecosystem (IoT agents, data collection, etc.) made possible thanks to compliance with the NGSI-LD standard, guaranteeing the internal and external interoperability of the platform.

The principles implemented in the Stellio context broker are those of a modern, responsive and scalable architecture:

- Business microservices, divided according to the main NGSI-LD API families, subscribing to their topics of interest, and using databases adapted to their business.
- A pub-sub-type central message bus (Kafka) ensures the exchange of messages and events between platform components in a decoupled, extensible, scalable, and reactive manner.

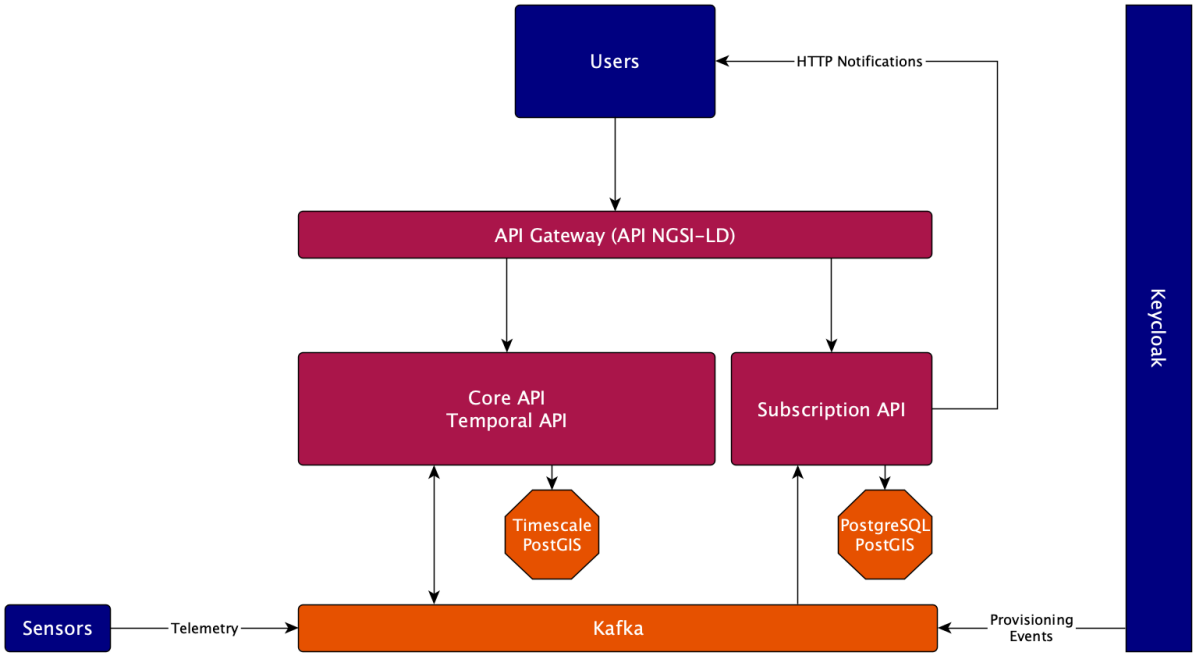


Figure 28. High level overview of the Stellio architecture

The data produced by the various external components integrated into the platform (sensors, external data flow) are stored by the context broker in a PostgreSQL database with 2 specific extensions:

- Storing the temporal evolution of the data is done in the TimescaleDB database. Designed as an extension to PostgreSQL, it is specialised in the storage of temporal data ("timeseries") and therefore perfectly suited, among other things, to the storage of sensor measurements, but more generally to any data that changes over time, for example, the location of a by-product being tracked, on which we can perform carbon footprint analysis at a specific point in time.
- Second, geo-information is managed with PostGIS (another extension to PostgreSQL) to allow spatial queries on the data stored in TimescaleDB (whether it is temporal or not).

Note that the TimescaleDB database, designed for the storage of time series and very large volumes of data, allows for the storage of petabytes of data distributed on several nodes while having one of the best levels of performance existing as of today for this type of database. It will therefore be able to safely support all the new storage needs that will be added during the project.

All the data stored in the different databases of the context broker is fully accessible via the NGSI-LD API.

5.4 Data processing

5.4.1 Decision support system development

The functioning of the proposed system is carried out in several stages:

- data import;
- formation of meta-information from the downloaded data;
- Data Mining tasks type and methods selection;
- specification of criteria and metrics for ML model quality assessing;
- obtained ML models creation and evaluation;
- ML models obtained results visualization and the issuance of a methods ranked list which provides the highest quality solution to the problem.

During the DSS creation several algorithm variants will be considered for ranking the models, considering the possibilities of their training and the implementation complexity, due to the need in each sign of the dataset and metadata. Mainly Gaussian processes for decision making will be used.

Decision-making is the process of making choices by identifying a problem. The process demonstrates the importance of troves of data collected and its analysis in having an ideal decision made for any business segment.

By gathering valuable data from various touchpoints and categorising options, a step-by-step decision-making process can assist in making more careful, considered actionable insights. Today for industries to grow, the most sought-after suggestive way in any business is to study your data and make accurate decisions.

5.4.2 The key steps in decision making

- Identify the problem
- Gather and analyze the most relevant data
- Discover alternatives to solve the problem
- Select the best alternative
- Convert the decision into action
- Evaluate and verify the decision

Data acts as the key to unlocking numerous avenues of opportunities, it helps to solve problems and make strategic decisions. Data is huge and ever-growing and also termed Big Data.

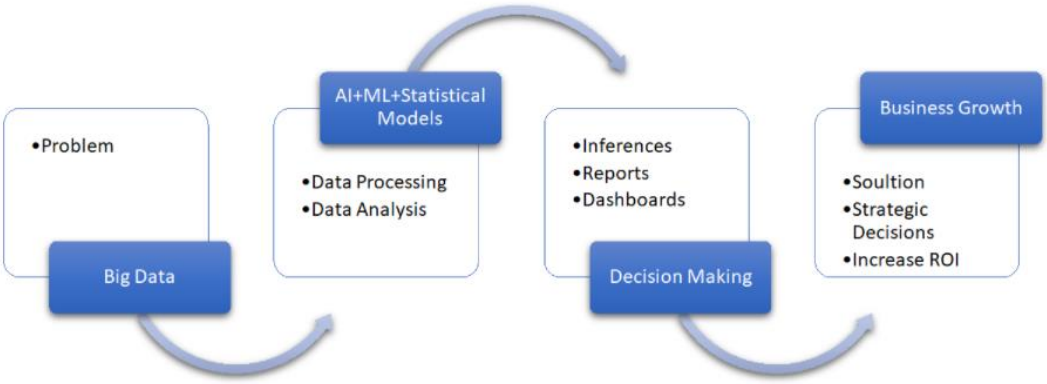


Figure 29. Key steps in decision making

Human decision-makers can make mistakes when faced with large-scale or complex decision-making problems. Manual data processes and report generation leads to errors increasing turn-around time. This happens due to the intrinsic limitations of their memory, attention, and limited knowledge.

Hence the role for data processing comes into the picture, the company needs to spend less time accumulating massive volumes of data and more time leveraging technologies. To detect and mitigate risks, as well as proactively uncover opportunities the use of AI, ML algorithms along statistical modelling tools help decision-makers overcome these limitations.

Statistics is a collection of tools that serve as the basis for the analysis and processing of data to transform raw observations which you can understand and is informative. Many performance indicators are used in the ML algorithm, such as precision, accuracy, recall, f-score, and root means square error, which is based on statistics. These indicators help in understanding the visual representation of the data and the performance algorithms used in it. Statistics helps to identify trivial patterns with perhaps outliers in the data metrics such as median, mean, and standard deviation of complex data sets.

Statistics is a significant source of an evident tool as it offers us clear objectives on numerical data on crucial areas of life such as business performance, population, growth and characteristics, economic performance, health and welfare, and the state of our environment.

In recent years, the growth of decision-making applications, where principled handling of uncertainty is of key concern, has led to increased interest in Bayesian techniques. By

offering the capacity to assess and propagate uncertainty in a principled manner, Gaussian processes (GPs) have become a key technique in areas such as Bayesian optimization, active learning, and probabilistic modeling of dynamical systems. In parallel, the need for uncertainty-aware modeling of quantities that vary over space and time has led to large-scale deployment of Gaussian processes.

Gaussian Processes are popular surrogate models for sequential decision-making tasks such as Bayesian Optimization and Active Learning. Such frameworks often exploit well-known cheap methods for conditioning a GP posterior on new data.

In typical machine learning, we specify a function with some free parameters (such as a neural network and its weights), and we focus on estimating those parameters, which may not be interpretable. With a Gaussian process, we instead reason about distributions over functions directly, which enables us to reason about the high-level properties of the solutions. These properties are controlled by a covariance function (kernel), which often has a few highly interpretable hyperparameters. These hyperparameters include the length-scale, which controls how rapidly (how wiggily) the functions are. Another hyperparameter is the amplitude, which controls the vertical scale over which our functions are varying. Representing many different functions that can fit the data, and combining them all together into a predictive distribution, is a distinctive feature of Bayesian methods. Because there is a greater amount of variability between possible solutions far away from the data, our uncertainty intuitively grows as we move from the data.

A Gaussian process represents a distribution over functions by specifying a multivariate normal (Gaussian) distribution over all possible function values. It is possible to easily manipulate Gaussian distributions to find the distribution of one function value based on the values of any set of other values. In other words, if we observe a set of points, then we can condition on these points and infer a distribution over what the value of the function might look like at any other input. How we model the correlations between these points is determined by the covariance function and is what defines the generalization properties of the Gaussian process. While it takes time to get used to Gaussian processes, they are easy to work with, have many applications, and help us understand and develop other model classes, like neural networks.

5.5 Visualisation and monitoring

We will explore two data visualisation platforms that will also serve for the monitoring basis: Superset and Grafana.

Superset[18] is a new-generation modern data exploratory analysis and visualisation platform, initially developed by AirBnB and now hosted by Apache Foundation under the Apache 2.0 licence. This project is sustained and regularly updated by an active community of over 500 contributors worldwide, making it a mature, stable, and scalable solution, perfectly adapted to any use type.

Its main functionalities for visualisation are:

- A complete interface and creatable interactive dashboards,
- A highly developed catalogue of designs and types of visualisation possibilities, such as line charts, bar graphs, numerical values, tabular data, geographical mappings, heatmaps, etc,
- An intuitive and accessible interface to all allowing the creation of new finely tunable visualisations,
- A highly performant SQL development environment including a rich metadata browser, integrating SQL Lab an advanced interface for data exploration, selection, and processing, to construct visualisations on the fly. SQL Lab includes the following capabilities:
    - A multi-tab environment to process multiple SQL queries simultaneously,
    - A smooth flow to view query results,
    - Ability to easily browse database metadata, i.e., tables, columns, indexes, and partitions,
    - A search engine to find queries executed in the past,
    - Support for creating models using the Jinja Modelling Language.
- A lightweight semantic layer for data analysts to rapidly define and customise dimensions and measures,
- A caching system and asynchronous requests, providing very high performance even for voluminous data,
- An extensible security model allowing very complex rules configuration on access to product features and datasets,
- Integration with standard authentication backends (including EGM's, based on OpenID Connect / OAuth2 standards),
- The possibility to develop and personalise visualisation plugins,
- An API for programmatic access to the main functionalities of the platform,
- A native cloud architecture,
- Reporting of a dashboard or specific figures/tables, as customizable time intervals via email groups or instant messaging services such as Slack.

[18] https://superset.apache.org/

In comparison, although **Grafana**[19] contains similar functionalities as Superset but with different design properties, it offers above all an automatic alert system, based on a value threshold with logging possibilities, with messaging via a wide range of possible channels. **Monitoring** will be performed with the coordination of such dashboards, to assess the real-time quality, frequency and distribution of the collected data, as well as their temporal evolution. Monitoring serves two purposes: warning and alerting. Warning signs are a tendency for data to move away from normal or expected conditions and engage the user to perform preventive actions to correct this drift and prevent an unwanted event. Alerts are instantaneous messages that inform the user that a threshold has been reached or that unwanted or unexpected data has been detected. We therefore monitor Key Performance Indicators (KPIs) or other user-defined indicators such as rates of change values, predictions vs reality, or simply raising a flag whenever an unwanted value appears (e.g., negatives, NaN's, etc.). KPIs for this project will be defined in WPs 2 and 3.

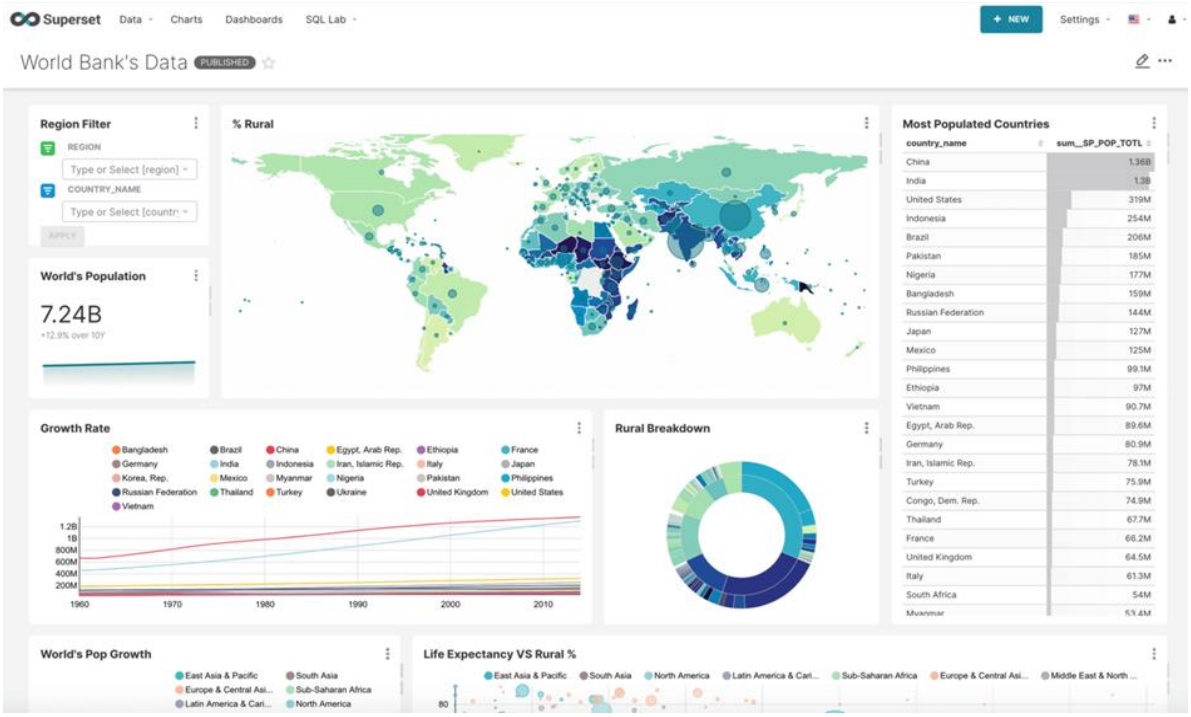


Figure 30. Example of an interactive dashboard welcome page, showing a diversity of data visualisation graphics, tables and statistics

[19] https://grafana.com/

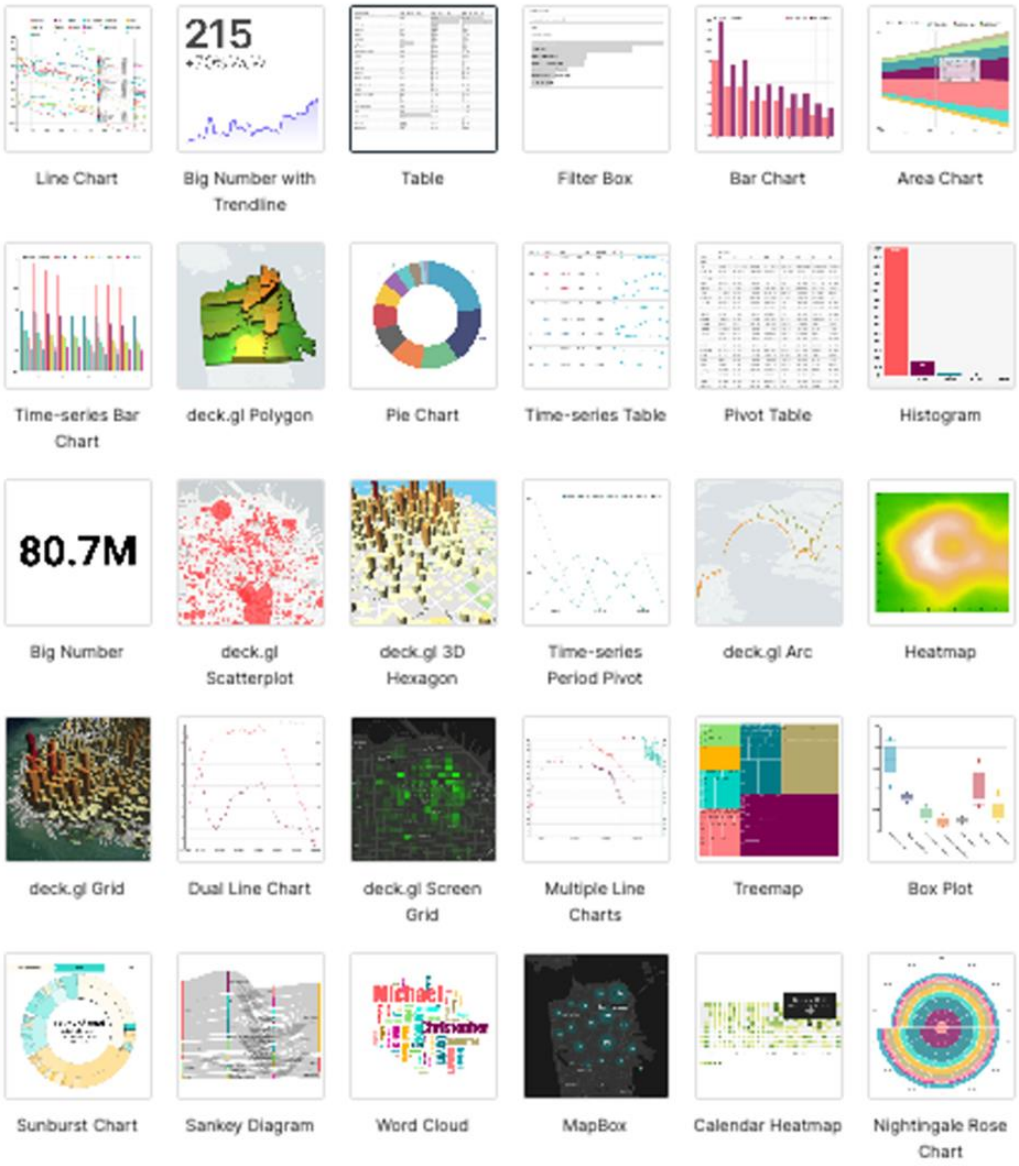


Figure 31. The non-exhaustive catalogue of the possible different data visualisation designs and types, including single- and multi-parameter visualisations

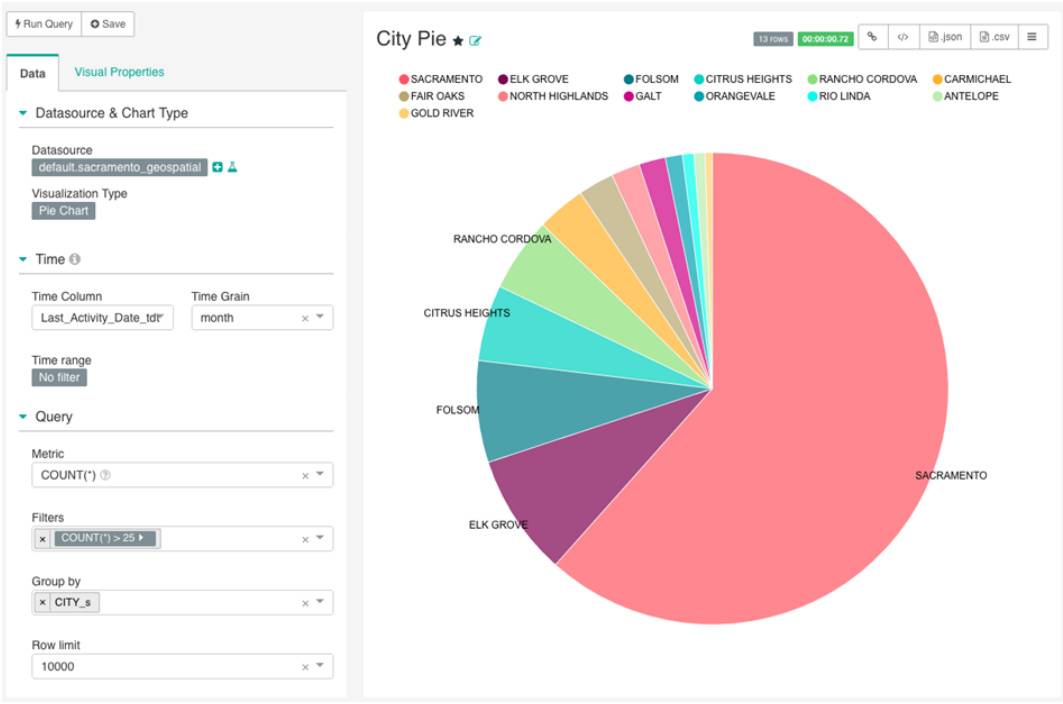


Figure 32. An example of an intuitive and accessible interface to create or modify visualisations (selected data, type of graph, colour, font, labels, etc.)

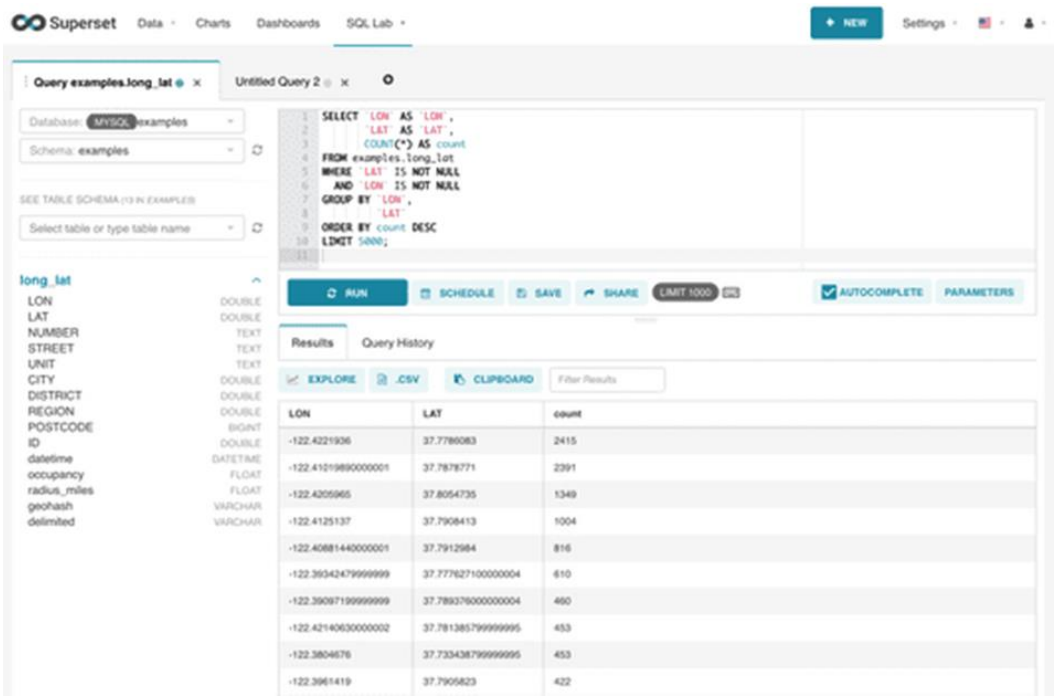


Figure 33. SQL Lab, a highly performant SQL development environment including a rich metadata browser, with SQL query result validation

5.6 BioReCer Web Applications

5.6.1 The BIT User Interface

The BioReCer ICT Tool user interface plays a pivotal role in the project, linking diverse actors in the bio-based industry, from biomass suppliers and technology providers to traders and consultants. The application serves as a key tool in achieving the project's goals of enhancing the environmental performance, traceability, and social acceptance of bioproducts. As a multifunctional web-based Decision Support System (DSS), the application brings together multiple stakeholders and public bodies, allowing them to explore multiple sustainability assessment alternatives for biological feedstocks. It equips these stakeholders with the necessary tools to fit the traceability requirements of certification schemes.

The application creates added value, converting insights and the Environmental and Circularity Assessment Framework into an accessible, easy-to-use format. The application's user-friendly interface fosters an inviting, engaging, and interactive environment for all users, providing a seamless and efficient experience.

Central to the application is the ability to list a vast array of bio-based goods. Through the product listing process, businesses can present their products, complete with descriptions and images.

The application further offers a platform for peer-to-peer interaction, discussion, and collaboration through the BioResources Stakeholders Platform (BRSP). This physical interaction level complements the digital interaction facilitated by the application itself, amplifying the scope of the project. It's also worth noting that the application takes into account the need for the security and protection of user data. All the data is stored in cloud-based servers, which adhere to industry-standard protection measures, ensuring that user data is secure and privacy is maintained.

A key feature of the BIT User Interface (UI) is the Track and Traceability capability. This functionality ensures the integrity and reliability of the bio-based supply chain, providing the ability to track the journey of goods, services, and information at each step. It promotes transparency by enabling stakeholders to monitor and validate the environmental performance of biological feedstocks throughout their lifecycle. From sourcing raw materials to the final end product, every transaction and movement is logged and monitored. The Track and Traceability feature enhances the marketplace's integrity by identifying any possible discrepancies, preventing fraud, and ensuring product safety and quality. It also facilitates compliance with certification schemes that require traceability information, such as sustainability, origin, and ecological footprint. Furthermore, this capability enables stakeholders to optimise their operations based on real-time and

historical data, making the bio-based supply chain more efficient, resilient, and sustainable.

Thanks to Stellio context broker and its exposed API (Application Programming Interface), the BIT web application can consume standardized data, populated by data producers within the BioReCer ecosystem. This approach facilitates seamless communication and data exchange in a secured environment, since all data producers and consumers components are following OAuth2 protocols. The primary objective is to enquire and obtain relevant data that contribute to our wider scientific objectives. Following this, the gathered data will undergo rigorous analysis and the measured criteria will be posted accordingly. This process not only allows for accuracy and efficiency but also ensures up-to-date and reliable data, creating an effective and integrated partnership ecosystem.

In the long run, the BIT UI is positioned to be connected with a global marketplace for bio-based products, expanding beyond its current capabilities and features. Named Circ4bio, this marketplace is being recognized as the application of choice by other Horizon programmes and is projected to undergo continuous development and enhancement.

5.6.2 Marketplace

The marketplace in the Circ4bio application is the nexus of all trading activities. It is an e-commerce platform that facilitates consumer-to-consumer and business-to-consumer transactions in bio-based goods and support services. The marketplace boasts a user-friendly interface, enabling easy navigation, efficient filtering of information, and convenient trade transactions.

The marketplace categorizes users into six main types: plant owners, biomass suppliers, technology providers, buyers/merchants/traders, experts/consultants and transport/storage. This categorization allows for efficient sorting and identification of users and their respective roles in the bio-based industry value chain.

One of the major highlights of the marketplace is its capability to host a broad array of bio-based products. Users can list their products complete with all relevant details, images, and even videos. The product listing process is flexible and user-friendly, with the option for bulk uploads using a CSV file. This allows suppliers to conveniently add multiple products to their listings. Moreover, the marketplace is equipped with a comprehensive search function. Users can search for products using keywords, suppliers' names, or categories. This function enables users to quickly and efficiently find the products or services they need.

The marketplace is designed to cater to all stakeholders in the bio-based industry, irrespective of their role or size. It offers opportunities for small and medium-sized

enterprises (SMEs) as well as large corporations, facilitating their access to wider markets and contributing to the bio-based economy's growth.

This platform ensures transparency in transactions. Each product listing includes complete details about the product, its supplier, and its sustainability, circularity and carbon footprint assessments. Users can review this information before making purchasing decisions, contributing to the overall goal of increasing social acceptance of bioproducts.

6 Platform deployment

Revision of the Cloud Architecture Plan with OVH cloud

In the following sub sections, we describe the architecture using Amazon Web Services (AWS) as the chosen cloud platform. However, we have since decided to switch to OVH Cloud instead. This decision was made because OVH Cloud is smaller in scale and easier to manage, making it better suited to our needs. Although the components will be packaged and deployed in a similar manner, some differences in implementation will apply.

This section presents the tools and technologies that will be used for the deployment of the BioReCer platform. The main objective is to design a cost-effective, state-of-the-art and scalable platform. Indeed, it must be able to handle the first use cases that will be integrated in the scope of the project without implying expensive infrastructure costs, but it must also be able to handle all the future users and use-cases without any degradation of the service and without implying too much effort, but still by keeping infrastructure costs under control.

That is why a choice has been made to deploy the platform onto the AWS infrastructure. It allows us to apply the cloud native principles and to scale up as needed as the load increases. It also provides advanced tools to guarantee the global security of the platform and to finely monitor the costs of the infrastructure.

6.1 Cloud native approach

The objective of the first deployment of the BioReCer platform is to quickly deploy a first operational version of the platform without requiring too many modifications on the existing applications that are to be integrated. Indeed, for this very first deployment, it is preferable to focus on the best practices required for a state of the art cloud deployment of the platform. Another focus will also be to gain some insights on the costs induced by the use of the platform, to finely analyse them, and ultimately to deliver a cost optimised version of the platform in the next iterations.

In this first iteration, we will thus focus on the preparation of the components so that they can be deployed in a pure cloud architecture. That mainly covers the following tasks:

- Package the components in a format that can be used in a cloud architecture (typically using Docker format but any OCI-compliant formats may be used).
- Ensure the components can be natively monitored, at least by exposing a health endpoint and preferably by also exposing more advanced metrics endpoints.
- Prepare the components so that they can horizontally scale, which typically implies that they become fully stateless.
- If possible, integrate the platform with a continuous integration and delivery tool that will automatically deploy and upgrade a component when it has a new version available.

The figure below illustrates the envisioned first version of the BioReCer platform deployment:

- An API gateway service is in charge of receiving external HTTP requests, checking and applying the security constraints, and routing them accordingly.
- A load balancer service is in charge of delivering the requests to one of the downstream services, according to the current load and target service.
- The core BioReCer platform is deployed into an AWS Fargate cluster. AWS Fargate is a serverless, pay-as-you-go compute engine for containers. Inside this cluster, we will make use of the Amazon Elastic Container Service (Amazon ECS) or Amazon Elastic Kubernetes Service (Amazon EKS) to execute the components. All the services that can be deployed as OCI-compliant images will be deployed inside the cluster. The cluster will be accompanied by the Amazon Managed Streaming for Apache Kafka. This service will be used to decouple the data exchange between the modules of the BIT platform, as it is typical in platforms made up of many interacting services.
- The monitoring of the platform is handled by the Amazon CloudWatch service, whose role is to collect and visualise real-time logs, metrics, and event data in automated dashboards to streamline infrastructure and application maintenance.
- If some BioReCer applications cannot be ready yet to be deployed into the AWS Fargate cluster, they will be deployed into Amazon EC2 instances that allow to deploy any application inside a standard virtual machine.
- In the same way, NiFi (the data ingestion solution) will be deployed on an Amazon EC2 instance as it is not yet fully ready to be deployed as a compliant containerized application (especially, its runtime configuration is currently difficult to manage in such a deployment).
- Finally, the Amazon Managed Service for Grafana is used to offer a lightweight, easy to use data monitoring solution. It is especially useful when sensor data must be visualised and monitored.

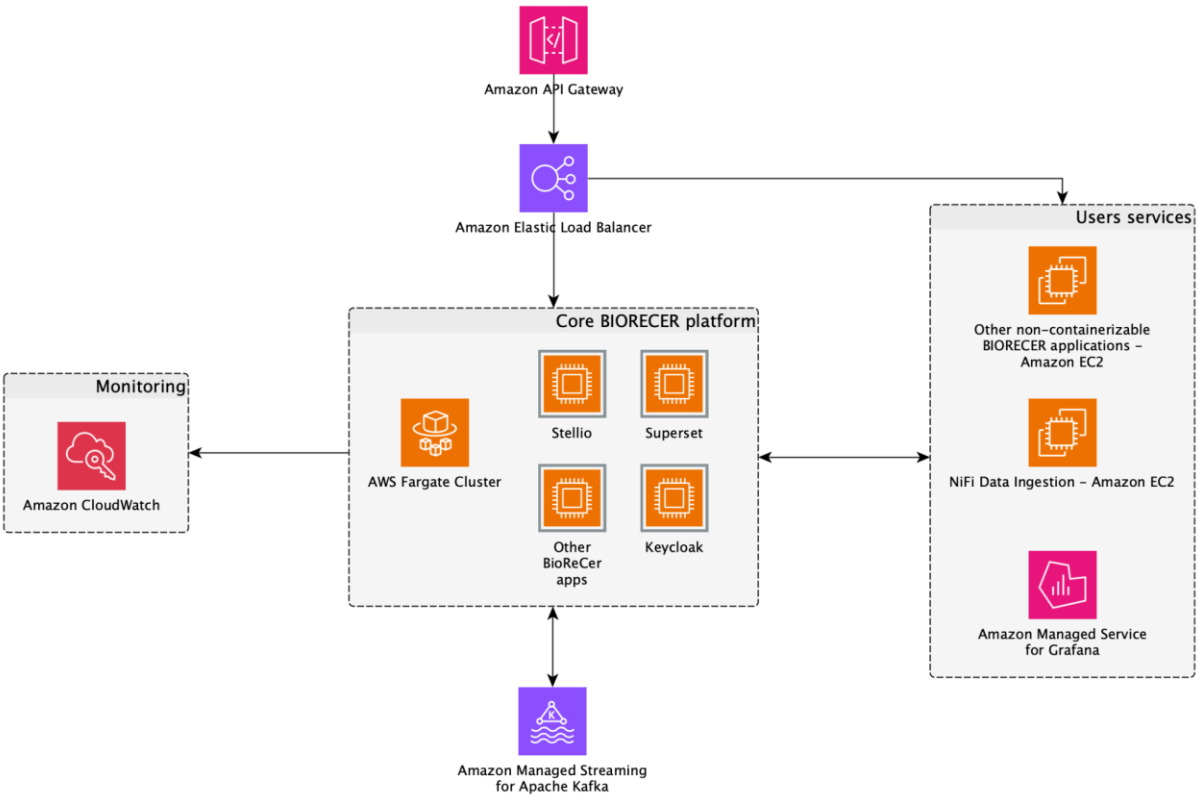


Figure 34. First iteration of the BioReCer platform deployed in AWS

The second iteration of the platform is based upon the first version and leverages it to improve the overall scalability and to optimise the costs. The two notable changes are the following:

- Most of the services run through the AWS Lambda service, a serverless, event-driven compute service. Using this service, services are only running when there is some action to handle, thus optimising the infrastructure costs, and also improving the scalability since as many lambdas as needed may run in parallel.
- In the same way, the Stellio database services will execute using the Amazon Aurora Serverless service. Indeed, Stellio uses the PostgreSQL database for data storage and the Amazon Aurora Service service is compatible with PostgreSQL. However, it does not support the Timescale extension which is used by Stellio to manage time series data, so a preliminary work is required on Stellio side to be able to use the Amazon Aurora Service.

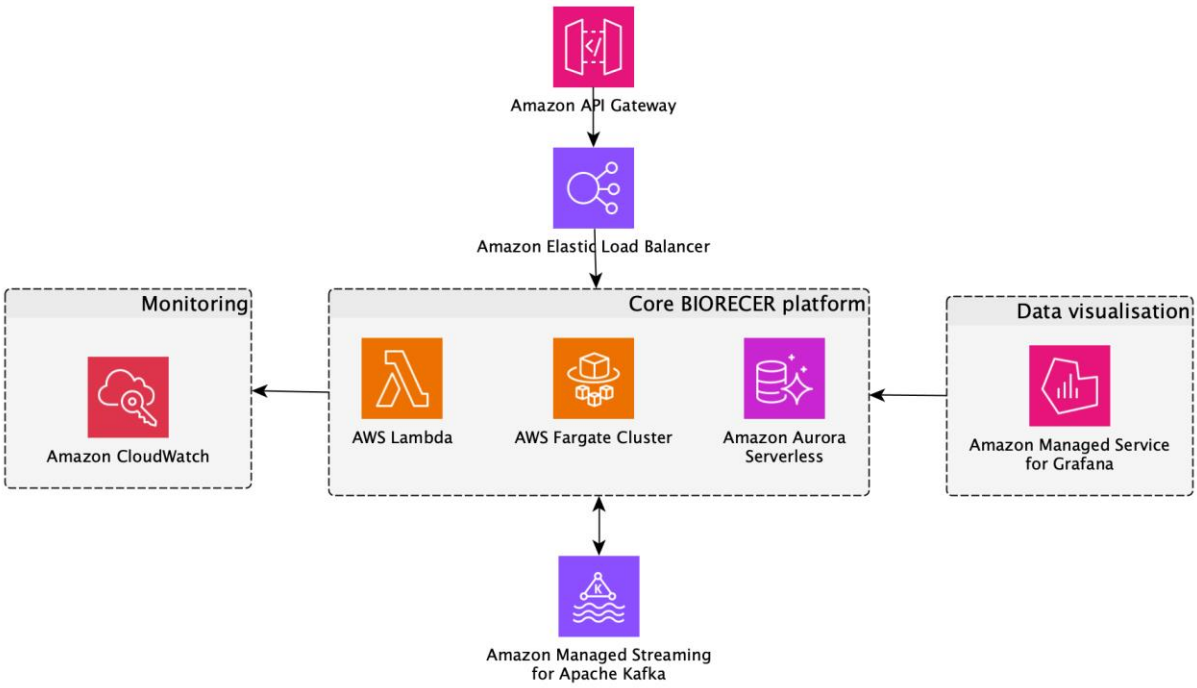


Figure 35. Final version of the BioReCer platform deployed in AWS

6.2 Scalability performance evaluation

It is important to ensure that the platform will be able to handle any future load and if not to know what the limitations are, who are the bottlenecks, and to identify ways to get rid of these limitations and bottlenecks. This can for instance be done by improving the code of a service deployed in the BIT platform, or by updating the deployment (for instance by adding a read only replica of a database instance or by allocating more resources to a service deployed in the AWS Fargate cluster).

This performance evaluation will be realised using the k6[20] tool. It is a powerful, mature, widely used and easy to set up tool for running performance evaluation tests suites.

Its main features are:

- It can run different kinds of performance tests: stress tests, spike tests, soak tests, smoke tests, ...
- It can mix browser and API testing.
- It can inject faults to test the resilience and tolerance of applications.
- It provides features to test how cloud native systems scale and to identify bottlenecks.

[20] https://k6.io/

- It has built-in integrations with external tools to visualise the results of a suite of tests (for instance, Grafana) or to send the results to an external component (for instance, Prometheus).
- It can easily be used in most of the usual tools used by developers (Postman, Visual Studio, etc.)

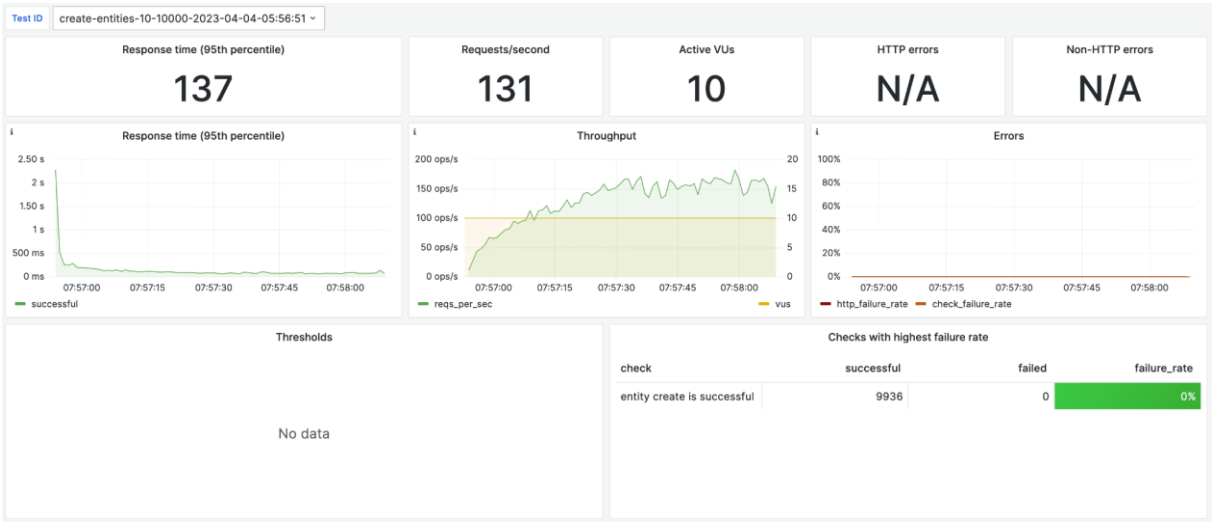


Figure 36. Sample Grafana dashboard showing the results of a performance tests

7 Conclusion

The BioReCer technical architecture is designed to provide a secure, scalable, and interoperable platform for collecting, storing, processing, and visualizing data related to the bio-based industry. It is an ecosystem meant to provide the data and tools needed to ensure that biological feedstocks are used in a way that minimizes their environmental impact.

The BioReCer technical ecosystem can be divided into the following layers:

- **The Data layer**: responsible for collecting, storing, and processing data. It uses a variety of data sources, including sensors, external APIs, and open data portals.
- **The Application layer**: provides a variety of applications for users to interact with the platform. These applications include data visualization tools, traceability tools, certification tools…
- **The Trust and Security layer**: protects the platform from unauthorized access and data breaches. Ensuring trust with certification schemes, it also uses a variety of security measures, such as firewalls, intrusion detection systems, and encryption.

By leveraging these different layers, the platform will indeed enable the three main pillars of the BioReCer project:

- **Traceability**: traceability of bioproducts from their origin to their end use. This will help to ensure the sustainability and environmental performance of bioproducts.
- **Certification**: the platform will support the development of complements for existing certification schemes for bio-circular products. This will help to increase the social acceptance of bioproducts.
- **Market intelligence**: the platform will provide market intelligence to bio-based industry stakeholders. This will help them to make informed decisions about production, procurement, and marketing.

At the heart of this multifaceted approach is the BioReCer ICT tool, which acts as the bridge between conceptual models and practical implementation. This digital interface serves as a catalyst for disseminating information and allowing acceptability through trust and security, and empowering stakeholders with near-real-time insights to guide their actions and choices.

As the project continues to unfold, this deliverable marks an important early milestone as a technical reference in the shaping of the future BIT. The cloud-native approach, with a focus on AWS, guarantees seamless deployment and scalability, accommodating the dynamic needs of the bio-based industries. Overall, the BIT ecosystem will pave the way

to a powerful platform, able to ensure long-term sustainability thanks to interoperability, scalability, and open standards.

8 List of abbreviations

BIT	BioReCer ICT tool
BWA	BioReCer Web Application
DID	Decentralized Identifiers
DLT	Distributed Ledger Technology
LCA	Life Cycle Assessment
MFA	Material Flow Analysis
MLOps	Machine Learning Operations
GP	Gaussian Process
VC	Verifiable Credentials
DSS	Decision Support System

9 References

Apache Software Foundation. (n.d.). *Exploration and visualization platform*. Retrieved from Apache Superset: https://superset.apache.org/

CIRCFORBIO. (2023). *Main page*. Retrieved from CIRC for BIO supplier shop: https://circ4bio.com

EBSI Verifiable Credentials Explained. (n.d.). Retrieved from ec.europa: https://ec.europa.eu/digital-building-blocks/wikis/download/attachments/597952490/Chapter%200%20-%20Verifiable%20Credentials%20An%20introduction.pdf?version=1&modificationDate=1676459051355&api=v2

ETSI. (n.d.). *Context Information Management.* Retrieved from https://www.etsi.org/deliver/etsi_gs/CIM/001_099/009/01.04.02_60/gs_CIM009v010402p.pdf

FIWARE. (n.d.). *What is the FIWARE foundation?* Retrieved from https://www.fiware.org/foundation/

Grafana Labs. (2023). *Load testing for engineering teams*. Retrieved from K6: https://k6.io/

Grafana Labs. (2023). *The open observability platform*. Retrieved from Grafana: https://grafana.com/

ISCC. (2023). *All certificates*. Retrieved from ISCC system: https://www.iscc-system.org/certification/certificate-database/all-certificates

Munteanu, A. (2022). *What is MLOps?* Retrieved from Canonical: https://canonical.com/blog/what-is-mlops

OpenID Foundation. (n.d.). *Explore all specifications.* Retrieved from https://openid.net/specs

Redhat. (2023). *Open Source Identity and Access Management*. Retrieved from Keycloak: https://www.keycloak.org/

Romero Ugarte, J. L. (2018). Distributed ledger technology (DLT): introduction. *Banco de Espana*.

UNECE. (2023). *Reference Data Models (RDM).* Retrieved from unece.org: https://unece.org/trade/uncefact/rdm

W3C. (n.d.). *Verifiable Credentials Data Model v2.0.* Retrieved from https://www.w3.org/TR/vc-data-model-2.0/

Yingzhong Zhang, X. L. (2015). LCA-oriented semantic representation for the product life cycle. *Journal of Cleaner Production*, 146-162.