D3.3 – Data Collection, Security, Storage, Governance & Management Services Bundles – Beta Release

Grant Agreement n°	**957020**
Project Acronym	**BEYOND**
Project Title	**A reference big data platform implementation and AI analytics toolkit toward innovative data sharing-driven energy service ecosystems for the building sector and beyond**
Starting Date	01/12/2020
Duration	36
EU Project Officer	Stavros STAMATOUKOS
Project Coordinator	UBITECH
Consortium Partners	VTT, FVH, CIRCE, Suite5, IGM, KONCAR, ARTELYS, MYTILINEOS, CUERVA, BELIT, URBENER, BEOELEK,
Project Website	beyond-h2020.eu
Cordis	https://cordis.europa.eu/project/id/957020

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement n° 957020.

Deliverable No.	**D3.3**
Deliverable Title	**Data Collection, Security, Storage, Governance & Management Services Bundles – Beta Release**
Work Package	WP3 - End-to-end Interoperable Big Data Management Platform
WP Leader	SUITE5
Due Date	31/01/ 2022
Actual Date of submission	01/02/2022
Version	V1.00
Status	Final Version
Dissemination Level	Public
Authors	Suite5, UBITECH
Reviewers	IGM, UBITECH

Disclaimer: *The present report reflects only the authors' view. The European Commission is not responsible for any use that may be made of the information it contains.*

BEYOND

Version	Modification(s)	Date	Author(s)
0.10	First ToC	17/09/2021	Suite5
0.20	Draft Chapter 1, 2, 3	11/10/2021	Suite5
0.30	Draft Chapter 4	14/10/2021	Ubitech
0.40	Updated Chapter 4, 5	16/11/2021	Suite5
0.50	Initial partner contribution (chapter 4)	03/12/2021	Ubitech
0.60	Updated Chapter 6	27/12/2021	Suite5
0.70	Updated partner contribution (chapter 4)	20/01/2022	Ubitech
0.80	Final draft ready for peer review	26/01/2022	Suite5
0.90	Updated as peer review comments	31/01/2022	Suite5
1.00	Final Version submitted to EC	01/02/2022	Suite5

EXECUTIVE SUMMARY

D3.3 documents the outputs of tasks "T3.3 - Platform Backbone Infrastructure, On-Premise and Secure Experimentation Playground Data Containers and Core Services Development" and "T3.4 – "Data Assets Security, Encryption and Privacy Mechanisms"; of WP3 "End-to-end Interoperable Big Data Management Platform",

As the actual deliverable is of type OTHER, this document comes as an accompanying report describing the design of the different components that have been developed and that are delivered as part of D3.3.

In this direction, the main scope of this deliverable is to report on the outputs of T4.1; presenting the early technical specifications and implementation activities for each of the modules/components forming these services and which are namely: a) the Data Ingestion Services, b) the Polyglot Data Storage Layer c) the Data Security services bundle and d) the Cloud Platform Operations Manager.

It shall be noted that this beta release corresponds only to the BEYOND Cloud based Platform Infrastructure, while the BEYOND Private infrastructure will be addressed in the forthcoming release. Additionally, all the services along with their envisioned functionalities described in this deliverable, are based on the latest BEYOND Architecture (as defined in D2.6 [3]) thus differences may be noticed by the readers in terms of their naming (in contrast to their original names stated in the DoA).

Towards this end, the deliverable at hand reports the preliminary development specifications and implementation details of the components/services by presenting:

- an overview of the services and the key functionalities delivered in their beta release,
- the implementation status of the respective functional requirements for each of these services, as defined in BEYOND deliverable D2.6.
- their internal architecture and the technology stack upon which they are built.
- any assumptions and restrictions considered in this beta release along with their accompanying licensing.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

List of Abbreviations

Abbreviation	Full text
ABAC	Attribute-Based Access Control
ABE	Attribute-Based Encryption
APIs	Application Programming Interface
BEYOND	A reference big data platform implementation and AI analytics toolkit toward innovative data sharing-driven energy service ecosystems for the building sector and beyond
BoB	Best of Breed
CIM	Common Information Model
CPOM	Cloud Platform Operation Manager
DIS	Data Ingestion Services
Dx.y	Deliverable x.y
DoA	Description of Action
ETH	Ethereum
HTTP	Hypertext Transfer Protocol
Mx	Month x
PEP	Policy Enforcement Point
PDP	Policy Decision Point
PDSL	Polyglot Data Storage Layer
SEC	Security Services
SSE	Server-Sent Events
Tx.y	Task x.y
UC	Use Case(s)
UUID	Unique User Identification
WPx	Work Package x
XACML	eXtensible Access Control Markup Language

1. Introduction

1.1. Scope and objectives

D3.3 distils the outcomes of "T3.3 - Platform Backbone Infrastructure, On-Premise and Secure Experimentation Playground Data Containers and Core Services Development" and "T3.4 - Data Assets Security, Encryption and Privacy Mechanisms; of WP3 "End-to-end Interoperable Big Data Management Platform", towards the delivery of the beta release of five distinct but interdependent services bundles offered through the BEYOND Platform, namely the Data Ingestion Services (DIS), the Polyglot Data Storage layer (PDSL), the Cloud Platform Operational Manager (CPOM) and a set of security components incorporated in the aforementioned services and which collectively form the Data Security services bundle.

In alignment with the BEYOND DoA, the Data Collection, Security, Storage, Governance & Management Services Bundles will be delivered in three distinct releases through the project's duration: in M14 (beta release – D3.3.), and in M18 (Release 1.00 – D3.6) As anticipated, the first release of the services Bundles, shall be built on the outcomes of this deliverable and will contain all the planned functionalities, as well as refinements and enhancements based on the updated outcomes of the design, specification and integration activities performed in WP3.

Finally, it shall be noted that this beta release corresponds only to the BEYOND Cloud based Platform Infrastructure, while the BEYOND Private infrastructure will be addressed in the forthcoming release. Additionally, all the services and components along with the envisioned functionalities described in this deliverable, are based on the latest BEYOND Architecture (as defined in D2.6 [3]) thus differences may be noticed by the readers in terms of their naming, in contrast to their original terms as stated in the DoA.

1.2. Relation to other tasks/deliverables

The present deliverable documents the activities performed in T3.3. and T3.4 and its main scope is to deliver the beta version of the BEYOND Data Collection, Security, Storage, Governance & Management services bundles. Towards this direction, for the design and implementation of the components described in this deliverable, D3.3 receives input from the following BEYOND tasks and associated deliverables:

- T2.1 - Definition of Business Scenarios, Use Cases and Elicitation of user & business requirements; where the updated technical requirements and the Use Cases (UCs) of the BEYOND project are defined and documented in D2.2 and which are used to drive the overall implementation activities for the services bundles presented in this document.
- T2.5 - Detailed architecture design, protocols, and interfaces specifications for Big Data-enabled Energy Services; where in D2.6 the initial design specifications for the services bundles forming the BEYOND integrated platform are documented.

The outcome of the activities performed in D3.3 will be also used as input in the following BEYOND tasks/deliverables and work packages:

- T3.5 - Platform and Services Bundles Continuous Integration and Open APIs,; where the Data Collection, Security, Storage, Governance & Management Services is naturally part of the system level software integration activities towards the delivery of the first release of the BEYOND big data platform.
- D3.5 - BEYOND Integrated Platform & Open APIs – Beta Release; where D3.3 will set the foundations for the delivery of the beta version of the integrated BEYOND Platform.
- D3.6 – Data Collection, Security, Storage, Governance & Management Services Bundles – Release 1.00, which will be based on the feedback received on its beta release and will include enhancements and updates

Moreover, D3.3 will provide a better understanding of the data provisioning alternatives, and semantic interoperability repercussions offered by the BEYOND Platform to all BEYOND End-User tools and services that are delivered in the context of WP5 “AI Analytics-based Decision Support Suite for Optimizing Energy Policy Planning, Infrastructure Sizing and De-risking Renovation Investments” and WP6 “AI Analytics-based Innovative Energy Services Suite towards Optimized Buildings Energy Performance Management”. The tools and services are also expected to provide feedback and lessons learnt from the real-life application of the BEYOND services towards the delivery of its first release initially in M18 and in its final release in M32.

1.3. Structure of the document

In order to address all the aspects relevant to the scope of D3.3, the remaining of this document has been structured as follows:

- Chapter 2 provides generic details on the bundles that have been developed and are provided as part of the beta release.
- Chapter 3 provides a comprehensive documentation of the Data Ingestion Services.
- Chapter 4 provides a comprehensive documentation of the Data Security services bundle.
- Chapter 5 provides a comprehensive documentation of the Polyglot data Storage Layer.
- Chapter 6 provides a comprehensive documentation of the Cloud Platform Operational Manager.
- In chapter 7 the next steps towards release 1.00 of the bundles are provided.

2. Beta Release of Data Collection, Security, Storage, Governance & Management Service Bundles

D3.3 reports on the early outcomes of T3.4 & T3.4 activities by M14 of the project's implementation; presenting the early technical specifications and implementation activities for each of the modules/components forming the BEYOND Data Ingestion Services, the Polyglot Data Storage Layer, the Cloud Platform Operational Manager, and the Data Security services bundle. In more detail, the following apply:

- **Data Ingestion Services** are utilized in the data check-in process, for the collection of data assets through various external sources into the BEYOND Cloud based Platform and, in their mapping, harmonisation and transformation to the appropriate formats so as to be available to any authorised platform user.
- **Polyglot Data Storage Layer**, acting as the platform's central repository also offering indexing capabilities.
- **Data Security services** bundles which ensure the security and privacy preservation of the data assets ingested by the platform,
- **Cloud Platform Operational Manager**, responsible for the management of the various platform's components and orchestration of the underlying resources, ensuring the overall platform's secure execution and efficient operation.

These components are highlighted in red in the following figure depicting the overall BEYOND Cloud based Platform architecture as defined in D2.6 [3].

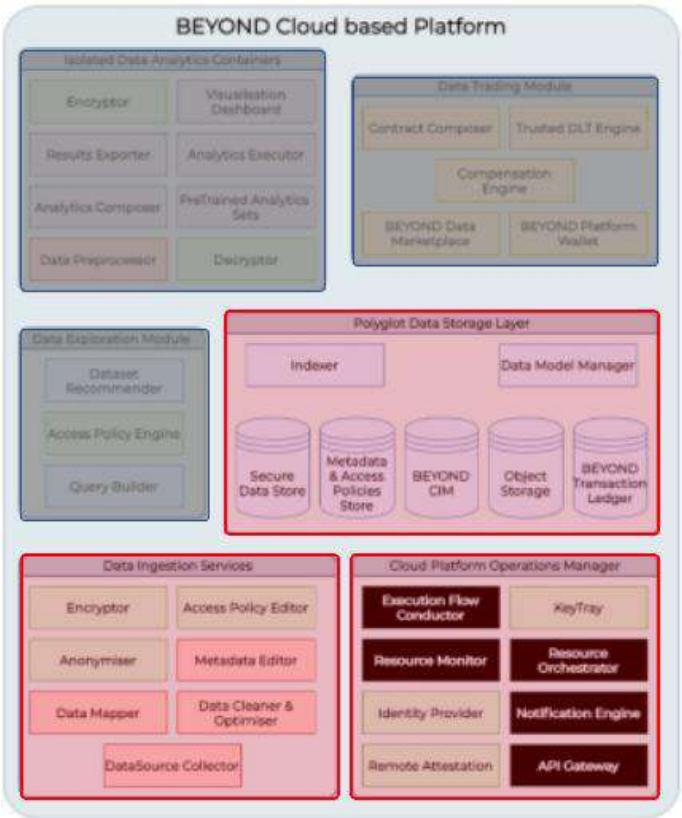

FIGURE 1 BEYOND CLOUD BASED PLATFORM ARCHITECTURE – HIGHLIGHTED SERVICES OF INTEREST

From an Asset Providers perspective, the aforementioned bundles have an instrumental role, as they are practically responsible for: a) collecting building-domain relevant data assets from external sources and in multiple modalities, b) processing them according to the provisions of the BEYOND Common Information Model (CIM), and c) finally storing and indexing the transformed data and their accompanying information (metadata) so as to be eventually available to all authorized BEYOND Platform actors (as described in D2.6 [3]) and BEYOND end users' tools.

For each of the aforementioned services bundles, this deliverable provides the documentation of the actual software that has been developed (for its beta release) and delivered in accordance with the BEYOND requirements and architecture, as defined in D2.6 [3]. In more detail, for each bundle, the following chapters

- provide an overview and key functionalities delivered form the various components/modules forming these services bundles,
- report on the implementation status of the functionalities delivered in this beta release and align with the respective components' functional requirements.
- define the technology stack upon which is based.
- identifies the accompanying licensing and access information of each of the services bundle.

2.1. Bundles API Documentation

The APIs responsible for the communication between the different BEYOND Platform services implemented under D3.3 are documented using Swagger and provided in the project's private repository.

2.2. Bundles Deployment instructions

The detailed deployment instructions for the different BEYOND bundles implemented under D3.3 are provided in the relevant private code repository.

3. Data Ingestion services (beta release)

3.1. Key Functionalities/Architecture

The Data Ingestion Services (DIS) are responsible for organising and undertaking the collection/ingestion and proper manipulation (such as pre-processing into an appropriate format for further analysis, adding of metadata, mapping to the BEYOND Common Information Model (CIM), cleaning, anonymization and encryption) of building domain-relevant data assets deriving from various external sources and in multiple modalities into the BEYOND Platform, so as to finally be available to any authorized BEYOND platform actor and BEYOND end users' tools.

As described in D2.6 [3], Data Ingestion Services facilitate the data asset collection into the BEYOND platform's; These services are responsible for the various data asset check-in processes (i.e., data mapping, data cleaning, data anonymization and data encryption) and are composed from the following four main data management subcomponents:

- The **DataSource Collector** component, facilitating the overall collection of data assets (i.e., data check-in process) from external sources into the BEYOND Platform, offering different data ingestion methods, such as batch data / file uploads, or via Application Programming Interfaces (APIs).
- The **Data Cleaner & Optimiser** delivering data cleaning functionalities, such as removing outliers or missing values.
- The **Data Mapper** undertaking the mapping of the ingested data to the concepts of the latest BEYOND CIM, carrying out also the necessary transformations.
- The **Metadata Editor,** enabling users to add key details (metadata) to their data assets.

It shall be noted that in order for the Data Ingestion Services to provide their full set of functionalities, a number of security subcomponents are also incorporated, namely: (a) the Access Policies Editor, (b) the Encryptor, and (c) the Anonymizer, which are described in detail in chapter 4 of this document.

The beta release of the Data Ingestion services aligns with the specifications and functional requirements defined in D2.6 [3]; the current status of their implementation is presented in the following figure.

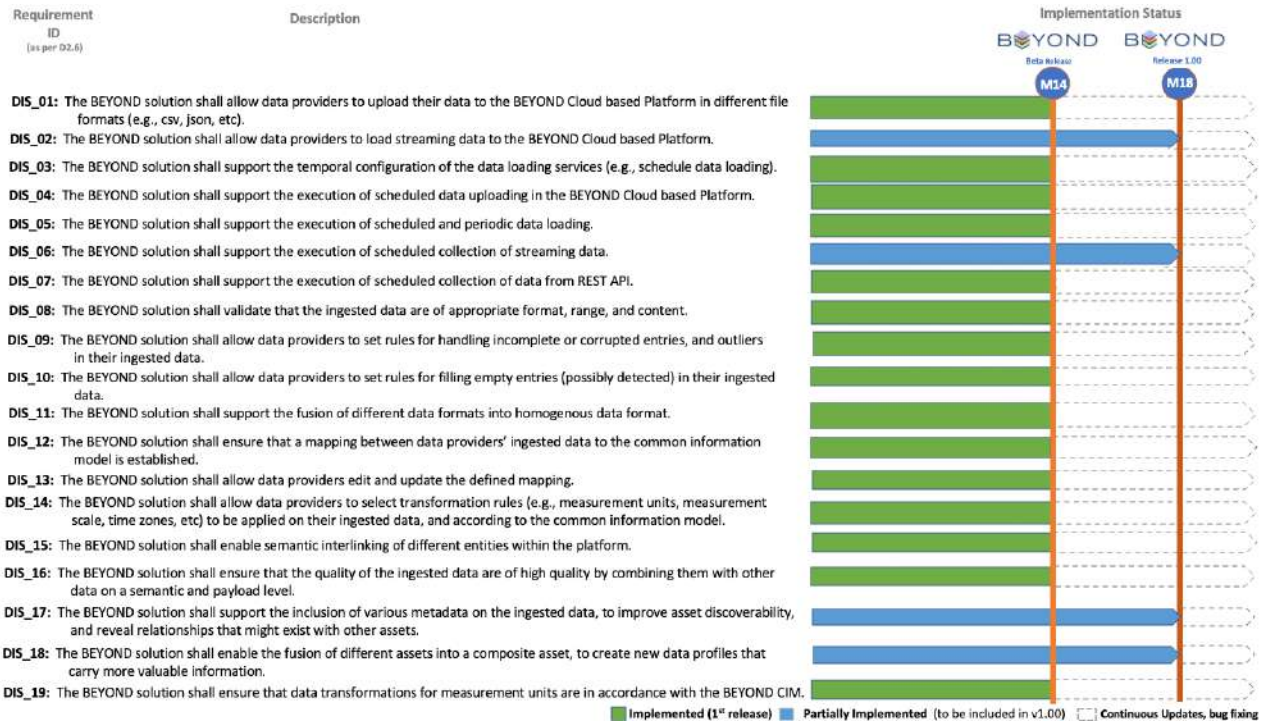

FIGURE 2 DATA INGESTION SERVICES - REQUIREMENTS BACKLOG AND FULFILMENT IN BETA RELEASE

As shown in Figure 3 the Data Ingestion Services are built on state-of-the art technologies across three layers:

- The *Presentation Layer*, containing the User Interface (UI) that is developed in VueJS[1] and TailwindCSS[2]
- The *Business Logic Layer*, containing the different packages of the Data Collection services backend and which is based on the Nest (Node JS)[3] and utilises also RabbitMQ[4] message broker.
- The *Data Access Layer* that essentially refers to the Polyglot Data Storage Layer (PDSL) (see chapter 4) that has been set up and utilizes MongoDB[5], MinIO[6] (as the data lake for the temporary/intermediate files management) and Vault[7] (as the secure database for sensitive and secret parameters).

[1] https://vuejs.org/
[2] https://tailwindcss.com/
[3] https://nestjs.com/
[4] https://www.rabbitmq.com/
[5] https://www.mongodb.com
[6] https://min.io/
[7] https://www.vaultproject.io/

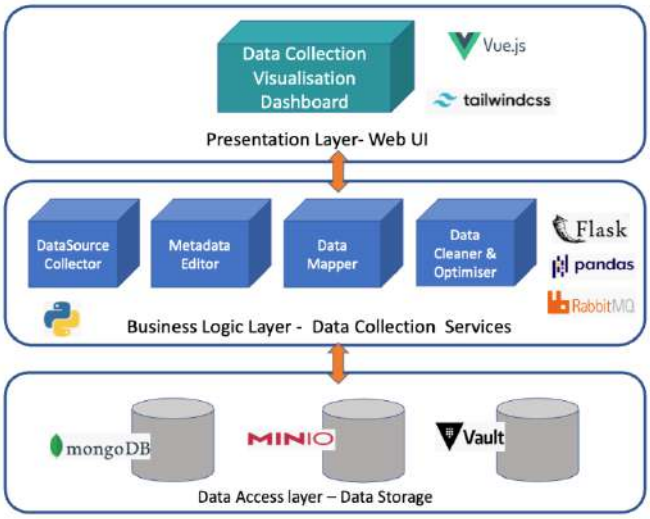

FIGURE 3 ARCHITECTURE OF THE DATA INGESTION SERVICES

It is noted that the above provided architecture is the full fledge architecture, that is envisioned to be delivered with release 2.00 of the platform. Updates to this architecture are to be considered with the delivery of the revised overall architecture of the BEYOND platform, under WP2 of the project.

3.2. Assumptions and Constraints

The following assumptions and constrains apply to the beta release of the components:

- The currently data collection methods available in this beta release include a) direct file upload b) upload through the BEYOND Platform's provided APIs c) upload through the Data Asset Provider's own APIs d) upload of file batch data through an FTP server.
- The collection of streaming data was not considered as a priority for this beta release, considering that the real-time sensor data will be available at a late stage of the project's implementation. However, recognising that there may be cases that real-time exchange of information between the BEYOND end-user tools (that are still at their early development phase at the moment this beta was released), the collection of streaming data will be implemented in the first release of the Data Ingestion services.

3.3. Licensing

The BEYOND Data Ingestion Services consist of closed source components and which will be available through the integrated BEYOND platform, as documented in D3.6.

4. Data Security Services bundle (beta release)

4.1. Key Functionalities/Architecture

The Data Security Services bundle is responsible to deliver a set of services regarding authentication, authorization, data access control policies, data encryption/decryption, data anonymization and remote attestation. A summary of the basic components and their role in the BEYOND ecosystem is given below:

- **Identity Provider**: This component is based on Keycloak service and is used for authenticating the data asset providers, before using the BEYOND cloud-based platform. It can possibly be extended to be used also for the data consumers as a centralized authentication solution.
- **KeyTray**: The main function of this component is to store and provide pairs of encrypted SSE Keys (encryption and verification key), as generated by the encryptor component. The encryption key corresponds to the unique user id (UUID) and the verification key to the Access Token, assigned to the user after successful user authentication with Keycloak.
- **Encryptor/Decryptor**: It implements two distinct flows:

  a) It receives a pair of unencrypted SSE Keys, encrypts them using attribute-based encryption (ABE) and stores them in the KeyTray. The attributes used will be defined in specified policies, after agreement with the rest of the consortium. The same functionality will be used for a data asset (instead of a key) that a data producer wishes to be encrypted. So finally, each data asset will be bound to a specific encryption and verification key.

  b) It requests a pair of encrypted SSE Keys from the KeyTray, tries to decrypt them and if successful, returns the unencrypted pair of SSE Keys. The same functionality will be used for a data asset/analytics result that a data consumer wishes to decrypt.

- **Access Policy Editor**: This component provides policy editors for developing ABAC and ABE policies, and a Policy Validation module for checking policy correctness, completeness and security awareness. As a first step it will support pre-defined policies upon agreement with the technical partners.
- **Access Policy Engine**: This component enhances the Data Handling Module with *Attribute-Based Access Control* (ABAC) authorization. It follows the paradigm of the eXtensible Access Control Markup Language (*XACML) Architecture*, which is the de facto standard for ABAC authorization. By implementing:

  ABAC client (provided as a library) is wired to the HTTP request processing flow of the protected application. This way ABAC client is able to intercept incoming requests, extract the needed information, and ask ABAC server if it is ok to let

request processing continue. ABAC client implements the *Policy Enforcement Point (PEP)* in XACML Architecture.

ABAC server is a standalone service, which accepts authorization requests from ABAC clients, and decides if they can be allowed on not, based on a number of XACML policies. ABAC server is the *Policy Decision Point (PDP)* in XACML Architecture.

- **Anonymizer**: This component enables data asset providers to safeguard the privacy of their data assets; it is responsible for handling any sensitive information that might be included in a data asset by enabling data providers to protect any sensitive information they do not want to expose, through anonymizing part of their dataset, (e.g., anonymization of specific concepts/fields). It shall be noted that even though the Anonymizer forms part of the Security services bundle, the configuration of the data anonymization is performed through the Data Processor during the process of a data check-in job.

- **Remote Attestation**: In the context of integrity, security and trust, between the local device of a BEYOND user and the BEYOND platform, it is important that the local device be able to attest to the remote server (platform) that it is both authorized and is also in a known configuration state. The local device needs to have access to a Trusted Platform Module (TPM) in order for the two subsequent phases to take place: the enrolment phase and the attestation phase.

The beta release of the Data Security Services aligns with the specifications and functional requirements defined in D2.6 [3]; the current status of their implementation is presented in the following figure.

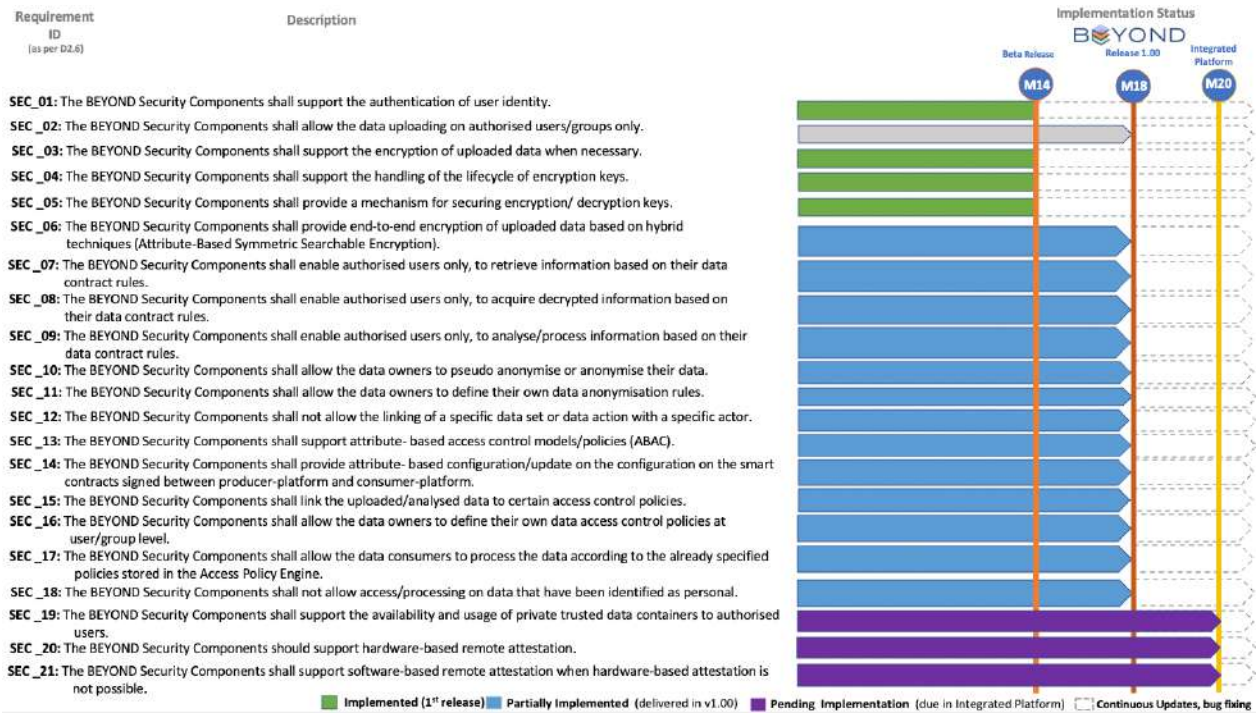

FIGURE 4 DATA SECURITY COMPONENTS - REQUIREMENTS BACKLOG AND FULFILMENT IN BETA RELEASE

4.2. Technology Stack

As shown in Figure 5 the Data Security Services are built on state-of-the art technologies across four layers:

- The *Presentation Layer*, containing the User Interface (UI) that is developed in VueJS[1] and TailwindCSS[2], as already shown in section 3.2.
- The *Authentication and User Configuration Layer*, containing the Identity Provider and the Access Policy Editor that are developed in Spring Boot[8].
- The Remote Attestation Layer, developed in C++98, using the IBM Trusted Software Stack (TSS)[9].
- The *Business Logic Layer,* containing the Encryptor, Decryptor, Anonymizer, Access Policy Engine and Keytray components, that are developed in Spring Boot[8].
- The Storage Layer which uses PostgreSQL[10] (as the database for the storage of data related to the business logic layer services) MinIO[6] (as the data lake for the temporary/intermediate files management such as the anonymized or encrypted data) and Vault[7] (as the secure database for sensitive and secret parameters such as passwords and tokens).

[8] https://spring.io
[9] https://develop.trustedcomputinggroup.org/2018/03/20/ibms-tpm-2-0-tss/
[10] https://www.postgresql.org

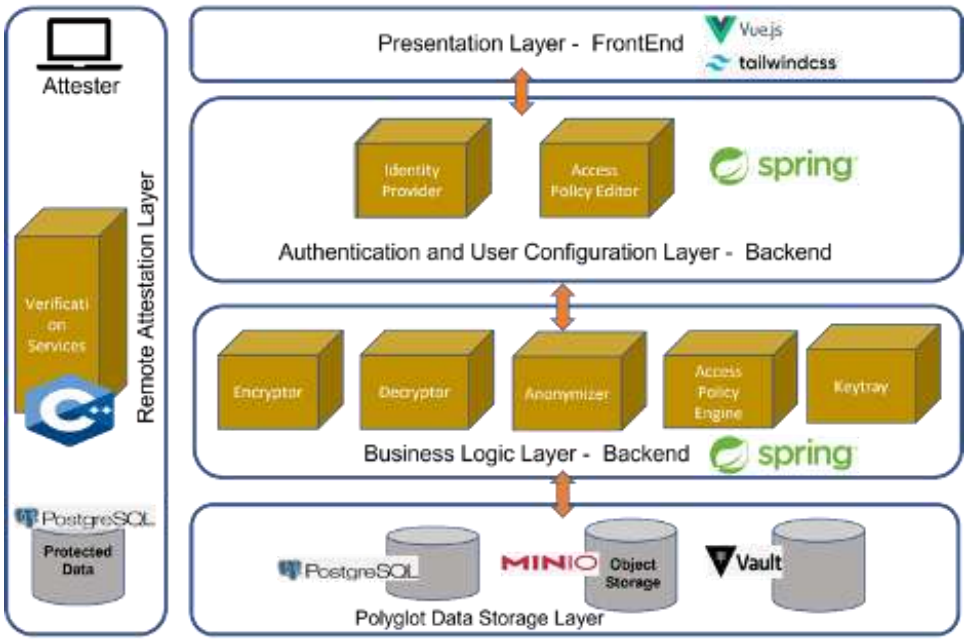

FIGURE 5 ARCHITECTURE OF THE DATA SECURITY SERVICES

It is noted that the above provided architecture is the full fledge architecture, that is envisioned to be delivered with release 2.00 of the platform. Updates to this architecture are to be considered with the delivery of the revised overall architecture of the BEYOND platform, under WP2 of the project.

4.3. Assumptions and Constraints

The following assumptions and constrains apply to the beta release of the components:

- The Keycloak configuration regarding the specifics on user roles, attributes is not completed at this stage as discussions with the stakeholders are pending.
- Due to the fact that the integration with the PDSL will be done at a later stage, in this beta stage only PostgreSQL will be considered.
- The deployment will be carried out using only docker containers (without Kubernetes orchestration).
- Because of the additional configuration needed regarding the TPM nodes, Remote Attestation Services will not be included for now in the solution.

4.4. Licensing

The beta release of the Data Security Services bundle is available through the integrated BEYOND Cloud based platform (see D3.3) and delivered as closed source components.

5. Polyglot Data Storage Layer (beta release)

5.1. Key Functionalities/Architecture

Once the data collection process is completed and the finalized data payload is available (validated and mapped to the BEYOND CIM), data assets are then securely stored in the BEYOND Cloud Based Platform. This is possible through the delivery of the BEYOND Platform's Polyglot Data Storage Layer (PDSL) offering secure data and metadata storage along with indexing capabilities, thus essentially addressing the platform's storage needs.

In general, the PDSL consist of a storage services bundle acting as the central storage layer of the BEYOND platform and is responsible for the persisting storage of all the various ingested data assets (e.g., datasets, analytics models, analytics results, etc.), along with their complementary information (such as access policies, sensitive data, data transactions information, etc.) in a secure and reliable manner.

As described in D2.6[3], the PDSL leverages the best of breed of different storage and indexing components; with the latter being responsible for organising the indexing of the collected data assets and their associated metadata, as well as for the management of the BEYOND CIM. These components are namely: a) the **Secure Data Store**, b) the **Metadata & Access Policies Store**, c) the **BEYOND CIM**, d) the **Data Model Manager**, e) the **Object Storage** and f) the **BEYOND Transaction Ledger**.

In more detail, the functionalities delivered through the components/stores forming the Polyglot Data Storage Layer are enlisted as follows:

- **Storage of the data collection jobs** (i.e., processed data assets and sample data) and their accompanying information in the platform's relational database.
- **Storage of a data asset's metadata**, while preserving its information and the links between the data assets stored and their accompanying metadata.
- **Storage of BEYOND Platform's operational data**, preserving all platform's operations data and accompanying metadata, as well as data related to users' and organisations' profiles.
- **Storage of analytics models**, in a dedicated storage container for pre-trained algorithms, model parameters etc.
- **Storage of data analytics jobs** and their configurations a space for storing configurations to be used when needed for logging, cloning, updating, or re-execution.
- **Storage of sensitive data and credentials**, such as tokens, usernames, passwords, and API keys in a dedicated secure database.
- **Temporary storage of intermediate files and objects**, such as intermediate configuration data files in an appropriate data lake, thus enabling the pause of an operation (such as a data collection job, a data analytics job, etc.) and its continuation in the future when required by the users. This is also very useful for

traceability purposes in case a specific job is interrupted or not successfully completed.

- **Storage of the BEYOND CIM** and its various versions along with its concepts and other domain vocabularies that are used within the BEYOND Platform.
- **Storage of the Data Asset Contracts** carried out in the BEYOND Platform in a distributed ledger, safeguarding the privacy of the contracts' info (such as pricing details, involved parties, etc.) while also ensuring, traceability and non-repudiation.

The beta release of the Polyglot Data Storage Layer aligns with the specifications and functional requirements in D2.6 [3]; the current status of their implementation is presented in the following figure.

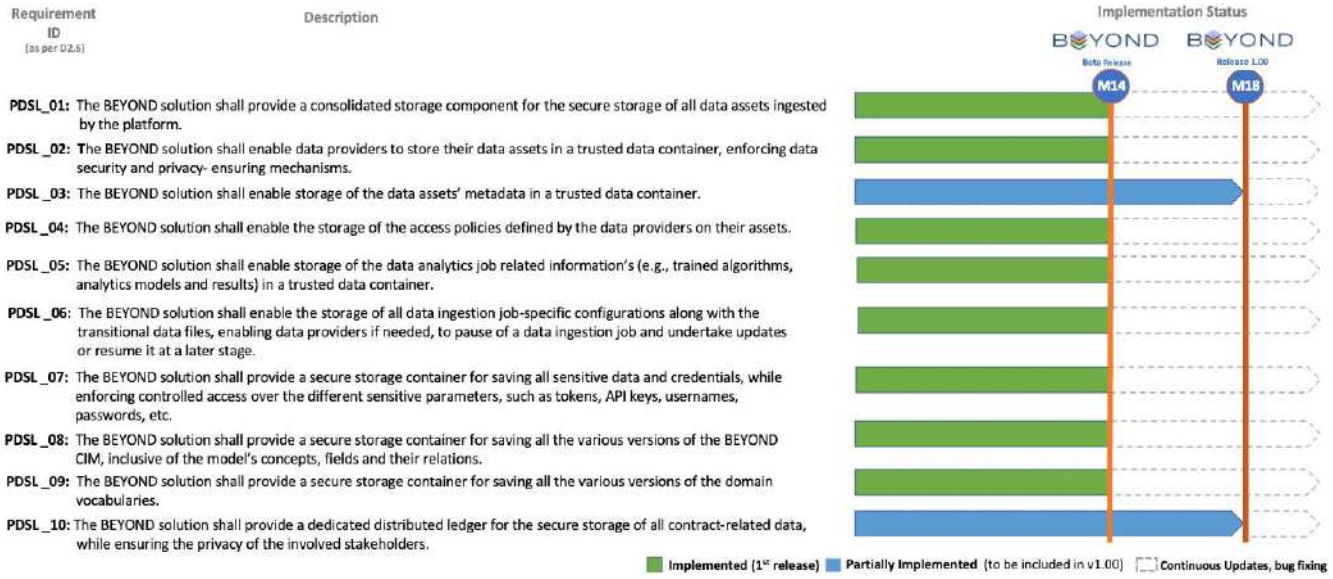

FIGURE 6 POLYGLOT DATA STORAGE LAYER - REQUIREMENTS BACKLOG AND FULFILMENT IN BETA RELEASE

In its beta release, the Polyglot Data Storage Layer is built on state-of-the-art database technologies including PostgreSQL[10] as the BEYOND Cloud based platform's relational database, MongoDB[5] as the NoSQL database for the storage of the various datasets, Elasticsearch[11]as the search optimisation and indexing engine, MinIO [6] as the temporarily storage data lake, Vault[7] for storage of sensitive data (e.g., passwords, APIs keys, etc.), Ethereum[12] as the transaction ledger technology, HDFS[13] as the distributed file system, and Gitlab[14] ,leveraged as the algorithms repository. These various storage containers along with the technologies leveraged for their implementation are depicted in the following figure.

[11] https://www.elastic.co
[12] https://ethereum.org/en/
[13] https://hadoop.apache.org/
[14] https:// gitlab.com/

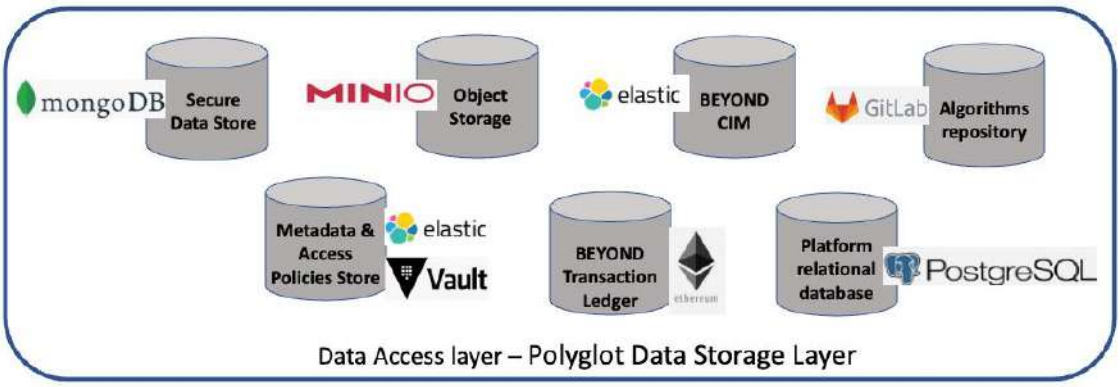

FIGURE 7 ARCHITECTURE OF THE DATA STORAGE SERVICES BUNDLE

It is noted that the above provided architecture is the full fledge architecture, that is envisioned to be delivered with release 2.00 of the platform. Updates to this architecture are to be considered with the delivery of the revised overall architecture of the BEYOND platform, under WP2 of the project.

5.2. Assumptions and Constraints

N/A

5.3. Licensing

The beta release of the Polyglot Data Storage Layer is available through the BEYOND Cloud based platform and delivered as a closed source component.

6. Cloud Platform Operations Manager (beta release)

6.1. Key Functionalities/Architecture

The Cloud Platform Operations Manager (CPOM) facilitates the effective orchestration of the various BEYOND platform's components (i.e., modules, layers, and containers), as well as the coordination of the underlying resources, and the management of appropriate notifications, while guaranteeing the BEYOND platform's effective operation.

As described in D2.6 [3], the CPOM consist of both security and system management components. In more detail, the CPOM consists of five core system management components:

- the **Execution Flow Conductor**, being responsible for the efficient orchestration of the various activities carried out in the BEYOND platform, also managing their internal interactions, by triggering the appropriate modules and providing them with the required information required for their effective operation.
- the **Resource Monitor**, offering monitoring of all the various data assets that reside in the BEYOND Cloud based Platform, delivering visualizations and aggregated statistics that can reveal insights regarding the usage of the different BEYOND Cloud based Platform components and services as well as insights of the various data assets usage.
- the **Resource Orchestrator**, responsible for the efficient operation of the BEYOND platform, ensuring the successful implementation of all data check-in/ data analysis related jobs, by triggering the different container and running the different processes, allocating the appropriate computing and storage resources required for their effective execution.
- the **Notification Engine,** which is responsible for informing the users of the BEYOND Platform, through real-time notification messages, about the status and advancement of the various activities taking place in the platform, and which are of their concern (e.g., progress of a data ingestion job and/or a data analysis job, etc.) as well as notify the responsible individuals of an organisation regarding the progress of data asset
- the **API Gateway**, which oversees and manages all the API requests from external applications, by acquiring and combining all the suitable data from the BEYOND Data Storage Services to appropriately respond to the requests.

It shall be noted that the CPOM also includes three distinct security components facilitating the security-related functionalities within the BEYOND Platform, enabling the registration, authorization, and authentication of organisations and users. These components are namely: the KeyTray, the Identity Provider, and the Remote Attestation that are described in detail in chapter 4 of this document.

The beta release of the Cloud Platform Operations Manager aligns with the specifications and functional requirements defined D2.6 [3]; the current status of their implementation is presented in the following figure.

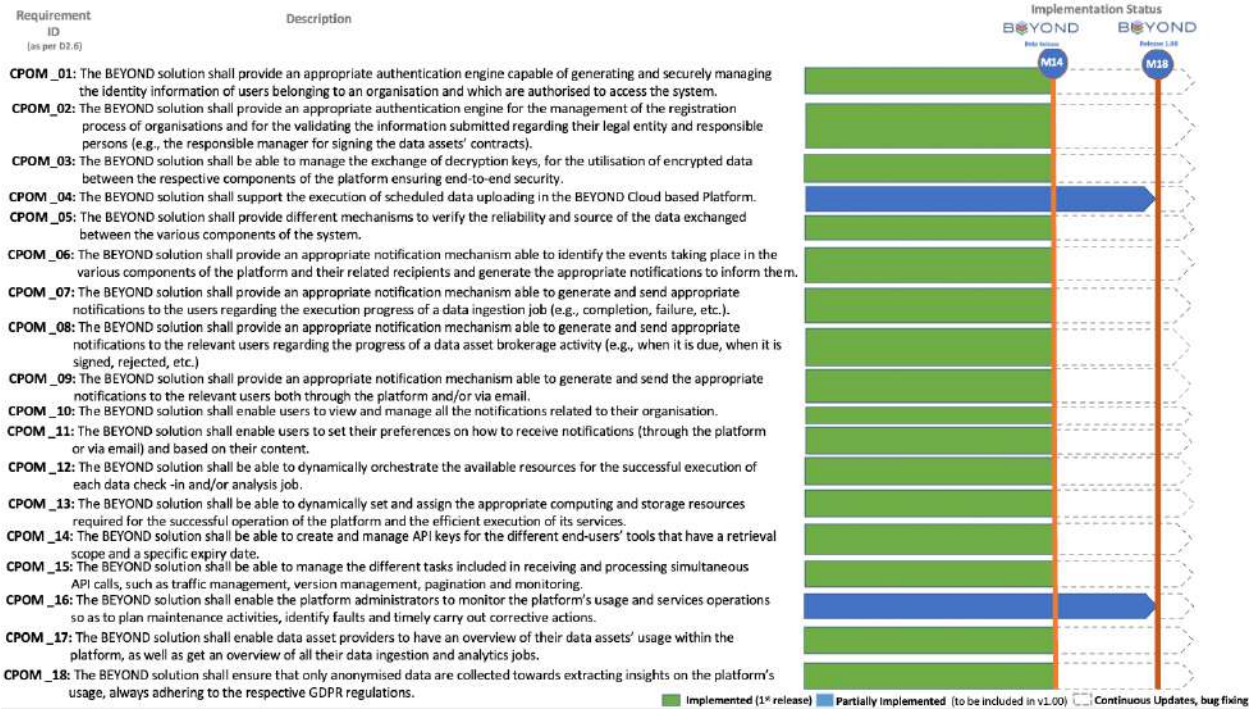

FIGURE 8 CLOUD PLATFORM OPERATIONS MANAGER - REQUIREMENTS BACKLOG AND FULFILMENT IN BETA RELEASE

As shown in Figure 9 , the beta release, the Cloud Platform Operations Manager is built on state-of-the-art technologies across three layers:

- The *Presentation Layer*, which consist of the BEYOND platform UI (currently under development and due to be released within the context of the beta version of the BEYOND Integrated Platform and described in D3.1) and that that is developed in VueJS[1] and TailwindCSS[2].
- The *Business Logic Layer - backend*, containing the different components' backend and which is built on the Nest (Node JS)[15] web framework for delivering efficient, reliable, and scalable server-side applications,
- The *Business Logic Layer* – Orchestrating layer containing the Resource Orchestrator, built on Kubernetes[16] portable platform for managing containerised workloads and services in different Kubernetes clusters, as well as on Docker for containerisation of the different services and components of the BEYOND Cloud based platform Moreover, the Server-Sent Events (SSE) technology is leveraged for enabling a client to receive automatic updates from a server via HTTP connection.

[15] https://nestjs.com/
[16] https://kubernetes.io/

- The *Data Access Layer* that essentially refers to the Polyglot Data Storage Layer (see chapter 4) that has been set up and utilises PostgreSQL[10] and Elasticsearch[11] as the main data storage mechanisms of the whole BEYOND platform.

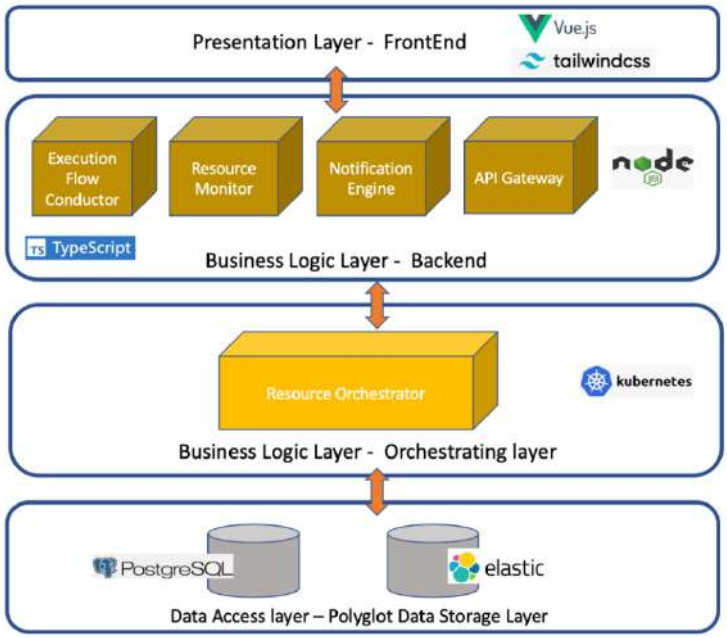

FIGURE 9 ARCHITECTURE OF THE CLOUD PLATFORM OPERATION MANAGER

It is noted that the above provided architecture is the full fledge architecture, that is envisioned to be delivered with release 2.00 of the platform. Updates to this architecture are to be considered with the delivery of the revised overall architecture of the BEYOND platform, under WP2 of the project.

6.2. Assumptions and Constraints

The following assumptions and constrains apply to the beta release:

- Users can retrieve data assets and results, through the provided API Gateway only for data assets that are already available (i.e., stored) in the BEYOND Cloud based platform.
- Users can retrieve the whole files through the BEYOND Platform APIs, since filtering cannot be applied over encrypted data.
- The BEYOND Platform's API gateway, in this beta release, generates non-expiring tokens for retrieving and uploading data assets; however, given a better understanding of the BEYOND end-users tools and platform's stakeholders (actors), such an approach may change in the future.

6.3. Licensing

The beta release of the Cloud Platform Operations Manager is available through the integrated BEYOND Cloud based platform (see D3.4) and delivered as a closed source component.

7. Next Steps towards release 1.00

In respect to the upcoming first release of the Data Collection, Security, Storage, Governance & Management Services bundles, the following pending features and extensions have been already considered/planned.

It is noted, that as shown in the various components, the features already delivered (and not part of the list below) will be subject to revisions and upgrades throughout the project, in order to improve the overall experience of users and enrich the functionalities of the platform.

TABLE 1 PENDING FUNCTIONALITIES FOR RELEASE 1.00

Component	Req. ID	Requirement Description
Data Ingestion Services	DIS_02	The BEYOND solution shall allow data providers to load streaming data to the BEYOND Cloud based Platform
	DIS_06	The BEYOND solution shall support the execution of scheduled collection of streaming data.
	DIS_17	The BEYOND solution shall support the inclusion of various metadata on the ingested data, to improve asset discoverability, and reveal relationships that might exist with other assets.
	DIS_18	The BEYOND solution shall enable the fusion of different assets into a composite asset, to create new data profiles that carry more valuable information
Polyglot Data Storage Layer	PDSL_03	The BEYOND solution shall enable storage of the data assets' metadata in a trusted data container
	PDSL_10	The BEYOND solution shall provide a dedicated distributed ledger for the secure storage of all contract-related data, while ensuring the privacy of the involved stakeholders.
Cloud Platform Operations Manager	CPOM _04	The BEYOND solution shall support the execution of scheduled data uploading in the BEYOND Cloud based Platform.
	CPOM_16	The BEYOND solution shall enable the platform administrators to monitor the platform's usage and services operations so as to plan maintenance activities, identify faults and timely carry out corrective actions.
Data Security	SEC_06:	The BEYOND Security Components shall provide end-to-end encryption of uploaded data based on hybrid techniques (Attribute-Based Symmetric Searchable Encryption).
	SEC_07	The BEYOND Security Components shall enable authorised users only, to retrieve information based on their data contract rules.
	SEC_08	The BEYOND Security Components shall enable authorised users only, to acquire decrypted information based on their data contract rules.
	SEC_09	The BEYOND Security Components shall enable authorised users only, to analyse/process information based on their data contract rules.
	SEC_10	The BEYOND Security Components shall allow the data owners to pseudo anonymise or anonymise their data.

services bundle	SEC_11	The BEYOND Security Components shall allow the data owners to define their own data anonymisation rules.
	SEC_12	The BEYOND Security Components shall not allow the linking of a specific data set or data action with a specific actor.
	SEC_13	The BEYOND Security Components shall support attribute-based access control models/policies (ABAC).
	SEC_14	The BEYOND Security Components shall provide attribute-based configuration/update on the configuration on the smart contracts signed between producer-platform and consumer-platform.
	SEC_15	The BEYOND Security Components shall link the uploaded/analysed data to certain access control policies.
	SEC_16	The BEYOND Security Components shall allow the data owners to define their own data access control policies at user/group level.
	SEC_17	The BEYOND Security Components shall allow the data consumers to process the data according to the already specified policies stored in the Access Policy Engine.
	SEC_18	The BEYOND Security Components shall not allow access/processing on data that have been identified as personal.

References

[1] BEYOND (2020) Description of Action (DoA)
[2] BEYOND (2021a): D2.1 - End-user & Business requirements analysis for big data-driven innovative energy services & ecosystems –
[3] BEYOND (2021b): D2.6 – BEYOND Framework Architecture including functional, technical and communication specifications - a