ACCURATE

D5.2 Compute-to-Data Environment: a secure Compute-to-Data (C2D) environment architecture integrated into the data space

Actual Submission Date:	31/01/2025
Produced by:	DAO: deltaDAO AG

Accurate

https://accurateproject.eu/

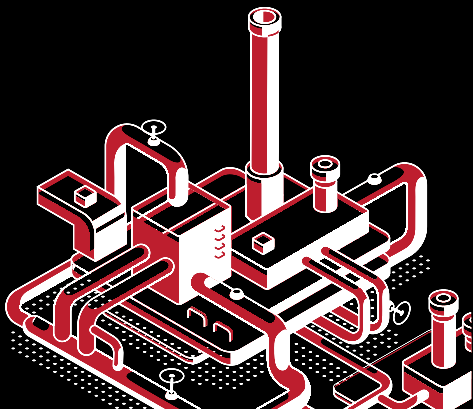
HORIZON-CL4-2023-TWIN-TRANSITION-01

Grant Agreement no.: 101138269

Start date of project: 01 12 2023 - Duration: 36 months

Funded by
the European Union

The ACCURATE project is funded by the European Union, under Grant Agreement number 101138269. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.



DELIVERABLE FACTSHEET

Deliverable D5.2	
Nature of the Deliverable:	Demonstrator
Due date of the Deliverable:	M14 – 31/01/2025
Actual Submission Date:	M14 – 31/01-2025
Produced by:	DAO: deltaDAO AG
Contributors:	
Work Package Leader Responsible:	DAO: deltaDAO AG
Reviewed by:	Shortname: SIMAVI, ES

Dissemination level	
	PU = Public
	PP = Restricted to other programme participants (including the EC)
	RE = Restricted to a group of the consortium (including the EC)
CO	CO = Confidential, only members of the consortium (including the EC)

Contents

Figures

Terms and abbreviations

C2D	Compute-to-Data
Compute Job	A process executing computations on private data without exposing it.
DDO	Decentralized Data Object
DID	Decentralized Identifier
FOSS	Free Open-Source Software
GDPR	General Data Protection Regulation
HTTP	Hypertext Transfer Protocol
IoT	Internet of Things
IPFS	Interplanetary File System
MaaS	Manufacturing as a Service
OEM	Original Equipment Manufacturer
SME	Small and Medium Sized Enterprise
SSI	Self-Sovereign Identity

Public Summary

The ACCURATE project [1] advances European manufacturing by enabling secure, data-driven collaboration across complex and dynamic value chains. A key enabler is the Compute-to-Data (C2D) environment, which allows sensitive manufacturing data to remain in its original location, or within a trusted infrastructure under the full control of the data owner, while still supporting computational result sharing, distributed analytics, and federated learning. [2]

Instead of transferring raw data—creating copies and potentially exposing sensitive information or risking breaches—C2D brings portable applications to the data. Computations are executed on-site, and only aggregated or anonymized results are returned to the requester. This approach preserves data sovereignty, ensures compliance with European data protection standards, and safeguards intellectual property. C2D aligns with ACCURATE's goals of improving supply chain resilience, sustainability, and competitiveness, as it allows participants to make previously unavailable data accessible, share insights and collaborate without compromising confidentiality or incurring unnecessary risk or data movements. [3]

By integrating C2D with tokenized smart-contract-based access and containerized software components, ACCURATE enables new kinds of data products and services. ACCURATE paves the way for an improved Manufacturing as a Service (MaaS) ecosystem. Manufacturers and Service providers can securely publish manufacturing services, software, and digital twins as services, while consumers can interact with these offerings to discover, evaluate, and select manufacturing partners or configure production processes—all without revealing sensitive operational data. This capability enhances flexibility, fosters dynamic supply chain relationships, and ensures that advanced analytics and optimization remain fully compatible with data sovereignty and IP protection.

In essence, C2D supports ACCURATE's vision: a federated, trusted and innovation-friendly European manufacturing ecosystem where data-driven collaboration boosts resilience, adaptability, and long-term strategic advantage.

1 Introduction

1.1 About this Deliverable

This deliverable documents the conceptual design, technical architecture, and integration of the C2D capabilities into the ACCURATE data space, as developed under Work Package 5 (Data space design and implementation). It details the key components, workflows, and services that enable secure computations on sensitive datasets, as well as the mechanisms for publishing, discovering, and consuming data and software assets. Additionally, it provides guidance on how to leverage C2D for various use cases, including federated analytics, resource matchmaking, and sustainability assessments within the manufacturing sector.

1.2 Document Structure

This deliverable is structured as follows:

Chapter 1: Introduction
Provides the context, objectives, and scope of this deliverable, as well as its relation to the ACCURATE project's broader goals and associated deliverables.

Chapter 2: A Secure Compute-to-Data Environment Architecture Integrated into the Data Space
This central chapter focuses on the technical design and implementation aspects of the C2D environment, aligning with the objectives of WP5 and the overall architecture of the data space developed in T5.2. It covers the following sections:

- 2.1 Overview: Introduces the C2D concept, its relevance to ACCURATE, and an outline of the key requirements and guiding principles.
- 2.2 Architecture: Details the decentralized C2D architecture, including its components (Operator-Service, Operator-Engine, Kubernetes, etc.) and workflows required to enable secure computations.
- 2.3 Data and Software Integration: Explains how C2D supports the seamless integration of datasets and containerized software services.
- 2.4 Publication and Consumption Flow for Data and Software Services: Describes how participants can publish datasets, algorithms, and SaaS offerings, and how consumers can discover, access, and use them.
- 2.6 Compute-to-Data Enabling Manufacturing as a Service (MaaS): Demonstrates how the C2D architecture enables MaaS scenarios by facilitating secure, federated analytics and federated learning across multiple stakeholders within a manufacturing ecosystem.

Chapter 3: Conclusion
Summarizes the key findings, highlights the benefits of the C2D approach, and outlines next steps.

1.3 Relation with Other Tasks and Deliverables

This deliverable supports Work Package 5 (WP5), which focuses on designing and implementing a decentralized data space that enables secure and privacy-preserving data sharing and usage. It aligns with the outcomes of earlier tasks and sets the stage for subsequent ones:

- Task 5.1 (Data spaces requirements engineering):
  Informed the initial requirements for data sovereignty, semantic models, and sustainability attributes, ensuring that the C2D solution meets identified ecosystem needs.
- Task 5.2 (Decentralized architecture and data technologies):
  Provided the initial data space architecture. This deliverable (D5.2) seamlessly integrates C2D into that architecture, which was thoroughly described in D5.1 Data Space Design and Architecture.
- Task 5.3 (Enabling Technical Sovereign Data Usage):
  Directly addressed here, the pilot partners will be supported among others to setup their own C2D environment to enable computations on sensitive data without moving it, fostering trust, technical data sovereignty, IP protection, and federated analytics.
- Task 5.4 (System integration and decentralization) and Task 5.5 (Testing and validation):
  C2D features will be incorporated into broader system integration efforts and thoroughly tested, ensuring scalability, robustness, and user acceptance.

Beyond WP5, this deliverable provides critical input for subsequent project deliverables and related efforts, notably in producing advanced data-sharing and MaaS solutions. D5.2's C2D framework is essential for:

- Sovereign Data Sharing (D5.3)
- High-level Ecosystem Architecture (D6.1)
- Digital Twin, Decision-Support System, and MaaS Solutions – First and Final Versions (D6.2 and D6.3)
- ACCURATE Framework – Initial and Final Versions (D6.4 and D6.5)

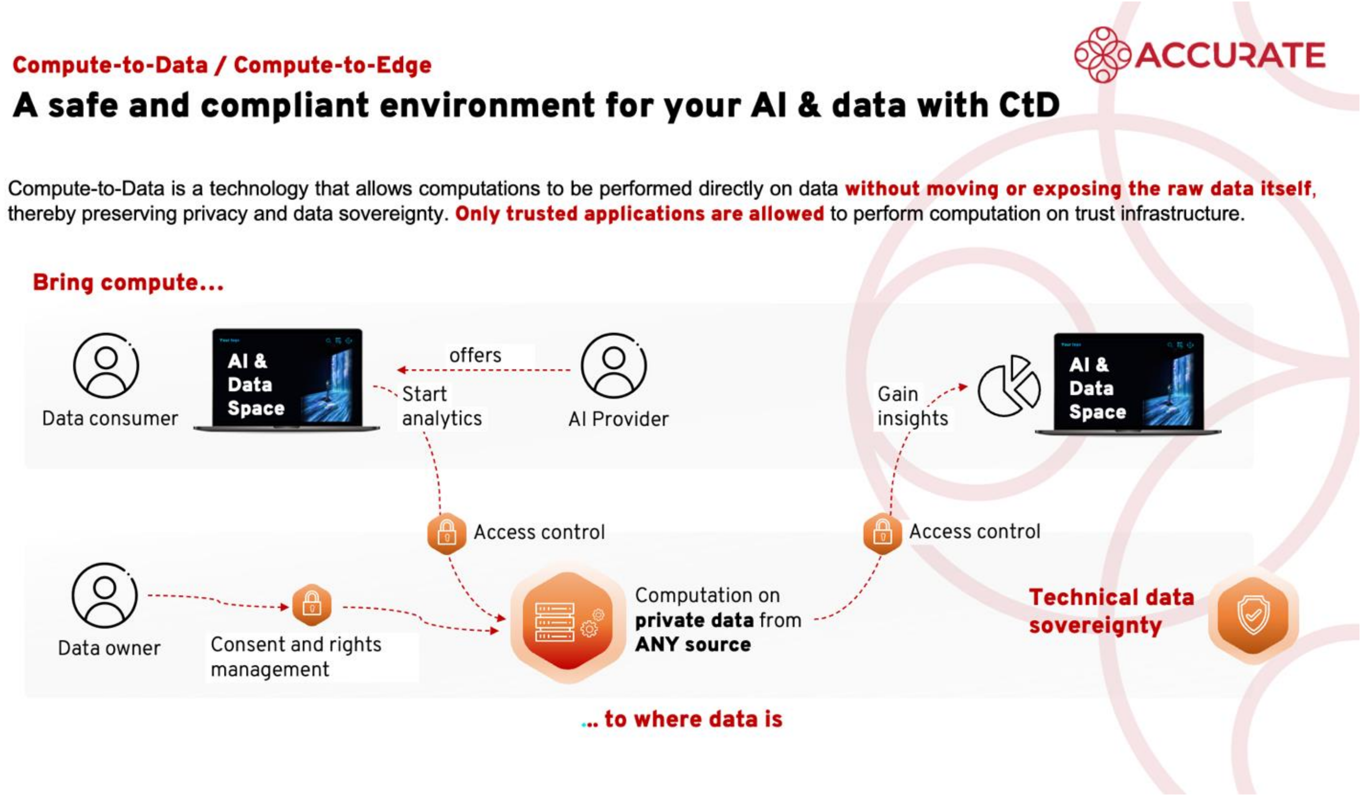
2 Compute-to-Data

2.1 Introduction to Compute-to-Data

Compute-to-Data (C2D) is a privacy and IP preserving computing paradigm designed to enable secure and controlled access to sensitive data with trusted portable applications. Compute-to-Data allows portable applications to be transferred to infrastructures controlled and/or trusted by the data owner, providing a safe room for the data to be processed. This avoids the creation of copies of data and data leakage outside the technical control of the data owners, and thus significantly reduces the trust required in other participants and their infrastructures. Compared to other contractual arrangements between participants, Compute-to-Data enables the technical enforcement of consent and other use-case specific restrictions on data use. [4]

In the ACCURATE data space, the Compute-to-Data feature is based on the free open-source software of the Ocean Enterprise framework and, in addition to traditional data transfer between participants, enables the publication of software services and the combination of these with data services to create new data products. [5]

The raw data does not need to be disclosed when orchestrating these compute jobs, because computation can take place in environments under control of the data owner. Furthermore, the Compute-to-Data mechanism can be combined with all other privacy-preserving or privacy-enhancing methods available, such as aggregation, anonymization, synthetic data generation, encryption, etc. As a key enabler of the ACCURATE project’s vision, C2D allows manufacturing stakeholders to collaborate on data-driven insights while fully retaining data sovereignty, ensuring compliance with European data protection regulations, and safeguarding intellectual property (IP). By leveraging Kubernetes for infrastructure-agnostic deployment, C2D seamlessly integrates with existing data environments, from on-premises installations to cloud and edge solutions, making it a highly scalable and adaptable approach for diverse manufacturing settings. [76]



Compute-to-Data / Compute-to-Edge

A safe and compliant environment for your AI & data with CtD

Compute-to-Data is a technology that allows computations to be performed directly on data **without moving or exposing the raw data itself**, thereby preserving privacy and data sovereignty. **Only trusted applications are allowed** to perform computation on trust infrastructure.

Figure 1: Compute-to-Data Overview (Source: deltaDAO)

C2D in the ACCURATE Context

In the ACCURATE ecosystem, which fosters the creation of resilient, sustainability-oriented Manufacturing as a Service (MAAS) networks, C2D provides the foundation for trusted data exchange, leveraging digital twins, building new data products and collaborative innovation. It enables participants—such as manufacturers, logistics providers, and service integrators—to perform computations, analytics, and even train machine learning models directly on proprietary datasets without exposing raw data, i.e. sensitive data about inventory, customer relations, current machine utilization rates, etc.. This approach is crucial for addressing the complexities of modern manufacturing value chains, where disruptions can arise unexpectedly, and operational resilience is paramount. C2D here aims to increase data availability and lead to more effective decisions, as data can now be made accessible without directly exposing it to new risks.

Key Features and Benefits:

1. **Secure Data Processing:**
   - Executes computations on-site, preserving raw data privacy.
   - Eliminates the need to transfer sensitive datasets, reducing transmission costs and security risks.
   - Allows data owners to retain full technical control over their data assets, granting or revoking compute access as needed.
   - Can be combined with all other forms of privacy-preserving mechanisms.
2. **Data Sovereignty:**
   - Ensures data remains under the data provider's governance and control.
   - Complies with stringent data protection regulations, a necessity in manufacturing domains where confidentiality and IP protection are critical.
   - Allows technical and automatic enforcement of terms of usage and data usage agreements.
3. **Building New Data Products:**
   - Compute-to-Data enables the publication of containerized software services and the efficient orchestration of compute jobs, fostering the creation of entirely new data products.
   - Allows to perform computations, analytics, and even train machine learning models directly on proprietary datasets without exposing raw data.
   - Allows the modular creation of new data products while retaining the right to exploit the core IP as a renewable resource through pay-per-use business models.
4. **Collaborative Innovation:**
   - Facilitates secure cross-organizational scenarios, including federated learning and collaborative analytics, without revealing confidential details.
   - Fosters new services, joint R&D, and scenario-based simulations for stress-testing supply chain resilience.
5. **Flexibility and Scalability:**
   - Built atop Kubernetes for adaptability in various IT environments—cloud, on-premises, edge, and IoT deployments.

- Scales effortlessly to support growing data volumes, complex computations, and an expanding network of ecosystem participants.

6. **Integrated Audit Trail:**
    - Employs smart contracts and blockchain integration to create immutable, tamper-proof audit trails of data usage, licensing, and service events.
7. **Manufacturing-Specific Applications:**
    - Enables the secure analysis of proprietary manufacturing processes and operational metrics.
    - Supports sustainability and circularity assessments by integrating resilience-focused KPIs and environmental indicators.
    - Encourages benchmarking and optimization efforts, enabling informed decision-making for competitive advantages.

Aligning C2D with ACCURATE's Ambitions:
C2D seamlessly fits into the ACCURATE project's broader objectives of enhancing European manufacturing competitiveness, improving sustainability, and achieving operational resilience against disruptions. By securely enabling data-driven planning, robust design strategies, federated learning, and scenario exploration, C2D helps ensure stable, future-proof MAAS value chains. Furthermore, as ACCURATE incorporates Gaia-X and industry standards, C2D ensures interoperability and trust, facilitating integration with prominent manufacturing and industry 4.0 data spaces (e.g., EuProGigant, COOPERANTS, DIONE-X, Flex4Res).

A Phased, Tailored Approach:
While the foundation of C2D is robust, its implementation within ACCURATE will be iteratively refined and validated in alignment with Task 5.3 and other related work packages. This phased approach ensures that the C2D architecture, policies, and workflows are specifically tailored to the unique challenges of the ACCURATE project and of European manufacturing contexts. [45]

2.2 C2D Architecture

This section provides a detailed view of the Compute-to-Data (C2D) architecture employed within the ACCURATE data space. The architecture builds upon the Ocean Enterprise framework [5], integrating seamlessly into the ACCURATE project's environment while maintaining data sovereignty, privacy, and security. By running computations directly on the data owner's premises, C2D ensures that sensitive manufacturing data and intellectual property remain protected throughout the entire process.

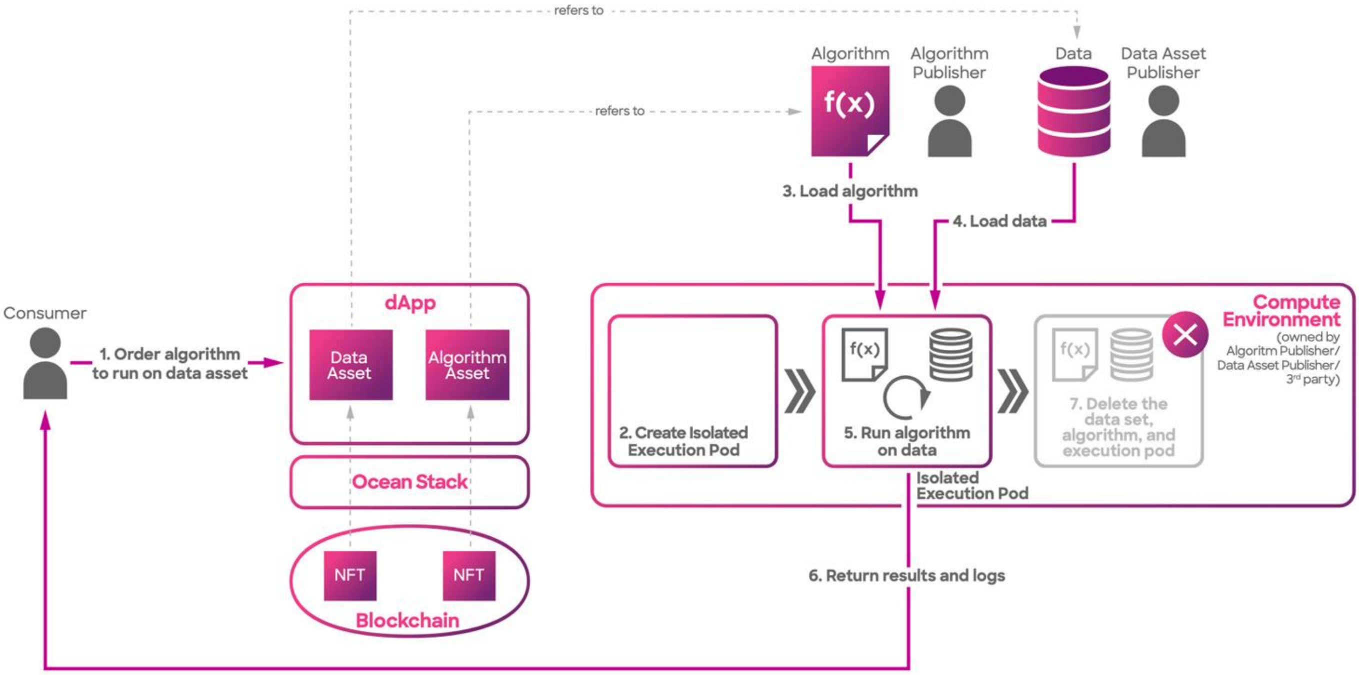


Figure 2: Compute Architecture Overview (Source: Ocean Protocol Docs)

The C2D workflow involves a series of steps that ensure secure, privacy-preserving computation on sensitive datasets:

1. The Service Consumer initiates a C2D job by selecting a desired data asset (identified by a DID) and an algorithm (identified by a DID) from the catalogue. The validity of orders (e.g., access permissions, payments) is verified through a decentralized application (dApp) and the access controller. [9]

2. A dedicated and isolated execution pod is created in the Kubernetes (K8s) environment for this specific compute job.

3. The execution pod loads the specified algorithm into its environment.

4. The selected dataset is loaded into the execution pod without exposing the raw data outside of the data owners C2D infrastructure.

5. The algorithm runs on the dataset inside the isolated execution pod.

6. The results and logs generated by the algorithm are stored in a dedicated secure storage next to the C2D infrastructure.

7. The execution pod deletes the dataset, algorithm, and itself, ensuring no data residue remains and thus maintaining strict data privacy.

8. A consumer can request and retrieve the results, after the identity has been confirmed to be identical to the party that requested the C2D job.

Compute Architecture Overview

The interaction between the Service Consumer and the Access Controller (Ocean Provider) component follows a structured flow:

- To start the process, the Consumer calls the start (did, algorithm, additionalDIDs) function on the Ocean Provider, specifying parameters such as the data identifier (DID), algorithm, and any additional DIDs if needed.
- The Ocean Provider responds by generating a unique job identifier (e.g., XXXX) and returns it to the Consumer. From this point onward, the Ocean Provider oversees the rest of the compute process.
- The Consumer can monitor progress by querying the Access Controller with getJobDetails(XXXX), where XXXX is the job identifier.
- It is possible to initiate a compute job using one or multiple data assets. Nautilus can be utilized to explore such functionalities.

Ocean Provider's Inner Workings

Initially, the Ocean Provider [13] verifies whether the Consumer has supplied the correct data tokens required to access the requested data and C2D process. Once confirmed, the Ocean Provider delegates the job to the Operator-Service, a dedicated microservice responsible for coordinating the execution.

The Operator-Service forwards the request to the Operator-Engine, the component that handles the actual compute operations (such as running Kubernetes jobs). The Operator-Engine executes the computations, continuously informing the Operator-Service about the job status. When completed, the Operator-Engine signals the Operator-Service, and in turn, the Ocean Provider is notified that the job finished successfully.

Actors/Components:

- **Consumers:** Organizations /institutions, that require access to computing services from the data Publisher's environment.
- **Ocean Provider**: Proxy and access controller granting access to specific data resources on behalf of the service provider. The Ocean provider, as data space connector, will only grant access if all conditions, as specified in the service metadata, are satisfied.
- **Operator-Service:** A microservice responsible for managing compute requests, interacting with the infrastructure using the Publisher's credentials, and providing APIs to run, stop, or retrieve the status of compute jobs.
- **Operator-Engine:** The system executing the compute jobs on top of Kubernetes. It orchestrates all stages of the job's lifecycle, from fetching data and algorithms to running computations and publishing results.

- **Kubernetes (K8s) Cluster:** The infrastructure layer where isolated pods run computations securely without exposing raw data to external environments.
- **Storage:** A result storage is responsible for storing compute results. Access to these results is given through the Ocean Provider, acting as access controller between Consumer and Provider.

Pre-Conditions:

Before initiating a compute job, the following conditions must be met:

- The Asset's Decentralized Document Object (DDO) includes a compute service [11].
- The Asset's DDO compute service explicitly permits algorithms to run on it.
- The Asset's DDO specifies an Ocean Provider endpoint exposed by the Publisher [12].
- The consumer can access the service in accordance with the policies set by the data Publisher.

Access Control using Ocean Provider

The compute service uses the Ocean Provider—the access controller component maintained by the data Publishers—to handle user interactions and facilitate secure integration with the ACCURATE data space. The Ocean Provider possesses the necessary credentials to communicate securely with the underlying infrastructure, which may be cloud-based or on-premises. This ensures that Publishers can leverage the compute service while retaining full control over their data and infrastructure. [13]

Operator Service

The Operator Service is a microservice charged with managing compute workflows and executing requests. Its main responsibilities include:

- Providing an HTTP API for data access and compute endpoints.
- Interacting with the infrastructure, using the Publisher's credentials, to initiate and manage compute resources.
- Starting, stopping, and executing computing instances with the user-provided algorithms.
- Retrieving logs generated during execution for auditing and debugging.

Typically, the Operator Service is integrated with the Ocean Provider. It establishes communication with the Kubernetes cluster, allowing it to:

- Register new compute jobs.
- List currently running jobs.
- Retrieve detailed results for specific jobs.
- Stop ongoing jobs as needed.

The Operator Service does not store any data internally. All state information is managed directly within the K8s cluster. [6]

Operator Engine

The Operator Engine orchestrates the compute infrastructure using Kubernetes as the backend. Each compute job runs in its own Kubernetes Pod, ensuring isolation and security. By retrieving workflows from the Operator Service, the Operator Engine manages all the necessary infrastructure resources to complete compute workflows.

The Operator Engine is responsible for:

- Orchestrating the execution flow.
- Initiating a configuration pod to download workflow dependencies, including datasets and algorithms.
- Starting the algorithm pod to execute computations.
- Launching the publishing pod to register newly created assets on the Ocean Protocol network.

Like the Operator Service, the Operator Engine does not store data internally; all states are kept in the K8s cluster.

Pod Configuration

The Pod-Configuration process is crucial for preparing the environment before a job begins. A node.js script dynamically manages the initialization, fetching datasets, algorithms, and necessary Decentralized Document Oriented Storage (DDOS) elements. By placing these assets in designated directories (e.g., /data/inputs/DID/ for datasets and /data/transformations/ for algorithms) and handling indexing and file organization, the Pod-Configuration ensures that the environment is correctly set up for execution.

In case of provisioning errors, the script updates job status in a PostgreSQL database and logs error messages. Once successfully completed, it signals the Operator Engine to start the algorithm pod, ensuring a seamless progression of the workflow.

Pod Publishing

Pod Publishing acts as a command-line utility that integrates with the Operator Service and Operator Engine to manage workflow outputs. Within the Kubernetes-based compute infrastructure, it facilitates the processing, logging, and uploading of results to external storage solutions such as AWS S3 or the Interplanetary File System (IPFS). It also updates a PostgreSQL database with logs and status information.

Pod Publishing plays a pivotal role in the publishing pod, which registers new assets on the Ocean Protocol network after the successful completion of a workflow. It does not provide storage capabilities itself; instead, it relies on K8s cluster states or external storage solutions. [6]

2.2.1 Detailed Compute-to-Data Workflow

Below is a detailed step-by-step outline of the Compute-to-Data (C2D) workflow, illustrating the entire process—from initiating a compute job to retrieving results.

Starting a C2D Job

1. The consumer selects a preferred computing environment from the Ocean provider's catalog and chooses a specific dataset-algorithm pair.
2. The Ocean provider verifies the orders on the blockchain.
3. If the dataset, algorithm, and compute environment fee orders are valid, the provider initiates the compute flow.
4. The provider notifies the consumer that the job has been successfully created, providing a unique job ID.
5. With the confirmed job ID and orders, the provider calls the operator service to start the job.
6. The operator service registers the new job in its local job queue.
7. The operator engine periodically checks the operator service's job queue. If resources are available, it requests a list of pending jobs to determine which one to launch.
8. The operator service returns the list of jobs, and the operator engine prepares to begin executing a selected job.

Creating the K8s Cluster and Allocating Job Volumes

9. When the job starts, the operator engine provisions storage volumes on the Kubernetes (K8s) cluster.
10. The K8s cluster creates and allocates the necessary volumes for the job.
11. The newly created volumes are allocated to a dedicated pod for the job.
12. After volume allocation, the operator engine initiates the "pod-configuration" phase by launching a new pod to handle pre-computation setup.

Loading Datasets and Algorithms

13. The pod-configuration process requests required datasets and the algorithm from their respective providers.
14. The provider supplies these files, which are downloaded into the cluster.
15. The downloaded datasets and algorithm files are written into the job's allocated volume.
16. Once all inputs are in place, the pod-configuration component notifies the operator engine that the environment is ready for computation.

Running the Algorithm on Dataset(s)

17. The operator engine deploys the algorithm pod onto the K8s cluster, mounting the prepared datasets and algorithm files.
18. Kubernetes runs the algorithm pod, executing the computation directly on the data.
19. The operator engine monitors the execution, enforcing any time limits or constraints specified by the selected environment.
20. After the algorithm completes and results are generated, the operator engine starts the "pod-publishing" stage.
21. Pod-publishing uploads the results, logs, and administrative records into the output volume.
22. Once the publishing phase finishes, the operator engine receives a notification, enabling it to proceed with cleanup operations.

Cleaning Up Volumes and Allocated Space

23. The operator engine deletes the K8s volumes associated with the job.
24. The Kubernetes cluster removes all used volumes, releasing the storage resources.
25. With all temporary resources cleared, the operator engine finalizes the job's lifecycle.
26. The operator engine informs the operator service that the job has concluded, and the results are ready.

Retrieving Job Details

27. The consumer requests job details from the provider using the get job details function.
28. The provider queries the operator service for the requested job details.
29. The operator service returns the job details to the provider.
30. The provider shares these job details with the dataset consumer, offering insights into the completed computation.

Retrieving Job Results

31. Armed with the job details, the consumer requests the final computation results from the provider.
32. The provider consults the operator engine to access the stored results.
33. Since the operator service cannot directly access the results, it relies on the output volume to retrieve them.
34. The output volume provides the results to the operator service.
35. The operator service then passes these results back to the provider.
36. Finally, the provider delivers the results to the dataset consumer, completing the C2D workflow. [8]

The C2D workflow is visualized in detail below:

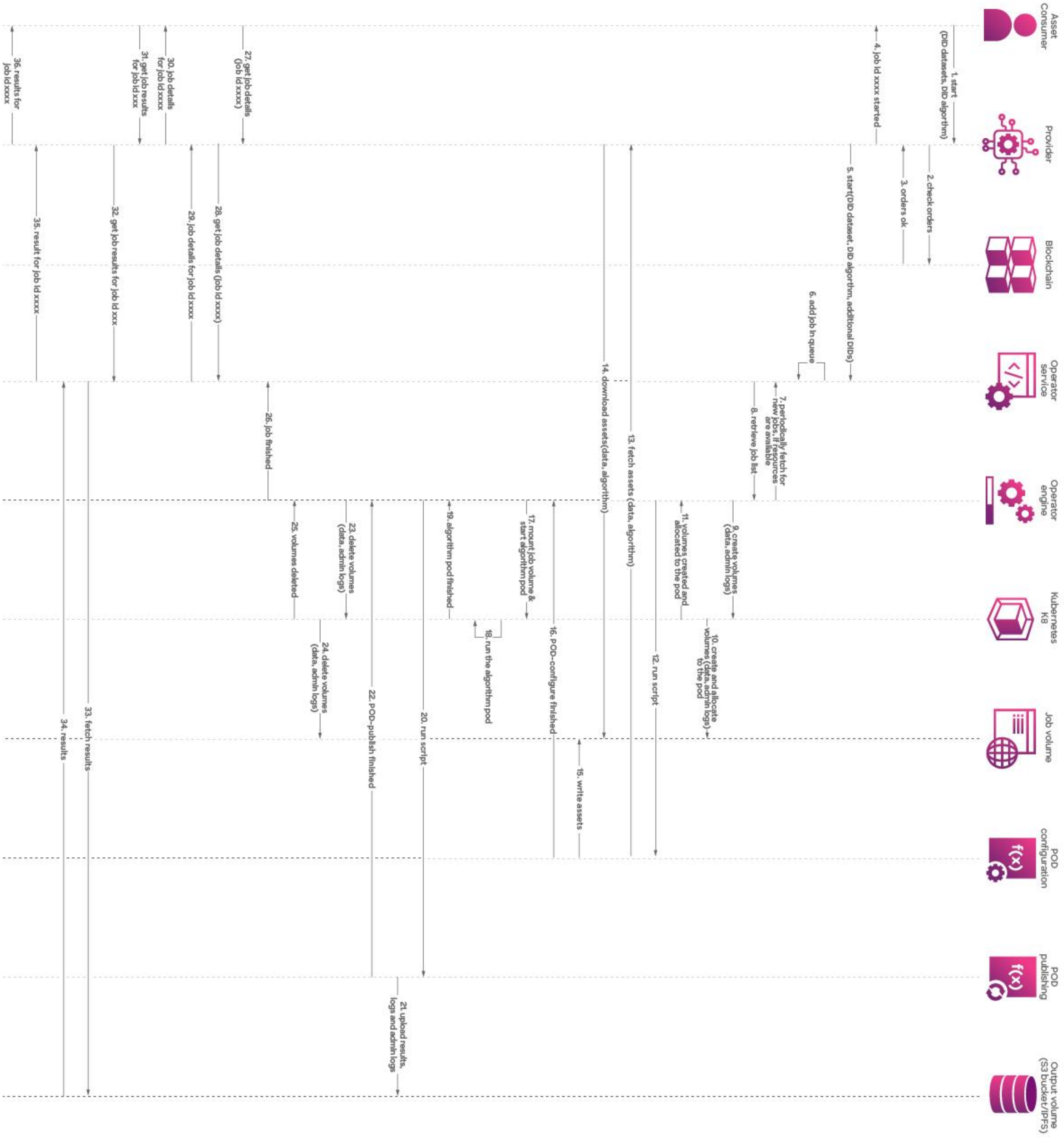

Asset Consumer
Provider
Blockchain
Operator service
Operator engine
Kubernetes K8
Job volume
POD configuration
POD publishing
Output volume (S3 bucket/IPFS)
1. start (DID datasets, DID algorthm)
2. check orders
3. orders ok
4. job id xxxx started
5. start(DID dataset, DID algorthm, additional DIDs)
6. add job in queue
7. periodically fetch for new jobs, if resources are available
8. retrieve job list
9. create volumes (data, admin logs)
10. create and allocate volumes (data, admin logs) to the pod
11. volumes created and allocated to the pod
12. run script
13. fetch assets (data, algorithm)
14. download assets(data, algorithm)
15. write assets
16. POD-configure finished
17. mount job volume & start algorithm pod
18. run the algorithm pod
19. algorithm pod finished
20. run script
21. upload results, logs and admin logs
22. POD-publish finished
23. delete volumes (data, admin logs)
24. delete volumes (data, admin logs)
25. volumes deleted
26. job finished
27. get job details (job id xxxx)
28. get job details (job id xxxx)
29. job details for job id xxxx
30. job details for job id xxxx
31. get job results for job id xxx
32. get job results for job id xxx
33. fetch results
34. results
35. result for job id xxxx
36. results for job id xxxx

Figure 3: Detailed C2D workflow (Source: Ocean Protocol Docs)

Summary

The C2D architecture in ACCURATE relies on a carefully orchestrated interplay of components—, Ocean Provider, Operator-Service, Operator-Engine, and Kubernetes—to securely execute computations on sensitive data. By leveraging pods for configuration, execution, and publishing tasks, the system ensures data privacy, integrity, and auditability, paving the way for compliant and collaborative data-driven innovation in the European manufacturing sector.

2.3 Data and Software Integration

Compute-to-Data (C2D) extends beyond traditional data sharing models by enabling participants in the ACCURATE ecosystem to publish not only datasets but also containerized software assets. This approach empowers organizations to securely integrate custom algorithms, simulation tools, AI models, and other specialized manufacturing applications into the data space, all without revealing underlying intellectual property or proprietary insights.

By leveraging C2D, data publishers can retain full sovereignty and control over their sensitive information while still offering value-added services that enhance supply chain resilience, production optimization, and collaborative innovation. In this context, both datasets and software assets become configurable, discoverable, and composable building blocks within the ACCURATE data space, supporting a wide range of integrated use cases tailored to the manufacturing sector. [7]

Key aspects of data and software integration include:

1. **Compatibility:**
   - Facilitates integration of standard manufacturing services and applications, including Manufacturing as a Service (MaaS), Production Data Sharing, Inventory Sharing, Product Information Sharing, Product Carbon Footprint Calculation, Asset Administration Shell (AAS) Integration, and Specification Document Exchange.
   - Accommodates various data types and formats—such as IoT sensor outputs, simulation model inputs, CAD files, production schedules, and quality control metrics—thereby enabling rich and versatile data-driven workflows.
   - Encourages the use of data analysis software, AI-driven models, and machine learning algorithms specifically designed for manufacturing process simulations and supply chain optimizations.
2. **Flexible Deployment:**

   - Supports deployment across cloud, on-premises, and edge environments to align with the infrastructural realities of diverse manufacturing operations.
   - Provides interoperability within hybrid infrastructures common in the industry, ensuring smooth transitions between different computational environments.
   - Allows consumer to perform ad hoc data production generation without the need to deploy own compute-environments.

3. Multi-Tier Visibility and Collaboration:
   - Facilitates seamless data and software service exchanges across complex, multi-tier manufacturing supply chains, including OEMs and multiple tiers of suppliers.
   - Enhances the ACCURATE project’s goal of fostering robust collaboration and improved visibility in environments characterized by numerous interconnected stakeholders.
4. Versatile Data and Service Management:
   - Empowers participants to publish, discover, and consume data sources via multiple protocols (URL, REST APIs, GraphQL, IPFS) and to incorporate containerized software services such as Docker-based applications directly into the data space.
   - Enables future orchestration of data and compute services to form new data products, workflows, and value-added services, leveraging C2D for secure, on-demand computations.
   - Allows the seamless combination of data and software assets—ranging from simulation models and data processing algorithms to proprietary AI engines—enabling participants to build integrated solutions that address a broad array of manufacturing challenges.
5. Flexible Data Hosting:
   - Supports data residency requirements by allowing data and software to remain at their original location—in the cloud, on-premises, or at the edge—integrating with existing enterprise infrastructures without disrupting established workflows or investments.
   - Preserves the ability of data holders to maintain strict control over their assets while still benefiting from secure and privacy-preserving collaborative computations.

By enabling both data and software to be published and consumed within the C2D framework, the ACCURATE project paves the way for a truly modular, composable, and secure ecosystem. This holistic integration approach ensures that manufacturing participants can leverage their existing data and tools, enrich them with cutting-edge software services, and collaboratively innovate within a trusted and compliant environment.

2.4 Publication and Consumption Flow for Data and Software Services

The ACCURATE data space provides a structured approach for publishing, discovering, and consuming both datasets and software assets (e.g., algorithms, containerized services, and SaaS offerings) via user interfaces, e.g., the ACCURATE marketplace, or programmatic tooling, e.g., nautilus. By leveraging Compute-to-Data (C2D), tokenization, and fine-grained access controls, participants can confidently share and utilize high-value data and software services without compromising privacy, data sovereignty, or intellectual property. This framework enables a secure, collaborative environment where manufacturing stakeholders can monetize their data and software assets, optimize supply chain operations, and foster innovation.

1. Connecting to the Marketplace

Participants, who want to use a user interface, start by connecting their identity wallets to the ACCURATE Marketplace [16]. This step provides them with decentralized access to publishing and consuming data and software assets. Through this connection, participants manage ownership, permissions, and financial transactions using secure distributed leger-based mechanisms.

2. Publishing Data and Software Assets

Participants can publish three main types of assets: Datasets, Algorithms, and Software as a Service (SaaS). Each publication process involves defining metadata, setting access controls, and establishing pricing models aligned with project objectives and Gaia-X guidelines.

- Metadata Definition:
  - o Asset Type: Selecting between Dataset, Algorithm, or SaaS.
  - o Title and Description: Supplying meaningful, descriptive information.
  - o Tags and Gaia-X Credentials (optional): Enhancing discoverability and compliance within the ACCURATE data ecosystem.

3. Docker Container Support for Algorithms

For algorithmic services—such as data models, simulation tools, or data analysis methods—the ACCURATE Marketplace supports the publication of Docker containers. By packaging algorithms as containerized solutions, participants ensure consistent and portable execution environments. When used in C2D scenarios, Docker containers can run computations directly on the data publishers' premises without exposing raw data, preserving confidentiality, and protecting sensitive manufacturing insights.

- Algorithm Asset Definition: Specifying the Docker image in the asset metadata, ensuring the C2D environment can securely retrieve and execute the algorithm.

4. Policies and Access Controls

Flexible policies and access controls enable asset publishers to define how their published data and software services can be consumed:

- Download Access: Allow users to download datasets or software for local use.

- **Compute Access (C2D):** Permit algorithms to run on data where the data never leaves its source.
- **Advanced Controls:** Configure allowlists, deny-lists, and user-specific parameters to ensure only authorized participants can access sensitive assets.

5. Software as a Service (SaaS) Integration

In addition to datasets and containerized algorithms, participants may publish SaaS offerings—external cloud-based services accessible via URLs. These services integrate seamlessly into the ACCURATE data space, governed by decentralized contracting and monetization. SaaS offerings do not require containerization; instead, they can link to third-party platforms or interfaces, broadening the range of available services. A subscription verifier component integrated into the external cloud service can enforce policies set in the ACCURATE data space.

6. Pricing Information and Tokenized Access

Participants define pricing models using Data Tokens, denominated in EURO or offered free of charge. These tokens, compliant with ERC-20 standards, enable transparent and automated settlements underpinned by smart contracts. This approach simplifies payment flows, ensures traceability, and can be adapted for testing and development environments using test tokens.

7. Consumption Flow

The ACCURATE Marketplace's catalog allows participants to discover, evaluate, and access published assets—datasets, algorithms, and SaaS offerings:

- **Discover Assets:** Use metadata filters, tags, and Gaia-X credentials to locate suitable assets.
- **Purchase or Access:** Acquire Data Tokens to access assets according to defined policies (download, compute, or SaaS invocation).
- **Secure Transactions:** Smart contracts enforce payment terms and access conditions, ensuring reliable and tamper-proof transactions.

8. Programmatic Publication and Automation with Nautilus

To support developers with streamlined publication and consumption via a CLI, deltaDAO has developed a data space toolkit called Nautilus [14]. Nautilus is a TypeScript-based library streamlining publication and consumption workflows for data and software assets. It automates metadata configuration, simplifies access control management, and enables event-driven actions (e.g., triggering model runs upon new data publication). By providing a robust developer experience, Nautilus encourages the rapid development of automated pipelines, enhances reusability, and allows for seamless integration with C2D-based workflows. [15]

Key Nautilus Features:

- **Simplified Metadata Configuration:** Utilize a builder pattern for consistent, maintainable asset definitions.
- **Complete Asset Management:** Manage full asset lifecycles, including C2D integration for secure computations.
- **Automated Data Flows:** Implement event listeners and webhooks to trigger actions on new data or algorithm updates.

- **Seamless Compute Jobs:** Start, monitor, and retrieve C2D job results directly from development environments.
- **Pricing and Access Control Automation:** Programmatically manage pricing models, allowlists, deny-lists, and user-specific controls.
- **Enhanced Developer Experience:** Focus on data-driven application logic rather than underlying smart contracts or tokenization details.

By providing a robust publication and consumption flow, ACCURATE ensures that data, containerized algorithms, and SaaS offerings are integrated securely and efficiently into the manufacturing data ecosystem, fostering a scalable, privacy-preserving, and innovation-friendly environment. [15]

2.5 Leveraging Compute-to-Data for Manufacturing as a Service

In a Manufacturing as a Service (MaaS) paradigm, the ability to discover, evaluate, and select suitable production resources from a distributed network of manufacturers is essential. Compute-to-Data (C2D) provides the technical foundation to enable these processes while preserving data sovereignty, confidentiality, and intellectual property.

Conceptual Overview:
Consider a scenario where a service consumer seeks to produce a custom component as a service. Instead of directly sharing their proprietary specifications with all potential manufacturers—or requiring manufacturers to reveal sensitive capacity, cost, or performance details—the C2D architecture allows computations to take place locally at the respective producers. The result is a federated analytics or federated learning setup where sensitive insights are never exposed outside the data owner's premises [2].

Federated Local Computation:

1. **Service Consumer Requirements:**
   The consumer defines manufacturing requirements, which can include technical specifications, quality standards, delivery times, cost targets, and sustainability criteria. These requirements are used as input parameters into a matchmaking algorithm.
2. **Local Evaluation at Producers:**
   Multiple producers run the matchmaking or evaluation algorithm locally (within their own C2D environments) against their confidential data—such as machine availability, internal cost structures, production capacity, and lead times.
   - **No Raw Data Exposure:** Thanks to C2D, producers do not have to share their proprietary metrics. The algorithm executes at the producer's location, consuming inputs and producing anonymized or aggregated performance metrics.
   - **Federated Analytics & Learning:** These local computations can leverage federated analytics or federated learning techniques, enabling optimization or model training across multiple producers' data without centralizing it.
3. **Algorithm Provisioning by a Certified Software Provider:**
   A certified software provider can publish the matchmaking and evaluation algorithms as containerized assets in the ACCURATE data space. These algorithms might incorporate

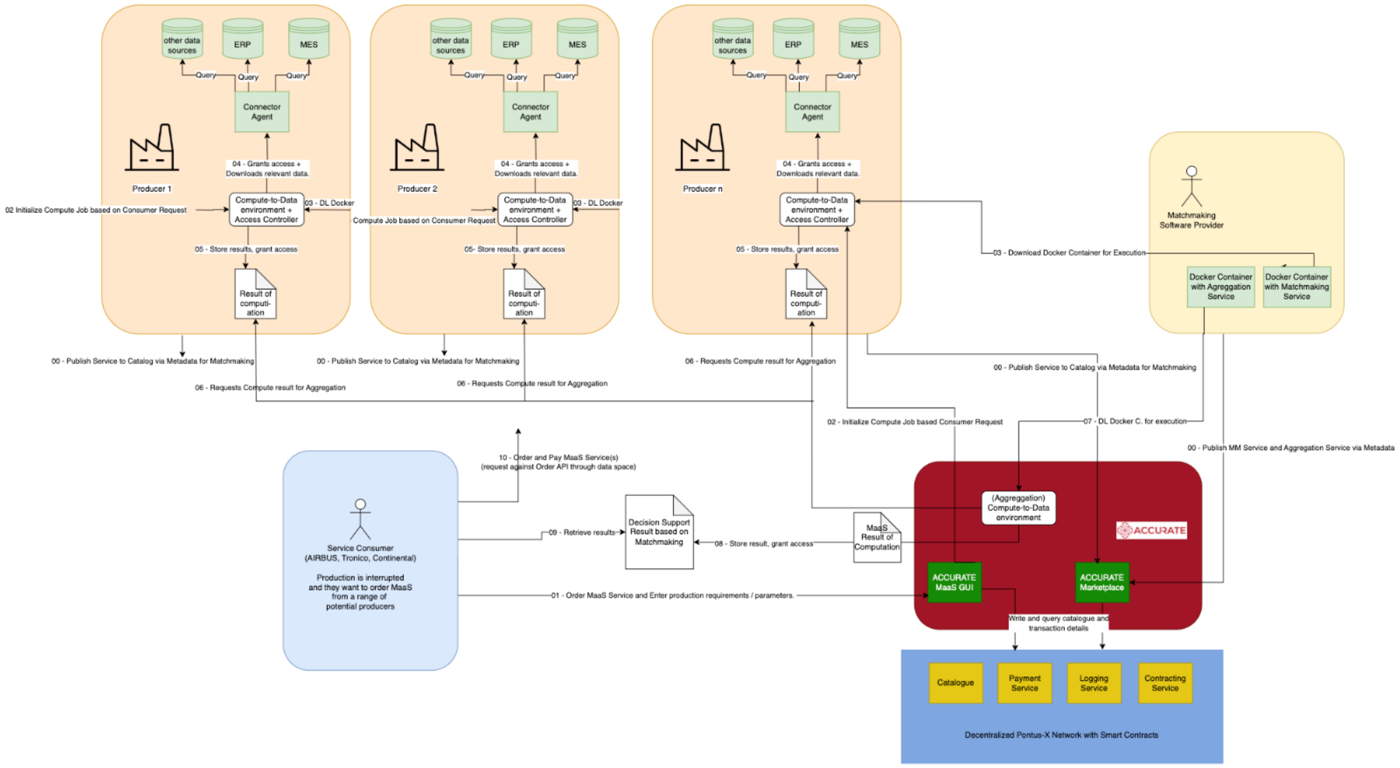
complex optimization logic, AI-driven analytics, or sustainability scoring models. The provider can monetize these algorithms by controlling their compute access, ensuring that each local evaluation run results in fair compensation.

Centralized Aggregation (Second C2D Stage):

After local computations are completed, each producer returns its anonymized, aggregated results—such as feasible production times, estimated costs, or sustainability scores—to a second C2D environment. This second environment serves as a trusted aggregator:

- Aggregated Matchmaking: The aggregator, running C2D tasks, consolidates the local results to rank or shortlist producers that best fit the consumer's specified criteria. This may involve multi-criteria decision-making—balancing economic, quality, and sustainability factors—without ever exposing the individual producers' sensitive details.
- Result Delivery to the Service Consumer:
  The outcome is a ranked set of potential manufacturing partners. The consumer receives these results from the aggregator. They can then select their preferred partner, initiate contract negotiations, and proceed with the MaaS transaction.

The architecture of this MaaS conceptual overview is visualized below:



other data sources
ERP
MES
Query
Query
Query
Connector Agent
Producer 1
04 - Grants access + Downloads relevant data.
02 Initialize Compute Job based on Consumer Request
Compute-to-Data environment + Access Controller
03 - DL Docker
05 - Store results, grant access
Result of computi-ation
other data sources
ERP
MES
Query
Query
Query
Connector Agent
Producer 2
04 - Grants access + Downloads relevant data.
Compute Job based on Consumer Request
Compute-to-Data environment + Access Controller
03 - DL Docker
05- Store results, grant access
Result of computi-ation
other data sources
ERP
MES
Query
Query
Query
Connector Agent
Producer n
04 - Grants access + Downloads relevant data.
Compute-to-Data environment + Access Controller
05 - Store results, grant access
Result of computi-ation
Matchmaking Software Provider
03 - Download Docker Container for Execution
Docker Container with Agreggation Service
Docker Container with Matchmaking Service
00 - Publish Service to Catalog via Metadata for Matchmaking
00 - Publish Service to Catalog via Metadata for Matchmaking
06 - Requests Compute result for Aggregation
06 - Requests Compute result for Aggregation
06 - Requests Compute result for Aggregation
00 - Publish Service to Catalog via Metadata for Matchmaking
02 - Initialize Compute Job based Consumer Request
07 - DL Docker C. for execution
00 - Publish MM Service and Aggregation Service via Metadata
10 - Order and Pay MaaS Service(s) (request against Order API through data space)
Service Consumer (AIRBUS, Tronico, Continental)
Production is interrupted and they want to order MaaS from a range of potential producers
09 - Retrieve results
Decision Support Result based on Matchmaking
08 - Store result, grant access
MaaS Result of Computation
(Aggreggation) Compute-to-Data environment
ACCURATE
ACCURATE MaaS GUI
ACCURATE Marketplace
01 - Order MaaS Service and Enter production requirements / parameters.
Write and query catalogue and transaction details
Catalogue
Payment Service
Logging Service
Contracting Service
Decentralized Pontus-X Network with Smart Contracts

Figure 4: MaaS architecture enabled by C2D (Source: deltaDAO)

Benefits of C2D-Enabled MaaS:

- Data Sovereignty: Producers maintain full control over their sensitive operational data.

- Competitive Advantage: Producers can participate in open markets without risking exposure of their internal strategies or cost structures.
- Supply Chain Resilience: Consumers gain insights into viable manufacturing options across different producers, fostering resilience and flexibility.
- Fair Monetization of Algorithms: Software providers can supply algorithms that power these matchmaking scenarios, monetizing their intellectual property through secure C2D computations, incentivizing the development of advanced analytics solutions for MaaS ecosystems.

By integrating Compute-to-Data within the MaaS landscape, the ACCURATE project enables a dynamic and confidential marketplace where stakeholders collaborate securely. Federated analytics and learning capabilities, driven by C2D and containerized algorithmic services, create an environment where trust, innovation, and value creation thrive—even in complex, multi-tier manufacturing networks. [2]

3 Conclusion

This deliverable has detailed the design, architecture, and operational principles of a secure Compute-to-Data (C2D) environment integrated into the ACCURATE data space. By aligning with Gaia-X principles, European data governance standards, and the project's overarching goals, this C2D environment enables federated analytics, federated learning, and the secure exchange of data and software assets—all while preserving sovereignty over sensitive manufacturing information.

Key features—such as on-site data processing, privacy-by-design computation, and containerized algorithm deployments—allow participants to run computations on proprietary data without ever revealing raw inputs. This approach fuels collaborative innovation in Manufacturing as a Service (MaaS) scenarios, where diverse stakeholders can securely interact, train AI models, simulate operational changes, and assess sustainability or circularity performance. Through configurable permissions and policies, tokenized access, and audit trails, C2D fosters a trusted environment that balances flexibility with robust compliance and IP protection.

The integration of C2D within the ACCURATE data space supports a wide range of use cases, from assessing resilience in complex supply chains to creating entirely new data products and services. Such versatility ensures that manufacturing companies can confidently engage in secure data-driven collaborations, obtain actionable insights, and leverage advanced analytics without undermining their competitive edge or violating regulatory constraints.

Looking ahead, the C2D architecture will continue to evolve, guided by feedback from project participants and alignment with emerging standards and technologies. Subsequent tasks and deliverables will validate the C2D implementation against real-world scenarios, refine configurations, and expand capabilities. Ultimately, the C2D environment lays a strong technical foundation for the ACCURATE vision: a future in which European manufacturing ecosystems are both resilient and adaptive, confidently harnessing the power of data to drive sustainability, innovation, and long-term competitiveness.

1. Bibliography

[1] ACCURATE Project Website, https://accurateproject.eu/

[2] Roman Gehrer, Stefan Dumss, Fabian Gast, Willi Wünschel, Frederic Schwill, Mateo Šoša, Shiyang Zhou, Gerald H. Ristow, Tatevik Gharagyozyan, Clemens Heistracher, Manfred Grafinger, Matthias Weigold, EuProGigant: A decentralized Federated Learning Approach based on Compute-to-Data and Gaia-X, https://doi.org/10.1016/j.procir.2024.07.060

[3] Siska, Veronika, Vasileios Karagiannis, and Mario Drobics. "Building a Dataspace: Technical Overview." *Gaia-X Hub Austria* (2023). https://www.gaia-x.at/wp-content/uploads/2023/04/WhitepaperGaiaX.pdf

[4] Ocean Protocol Compute-to-Data Documentation, https://docs.oceanprotocol.com/developers/compute-to-data

[5] Ocean Enterprise, https://www.oceanenterprise.io/

[6] Ocean Protocol Compute-to-Data Architecture Documentation, https://docs.oceanprotocol.com/developers/compute-to-data/compute-to-data-architecture

[7] Ocean Protocol Compute-to-Data Dataset & Algorithms Documentation, https://docs.oceanprotocol.com/developers/compute-to-data/compute-to-data-datasets-algorithmshttps://docs.oceanprotocol.com/developers/architecture

[8] Ocean Protocol Compute-to-Data Workflow Documentation, https://docs.oceanprotocol.com/developers/compute-to-data/compute-workflowhttps://docs.oceanprotocol.com/developers/architecture

[9] W3C Decentralized Identifiers (DIDs) v1.0 Specifications, https://www.w3.org/TR/did-core/

[10] ERC-20 Token Standard, https://ethereum.org/de/developers/docs/standards/tokens/erc-20/

[11] Ocean DDO Specification, https://docs.oceanprotocol.com/developers/old-infrastructure/aquarius

[12] Ocean Enterprise and Pontus-X Draft DDO Specifications, https://docs.pontus-x.eu/docs/ddo_credential/ddo_intro

[13] Ocean Provider Documentation, https://docs.oceanprotocol.com/developers/old-infrastructure/provider

[14] Nautilus Documentation, https://nautilus.delta-dao.com/

[15] Fabian Gast, Viktor Berchtenbreiter, Stefan Dumss, Roman Gehrer, Paul Weißenbach, Matthias Weigold, Manfred Grafinger, Automatic Publication of Data to Data- and Service Ecosystems from the Shopfloor, https://doi.org/10.1016/j.procir.2024.10.287

[16] ACCURATE Portal GitHub Repository, https://github.com/deltaDAO/portal-accurate

[17] Ocean Enterprise GitHub Repository, https://github.com/OceanProtocolEnterprise