NARRATE

Regenerative Resilient Smart Manufacturing Networks

D3.6 DESIGN & IMPLEMENTATION OF CONTEXTUALIZATION SERVICES (B)

2025/02/25

NARRATE
Regenerative Resilient Smart
Manufacturing Networks

D3.6 DESIGN & IMPLEMENTATION OF CONTEXTUALIZATION SERVICES

Work package	WP 3
Task	T3.6
Due date	30/11/2025
Submission date	30/11/2025
Deliverable lead	SERVTECH
Version	1.0
Authors	Michael Papazoglou, Amal Elgammal, Bernd Kraemer, Mira Raheem (SERV)
Reviewers	Maurizio Griva (REP)
Abstract	This deliverable presents an early version of the Proof-of-Concept (PoC) implementation of the Federated Data Integration Framework (FDIF) developed within the NARRATE project. The FDIF enables the harmonization, validation, semantic enrichment, and federation of heterogeneous industrial data from IoT, ERP, and production systems. Implemented using FastAPI, the PoC adopts a modular, service-oriented architecture in which each transformation stage—harmonization, validation, semantic tagging, and blueprint mapping—is provided as an independent REST endpoint. A dedicated /run service executes the complete pipeline from raw data to Digital Twin Blueprint representation in a single call.

Funded by
the European Union

Funded by the European Union under GA No 101138094. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. • Grant Agreement No.: 101138094
Call: HORZON-CL4-2023-TWIN-TRANSITION-01 • Topic: HORIZON-CL4-2023-TWIN-TRANSITION-01-07 • Type of action: HORIZON Research and Innovation Actions

UK Research and Innovation

UK Research and Innovation,
Horizon Europe Guarantee.

1

	This deliverable aims to validate the FDIF architecture and its operational implementation as a foundational element of the NARRATE framework. The FDIF provides the semantic integration and data federation layer that links IoT, ERP, BOM/MBOM, and CEP data sources to the IMC. Its implementation demonstrates how standardized, semantically enriched data can support intelligent analytics, decision-making, and resilience within Smart Manufacturing Networks (SMNs). The work establishes a practical reference model for future FDIF extensions and integration into NARRATE's IMC environment. .
Keywords	Data Homogenization, Semantic Integration, Production Intelligence, Contextualisation, Federated data architecture, data mesh, proof of concept implementation

Project co-funded by the European Commission in the H2020

Nature of the deliverable:	R, DEM
Nature of the deliverable:	P

DOCUMENT REVISION HISTORY

Version	Date	Description of change	List of contributor(s)
0.1	2025/10/20	Definition of the Table of Contents	SERV
0.2	2025/10/31	First draft	SERV
0.3	2025/11/04	Partner contributions	BLU,
0.4	2025/11/20	Review of V0.3	REP
0.5	2025/11/27	Final version incorporating reviewers' comments,	SERV
1.0	2025/11/27	Final revision by coordination	AID

STATEMENT ON MAINSTREAMING GENDER

The NARRATE consortium is committed to including gender and intersectionality as a transversal aspect in the project's activities. In line with EU guidelines and objectives, all partners – including the authors of this deliverable – recognise the importance of advancing gender analysis and sex-disaggregated data collection in the development of scientific research. Therefore, we commit to paying particular attention to including, monitoring, and periodically evaluating the participation of different genders in all activities developed within the project, including workshops, webinars and events but also surveys, interviews and research, in general. While applying a non-binary approach to data collection and promoting the participation of all genders in the activities, the partners will periodically reflect and inform about the limitations of their approach. Through an iterative learning process, they commit to plan and implement strategies that maximise the inclusion of more intersectional perspectives in their activities.

DISCLAIMER

COPYRIGHT NOTICE

The NARRATE Consortium is the following:

Participant number	Participant organisation name	Short name	Country
1	INSTITUTO TECNOLOGICO METALMECANICO, MUEBLE, MADERA, EMBALAJE Y AFINES-AIDIMME	AID	ES
2	SCIENTIFIC ACADEMY FOR SERVICE TECHNOLOGY EV	SERV	DE
3	FRAUNHOFER GESELLSCHAFT ZUR FORDERUNG DER ANGEWANDTEN FORSCHUNG EV	FhG	DE
4	INSTITUT NATIONAL DES SCIENCES APPLIQUEES DE LYON	INSA	FR
4.1	INSAVALOR SA	INSA-V	FR
5	SOFTWARE AG	SAG	DE
6	F6S NETWORK IRELAND LIMITED	F6S	IE
7	SYNESIS-SOCIETA CONSORTILE A RESPONSABILITA LIMITATA	SYN	IT
8	MEDITERRANEAN WOOD FACTORY S.L. - MEDWOOD	MED	ES
9	DHL EXEL SUPPLY CHAIN SPAIN SL	DHL	ES
10	NUNSYS SA	NUN	ES
11	BUDATEC GMBH	BUD	DE
12	BRUNEL UNIVERSITY LONDON	BUL	UK
13	POLICY LAB	PL	EE
14	BLUEBRIDGE SOLUTIONS UG	BLU	DE
15	SANTER REPLY SPA	REP	IT

Table of Contents

List of figures

List of tables

Abbreviations

AAS	Asset Administration Shell
AI	Artificial Intelligence
API	Application Programming Interface
BOM	Bill of Materials
CA	Consortium Agreement
CDIF	Cross-Domain Interoperability Framework
CEP	Complex Event Processing
CFX	Connected Factory Exchange
DL	Deliverable leader
DM	Data Manager
DT	Digital Twin
EC	European Commission
ERP	Enterprise Resource Planning
FAIR	Findability, Accessibility, Interoperability, and Reusability
FDIF	Federated Data Integration Framework
IMC	Intelligent Manufacturing Custodian
IoT	Internet of Things
IIoT	Industrial Internet of Things
ISO	International Organization for Standardization
KPI	Key Performance Indicator
MaaS	Manufacturing as a Service
MBOM	Manufacturing Bill of Materials
MES	Manufacturing Execution System
PLM	Product Lifecycle Management
REST	Representational State Transfer
SHACL	Shapes Constraint Language
SKOS	Simple Knowledge Organization System
SMN	Smart Manufacturing Network

Tx.x	Task x.x
WP	Work Package

EXECUTIVE SUMMARY

The Federated Data Integration Framework (FDIF) that was presented in D3.1 is a core enabler within the NARRATE architecture, designed to achieve seamless data federation and semantic interoperability across the manufacturing ecosystem. FDIF harmonizes information originating from Enterprise Resource Planning (ERP) systems, Bill of Materials (BOM) and Manufacturing Bill of Materials (MBOM) structures, Internet of Things (IoT) sources, and Complex Event Processing (CEP) streams. Through its federated approach, FDIF enables comprehensive visibility, traceability, and intelligent orchestration across the Smart Manufacturing Network (SMN).

FDIF's central mission is to integrate, harmonize, and semantically enrich heterogeneous manufacturing data drawn from distributed systems and production environments. By providing a hybrid federated data mesh model, FDIF allows consistent interpretation and interoperability of information across enterprise, engineering, and operational levels. D3.6 focuses on an early implementation of the FDIF that performs the following key functions:

- Data Harmonization: supports the integration of structured (ERP, BOM) and dynamic (IoT, CEP) data streams into a unified schema.

- Semantic Annotation: Uses JSON-LD metadata to capture contextual meaning, functional roles, and relationships among data entities.

- Standardized Exchange: Implements IPC-2591 (CFX) messaging for process-level and equipment-level data exchange.

- Digital Twin Integration: Supplies harmonized and semantically tagged data to Blueprint Frames — the Digital Twin representations that mirror production assets, processes, and supply chain entities.

This deliverable builds upon and extends the work presented in Deliverable D3.1a by translating the FDIF design specifications into concrete implementation decisions and development activities. In the course of this transition from design to implementation, several architectural elements were streamlined and adapted to ensure coherence with pilot requirements and prototype development priorities. These adjustments were guided by practical considerations arising from the implementation context and aimed at enhancing the overall feasibility and performance of the FDIF framework.

1. INTRODUCTION

1.1. FEDERATED DATA INTEGRATION FRAMEWORK (FDIF) — PURPOSE WITHIN NARRATE

The Federated Data Integration Framework is the data harmonization and interoperability core of the NARRATE architecture. It establishes a hybrid federated data mesh, that seamlessly connects, integrates, and contextualizes heterogeneous manufacturing and enterprise data sources across the Smart Manufacturing Network.

Rather than consolidating data into a single repository, FDIF applies data federation [1], [2] and data mesh [3] principles to achieve a distributed yet semantically unified data space. Each connected system — such as ERP, BOM/MBOM, IoT, and Complex Event Processing (CEP) platforms — maintains its autonomy while exposing harmonized, metadata-rich data entities that can be queried, exchanged, and reused across the manufacturing network.

FDIF's hybrid approach combines the governance and consistency of data federation with the scalability and domain ownership of data mesh architectures, enabling manufacturing enterprises to maintain data sovereignty while achieving interoperability and cross-domain coordination.
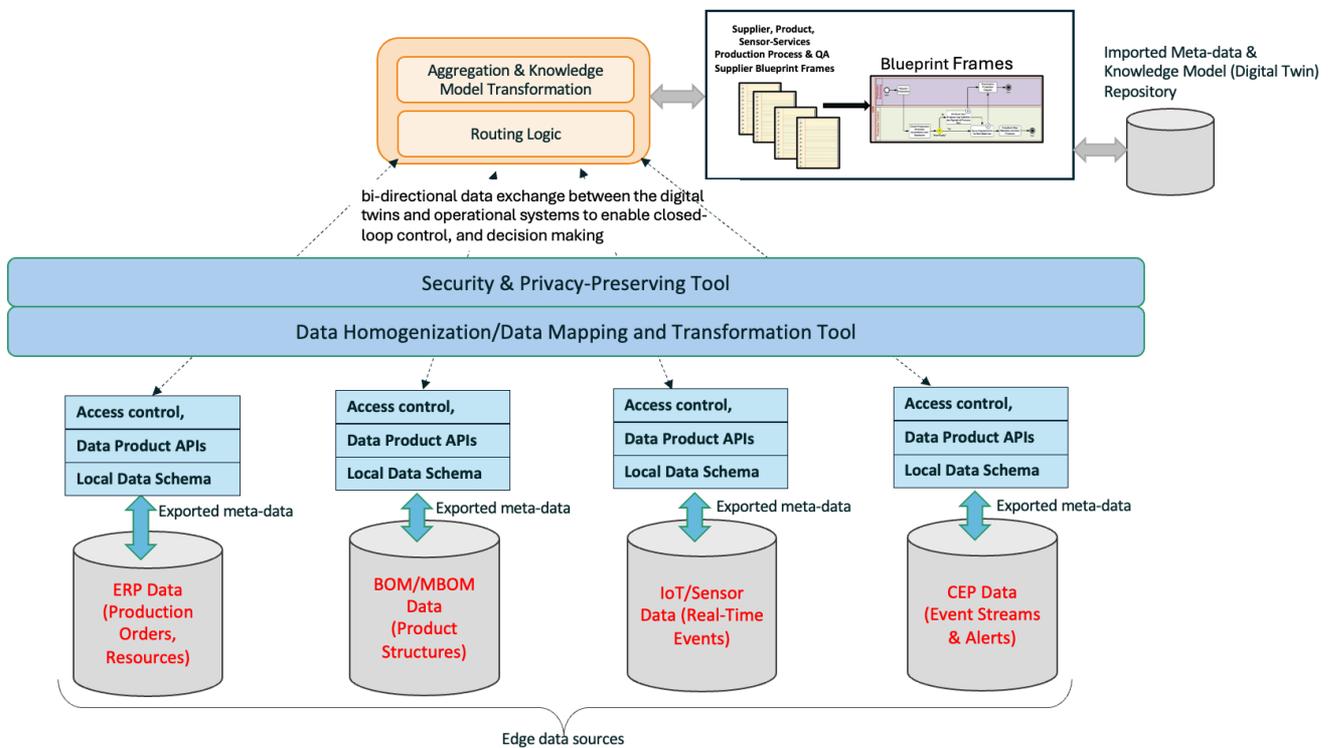

FIGURE 1 DATA HARMONIZATION IN THE FEDERATED DATA INTEGRATION FRAMEWORK.

Through data categorization, metadata annotation, and ontology-driven mapping, FDIF transforms raw industrial data into semantically harmonized formats aligned with open standards such as IPC-2591 Connected Factory Exchange (CFX) [4]. This ensures consistent representation of entities such as Production.WorkOrder, ProductData.Definition, ProcessData.SensorReading, and Traceability.Event across different systems.

Within NARRATE, FDIF plays a pivotal role in linking the operational and intelligence layers of the architecture as shown in Figure-1. To this end, the FDIF:

- collects and normalizes data from various manufacturing and enterprise domains.

- enriches and annotates this data with semantic metadata to maintain traceability and contextual meaning.

- feeds the Blueprint Frames, which compartmentalise, transform this enriched data into context-aware knowledge required by the blueprint frames, power real-time visibility, and process it to support the Intelligent Manufacturing Custodian (IMC) perform real-time monitoring, disruption prediction, and sustainability assessment and circularity stress testing.

Each incoming data element (from ERP, MBOM, IoT, etc.) is enriched with:

- Semantic category: identifies the business or operational domain (e.g., Product, Process, Resource, Sensor, Event).

- Functional role: clarifies its use (e.g., bill-of-material definition, work order, environmental sensor reading, machine alert).

- Contextual metadata: adds details such as timestamp, source system, data owner, unit, or confidence level.

In essence, FDIF transforms fragmented production and enterprise data into actionable, standardized, and context-aware knowledge, forming the foundation for NARRATE's cognitive, resilient, and AI-enhanced Manufacturing-as-a-Service ecosystem.

In summary the FDIF is the data unification and synchronization layer that:

- Integrates heterogeneous data sources across the Smart Manufacturing Network.

- Harmonizes data models into a Blueprint-compatible structure.

- Enables bi-directional data flow between enterprise (ERP), engineering (BOM/MBOM), and operational (sensor/production) layers.

- Supports the Intelligent Manufacturing Custodian (IMC) by ensuring data consistency and interoperability across all Blueprint Frames.

1.2. DATA SOURCES AND INTEGRATION MECHANISM

In its current form FDIF harmonizes information from four principal data domains:

- ERP Systems:
  Provide production orders, inventory levels, scheduling data, and cost structures. ERP records are annotated with JSON-LD semantics identifying entities such as Work Orders, Resources, and Production Lots, enabling traceability within Digital Twins.

- BOM and MBOM Structures:
  Represent the product definition and the corresponding manufacturing process

plans. BOM data define the product composition, while MBOM extends this to manufacturing operations, tooling, and resources. FDIF links these layers through JSON-LD relations, enabling Blueprint Frames to maintain a synchronized view of both design intent and manufacturing execution.

- IoT Data Streams:Capture real-time machine and sensor data (e.g., temperature, vibration, performance). These are normalized and formatted into IPC-2591 (CFX) messages, ensuring standardized machine-to-system interoperability. JSON-LD tagging provides semantic context, associating each data stream with its asset, process step, or condition state within the relevant Blueprint Frame.

- CEP (Complex Event Processing) Data: Monitors and detects composite events across the production network, such as deviations, delays, or quality anomalies. FDIF semantically annotates CEP outputs with JSON-LD metadata, specifying event types, triggered assets, and affected process steps, which are then linked dynamically to corresponding Digital Twin entities.

1.3. USE OF STANDARDS AND TECHNOLOGIES

FDIF leverages an implementation of the IPC-2591 (CFX) standard to ensure syntactic interoperability in equipment communication and process reporting. CFX messages enable:

- Standardized event and process notifications,

- Equipment state and performance reporting, and

- Data consistency across multi-vendor environments.

To achieve semantic interoperability, FDIF applies JSON-LD (JavaScript Object Notation for Linked Data) annotations that:

- Define the contextual meaning of each data entity (e.g., product, process, event).

- Establish interrelationships between data sources (e.g., linking a sensor reading to a process step and product component).

- Map to established ontologies such as the Asset Administration Shell (AAS) [5] and CFX data models.

This dual use of CFX for structure and JSON-LD for semantics allows FDIF to unify data streams into coherent, machine-interpretable representations for Blueprint Frames.

2. CURRENT IMPLEMENTATION STATUS

An early and partial implementation of the Federated Data Integration Framework is currently operational, serving as a proof-of-concept prototype to validate its foundational principles and technical feasibility. This early experimentation phase was deliberately designed to concentrate on a small but focused set of core functionalities, enabling architectural validation, early experimentation, incremental prototyping, and risk reduction before large-scale deployment. The ongoing implementation emphasizes the following key capabilities:

- Data Federation and Harmonization: Integration of heterogeneous datasets from ERP, BOM/MBOM, IoT, and CEP sources into a unified, semantically coherent data layer. This ensures end-to-end continuity between business planning, design, production, and operational monitoring.

- Standardized Data Exchange: Generation of IPC-2591 (CFX)–conformant messages for selected production processes, enabling consistent and vendor-neutral communication between equipment, systems, and higher-level applications.

- Semantic Enrichment with JSON-LD: Embedding of JSON-LD metadata [6] to encode contextual, relational, and functional meaning within datasets—allowing cross-domain mapping, traceability, and intelligent data discovery across the federated environment.

- Digital Twin Integration through Blueprint Frames: Linking harmonized and semantically annotated data streams to Blueprint Frames [7], which serve as the Digital Twin representations of assets, products, and processes. These frames leverage the federated data for real-time monitoring, validation, predictive simulation, and continuous optimization.

This implementation substantiates the practicality and value of federating heterogeneous manufacturing data through a standards-based and semantically rich approach. It demonstrates how FDIF can underpin Blueprint Frame–driven Digital Twin environments that accurately mirror real-world operational states and support data-driven decision-making across the manufacturing network.

3. WHAT IS IPC-2591 (CFX)

CFX (Connected Factory Exchange) is an open, vendor-neutral communication standard for smart manufacturing systems, officially defined by the IPC-2591 standard. CFX, although used predominantly in the electronics manufacturing industry, can be readily adopted to broader digital and smart manufacturing environments due to its open, vendor-agnostic communication model. The IPC-CFX standard provides a unified, message-based protocol that enables seamless interoperability between heterogeneous machines and systems. In Smart Manufacturing, CFX can:

- Unify equipment data acquisition: Connect diverse machines and sensors without custom adapters.

- Provide real-time operational visibility: Stream standardized machine, process, and quality data into MES, Digital Twins, or analytics platforms.

- Enable closed-loop optimization: Feed live CFX events into AI, CEP engines, or orchestration systems for predictive maintenance, anomaly detection, or dynamic scheduling.

- Strengthen traceability and compliance: Capture consistent material, process, and quality information required for audits and regulatory frameworks.

- Reduce integration complexity across sites: The standardized interface accelerates deployment of distributed or multi-site Smart Manufacturing Networks.

Applied beyond electronics, CFX can serve as a robust equipment-level data foundation for systems like the NARRATE IMC, BMS, and SCST, supporting interoperable, resilient, and data-driven production ecosystems. It provides a universal language and data

exchange protocol for manufacturing equipment, software, and systems (like MES, ERP, PLM, and IoT platforms) — allowing them to communicate seamlessly without custom interfaces. CFX is part of the "Connected Factory Initiative" (https://www.electronics.org/committee-detail/2-17), aimed at creating plug-and-play interoperability in Industry 4.0 environments.

3.1. CFX IN SIMPLE TERMS

CFX is the "grammar" of digital manufacturing — a standardized way for machines, sensors, and enterprise systems to exchange structured, meaningful information about:

- What is being made (ProductData.Definition)
- How it's being made (ProcessData.Operation)
- Who/What made it (EquipmentInformation)
- When and where it happened (Traceability.Event)
- Why something occurred (e.g., fault, alarm)

All of this data is exchanged as self-describing JSON messages, transmitted over modern lightweight protocols like AMQP or MQTT.

3.2. KEY CHARACTERISTICS

Feature	Description
Format	JSON-based message structures using MQTT transport
Scope	Covers production operations, process data, traceability, and analytics
Granularity	From high-level work orders to low-level sensor readings
Semantics	Provides standardized message schemas for all factory events
Extensibility	Modular namespaces (e.g., Production.WorkOrder, ProcessData.SensorReading)
Transport Layer	Supports lightweight, event-driven, publish–subscribe data exchange

TABLE 1 CFX KEY CHARACTERISTICS.

Category	Example Message	Description
Production	Production.WorkOrder	Work order and job definition
ProductData	ProductData.Definition	BOMs, part structures, recipes
ProcessData	ProcessData.SensorReading	Equipment telemetry and IoT sensor data
Traceability	Traceability.Event	Product or process trace logs

Material	Material.CarrierInformation	Material and logistics information
Quality	Quality.InspectionResult	Inspection or quality outcomes
Maintenance	Maintenance.EquipmentStatus	Machine health, status, and alerts

TABLE 2 TYPICAL CFX MESSAGE CATEGORIES (PER IPC-2591)

CFX Message Type	Function
Production.WorkOrder	Defines production tasks and scheduling from ERP/MES
ProductData.Definition	Contains BOMs, product structure, revisions
ProcessData.Operation	Represents execution details of a process step
ProcessData.SensorReading	Encapsulates real-time IoT or equipment telemetry
Traceability.Event	Logs quality, process, or disruption events
Material.CarrierInformation	Describes material logistics and flow
Quality.InspectionResult	Communicates quality inspection outcomes

TABLE 3 EXAMPLES OF CORE CFX MESSAGE CATEGORIES.

The three tables above summarise the CFX key characteristics, typical message categories, and provide examples of CFX message categories.

3.3. EXAMPLE OF CFX MESSAGE SNIPPETS

CFX provides the standardized message vocabulary and exchange protocol that FDIF uses to externalize and exchange this harmonized data across the Smart Manufacturing Network.

a. Production.WorkOrder

```
{
  "MessageName": "Production.WorkOrder",
  "WorkOrderId": "WO-2025-00045",
  "ProductId": "PRD-9001",
  "Quantity": 500,
  "StartTime": "2025-10-20T08:00:00Z",
  "DueDate": "2025-10-22T18:00:00Z",
  "Status": "Released"
}
```

b. ProductData.Definition

```
{
  "MessageName": "ProductData.Definition",
  "ProductId": "PRD-9001",
```

```
  "Revision": "A2",
  "Components": [
    {"PartNumber": "CMP-001", "Description": "Base Frame"},
    {"PartNumber": "CMP-002", "Description": "Sensor Housing"}
  ],
  "BOMVersion": "v3.4"
}
```

c. ProcessData.Operation

```
{
  "MessageName": "ProcessData.Operation",
  "OperationId": "OP-1001",
  "ProcessName": "Laser Cutting",
  "Parameters": {
    "Speed": "1200mm/s",
    "Power": "250W"
  },
  "OperatorId": "MACH-03"
}
```

d. ProcessData.SensorReading

```
{
  "MessageName": "ProcessData.SensorReading",
  "SensorId": "TEMP_01",
  "Value": 85.3,
  "Unit": "Celsius",
  "Timestamp": "2025-10-20T10:15:00Z"
}
```

4.   FDIF TO CFX INTEGRATION WORKFLOW

The FDIF–CFX Integration Workflow establishes a seamless data transformation and exchange pipeline that bridges heterogeneous manufacturing and enterprise systems with standardized, interoperable communication formats. The process connects multiple data sources — ERP, BOM, MBOM, IoT, and CEP — through a series of harmonization and standardization steps, ensuring consistent semantics and traceable data flow into the Blueprint Frames and the Intelligent Manufacturing Custodian analytics layer.

The FDIF–CFX integration workflow operationalizes semantic manufacturing interoperability by connecting enterprise and operational systems through a federated, ontology-driven data mesh. By leveraging the IPC-2591 CFX standard, FDIF ensures that

manufacturing data becomes consistent, traceable, and exchangeable, providing the essential data backbone for resilient, intelligent, and sustainable manufacturing operations within the NARRATE ecosystem. The implementation of FDIF to CFX Integration Workflow steps are outlined below and shown in Figure-2.
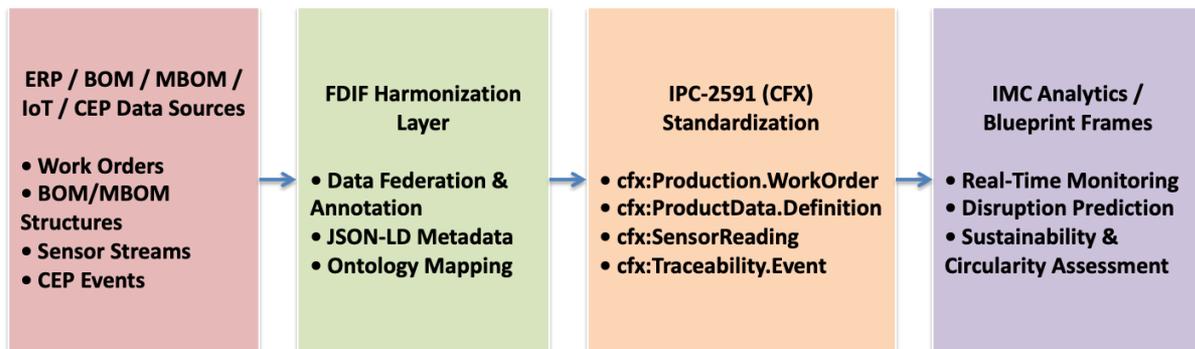

FIGURE 2 FDIF TO IPC-2591 (CFX) INTEGRATION WORKFLOW.

Step 1. FDIF Data Federation and Harmonization

The FDIF acts as the unifying layer that collects, aligns, and contextualizes data from distributed sources across the manufacturing ecosystem. It operates using a hybrid federated data mesh approach — enabling each source to retain ownership of its data while contributing harmonized and semantically annotated datasets to the common integration layer.

Data Sources Integrated:

- ERP Systems: Capture operational and administrative data such as work orders, production schedules, and inventory levels.

- BOM / MBOM Systems: Define the hierarchical structure of products, assemblies, and corresponding manufacturing processes.

- IoT Platforms: Generate continuous streams of telemetry and condition-monitoring data from sensors, machines, and devices.

- CEP Engines: Produce complex event notifications derived from rule-based or AI-driven analysis of streaming data (e.g., anomaly detection, predictive alerts).

All these inputs are semantically annotated with JSON-LD metadata and represented as instances of the fdif:DataEntity class (and its specializations such as fdif:ERPDataEntity or fdif:IoTDataEntity).
This ensures a consistent data vocabulary, making entities machine-understandable and ready for ontology-based transformation.

Step 2. Mapping to IPC-2591 (CFX) Schema

Once harmonized within FDIF, the semantically enriched data is mapped to the IPC-2591 Connected Factory Exchange (CFX) standard. This transformation aligns each FDIF data entity with a corresponding CFX message type, enabling interoperability with industrial systems supporting open CFX communication. Table-4 outlines the FDIF entity to CFX message mapping.

FDIF Entity	Mapped CFX Message	Semantic Purpose
fdif:ERPDataEntity	cfx:Production.WorkOrder	Describes job and order information (e.g., order ID, quantity, due date).
fdif:BOMDataEntity	cfx:ProductData.Definition	Defines product structure, components, and material hierarchy.
fdif:MBOMDataEntity	cfx:ProcessData.Operation	Represents process execution details, equipment setup, and parameters.
fdif:IoTDataEntity	cfx:ProcessData.SensorReading	Reports real-time sensor or telemetry data (temperature, vibration, etc.).
fdif:CEPDataEntity	cfx:Traceability.Event	Captures complex events, alarms, and fault notifications.

TABLE 4 ENTITY TO MESSAGE MAPPING.

Through this mapping, FDIF produces *syntactically standardized and semantically coherent CFX messages* that are fully compatible with industry-wide machine-to-machine communication protocols.

Step 3. CFX Message Exchange and Communication Layer

Once converted into CFX-compliant messages, the data flows through the *CFX communication bus* — an open, vendor-neutral messaging layer. Each JSON message is *self-describing*, containing:

- **Metadata** for schema versioning, timestamps, and traceability;
- **Payloads** representing harmonized production data;
- **Source identifiers** linking back to FDIF entities for provenance tracking.

This message exchange mechanism enables **plug-and-play integration** among participating systems, allowing seamless connectivity across the Smart Manufacturing Network.

Step 4. Integration into Blueprint Frames and IMC Analytics

The standardized CFX messages produced by FDIF are then consumed by Blueprint Frames within the **Intelligent Manufacturing Custodian**. These frames act as the *data intelligence layer* — organizing, correlating, and analyzing the incoming data streams to enable:

- **Real-time production monitoring** and system state visualization;
- **Disruption prediction** using CEP and AI-driven analytics;
- **Sustainability and circularity assessments** through the SCST (Sustainability & Circularity Stress Testing) Tool.

Blueprint Frames provide a modular, semantic interface between raw manufacturing data and higher-level decision-support services within NARRATE.

5.   BENEFITS OF USING IPC-2591 CFX WITHIN FDIF

The following table summarizes the benefits of using IPC-2591 CFX within FDIF.

Benefit	Description
Semantic Interoperability	Establishes a shared data vocabulary and structure for all systems, eliminating ambiguity across domains.
Data Traceability	Ensures end-to-end visibility from ERP order creation to IoT sensor output and event logs.
Plug-and-Play Integration	Open, JSON-based standard allows systems to interoperate without vendor lock-in or proprietary connectors.
Event-Driven Data Flow	Enables real-time updates and intelligent event correlation between production and analytics systems.
Data Type Flexibility	Allows to describe different data categories (static and dynamic information such as BOM, components, products, physical and virtual sensor values etc.) in a single, unified human-readable format.
Open Standards Alignment	Natively aligns with FDIF's semantic metadata, IMC's ontology, and IPC-2591 CFX schema definitions, ensuring future-proof interoperability.

TABLE 5 BENEFITS OF USING IPC-2591 CFX WITHIN FDIF.

Using IPC-2591 within FDIF enables:

- Standardized data exchange across heterogeneous systems and partners

- Machine-readable interoperability for ERP–MES–IoT integration

- Plug-and-play communication in Smart Manufacturing Networks

- Traceable data lineage, from product definition to real-time execution and event handling
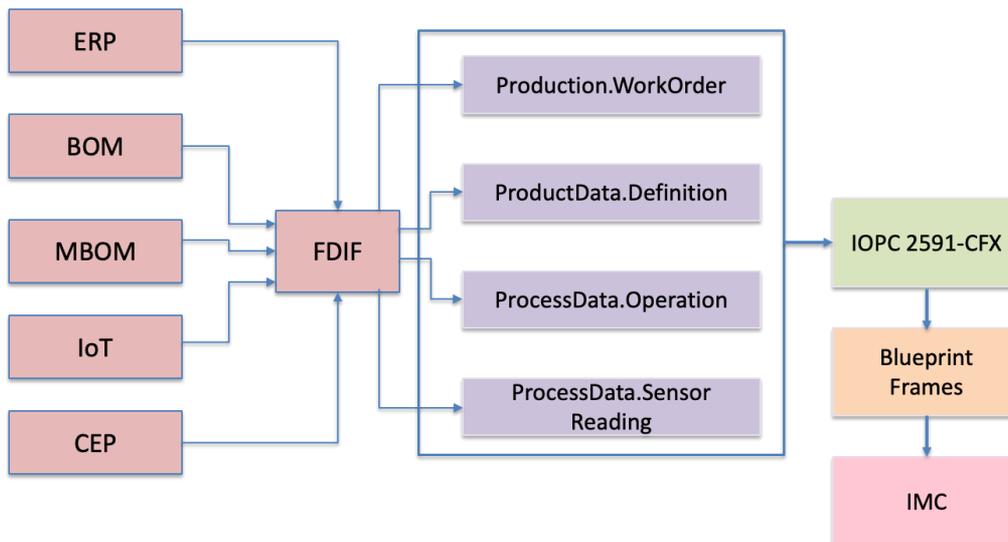

FIGURE 3 FDIF DATA FLOW AND MAPPING FROM (ERP/BOM/IOT/CEP) INTO IPC-2591 CFX MESSAGE TYPES

In essence, CFX is the data transport and structure standard, while FDIF is the semantic and integration layer that harmonizes, annotates, and contextualizes the data. Figure-3 shows the FDIF data flow and mapping from enterprise and operational sources (ERP/BOM/IoT/CEP) through the FDIF harmonization layer, into IPC-2591 CFX message types, and finally reaching the Blueprint Frames and the Intelligent Manufacturing Custodian (IMC) for orchestration, analytics, and traceability. Figure-4 illustrates text-based ontology hierarchy diagram showing mappings, namespace alignment, and semantic chaining between the FDIF, CFX, and Blueprint layers. This is the semantic continuity layer that shows how FDIF, CFX (IPC-2591), and Blueprint layers form a unified ontology chain from *raw data → standardized message → orchestrated manufacturing intelligence* that can be employed by the IMC.
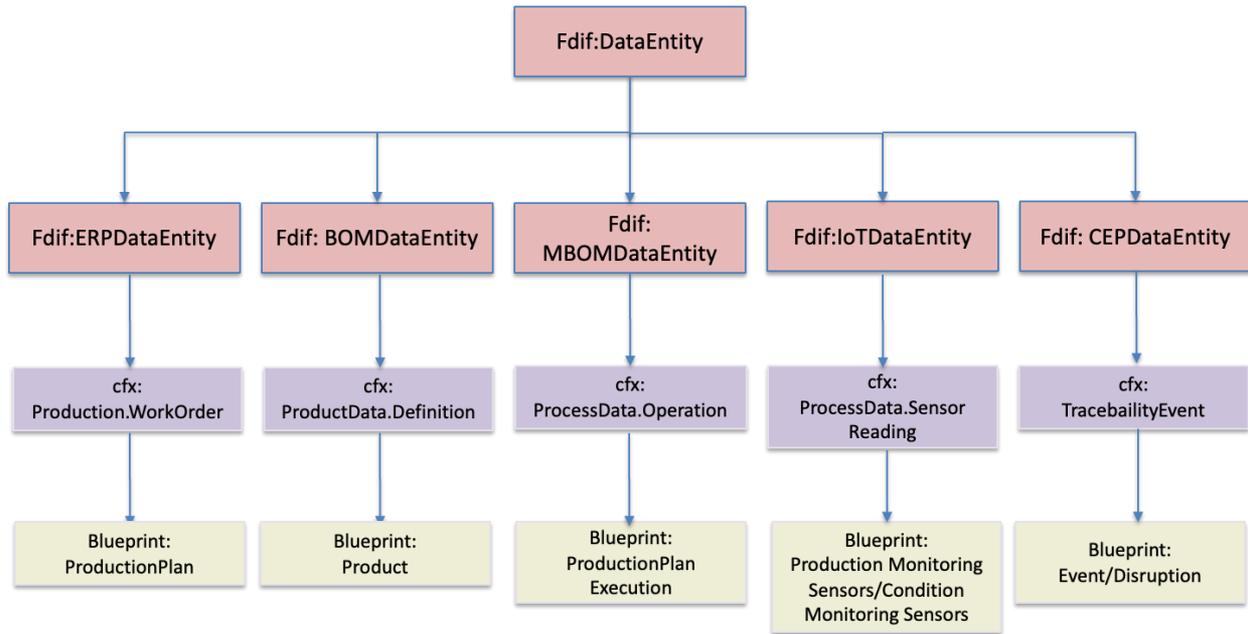
FIGURE 4 ONTOLOGY AND SEMANTIC CHAINING ACROSS MANUFACTURING LAYERS.

6. FDIF AND ITS INTERACTION WITH ERP, BOM/MBOM, IOT, AND CEP DATA

This section details how, in the current implementation, the Federated Data Integration Framework orchestrates data interaction and semantic alignment across ERP, BOM/MBOM, IoT, and CEP systems. It illustrates how FDIF acts as the central harmonization layer—ingesting, contextualizing, and federating diverse information sources into a unified, standards-based data fabric. Through this integration, FDIF enables seamless interoperability between enterprise planning, product definition, manufacturing execution, and real-time operational intelligence, providing the digital backbone that connects all functional layers within the Smart Manufacturing Network.

6.1. ERP (ENTERPRISE RESOURCE PLANNING) DATA

Concept:

ERP systems manage core business operations — including production planning, resource scheduling, procurement, inventory, and order management.

Function:

They provide structured, transactional data that defines *what should be produced, when, and with what resources* (materials, machines, and labor).

FDIF Interaction:

- ERP exports production orders, resource allocations, and process schedules to FDIF.
- FDIF enriches ERP data with metadata (e.g., context, time, process linkage) and maps it to IPC-2591 Production.WorkOrder structures.
- The data becomes available for synchronization with BOM definitions, IoT sensor feedback, and CEP event streams.

Distinct Role: ERP provides the *planning and coordination layer*, defining manufacturing intent and resource allocation. Within FDIF, ERP acts as the *authoritative source of production logic* and operational reference.

6.2. BOM / MBOM DATA

Concept:

The Bill of Materials (BOM) and Manufacturing Bill of Materials (MBOM) describe *product structure and composition*, defining the hierarchy of components, subassemblies, and materials required for production.

Function:

BOM represents the *design view* of the product, while MBOM defines the *manufacturing view,* including routing, process dependencies, and assembly order.

FDIF Interaction:

- FDIF ingests BOM and MBOM data to construct **ProductData.Definition** and **ProcessData.Operation** entities aligned with IPC-2591 standards.
- The framework enriches these structures with contextual metadata and links them with real-time production and quality events.
- Cross-references between ERP work orders and MBOM structures enable *digital continuity* across design, planning, and execution stages.

Distinct Role: BOM and MBOM form the *product and process knowledge backbone* of the federated data mesh, ensuring that all data-driven actions within NARRATE are contextually tied to specific product configurations and lifecycle stages.

6.3 IOT (INTERNET OF THINGS) DATA

Concept:

IoT devices capture *real-time sensor data and machine telemetry* from the production environment.

Function:

They provide dynamic feedback on machine status, environmental conditions, production rates, energy consumption, and quality parameters.

FDIF Interaction:

- IoT gateways stream data to FDIF in near real time.

- FDIF applies *Complex Event Processing (CEP) and semantic annotation* to transform raw readings into **ProcessData.SensorReading** or **Traceability.Event** entities.

- IoT data is aligned with ERP and MBOM entities to close the loop between *planning and execution.*

Distinct Role:

IoT acts as the *real-time operational data source*, enriching the federated mesh with continuous production insights and enabling adaptive, event-driven control mechanisms.

6.4 CEP (COMPLEX EVENT PROCESSING) DATA

Concept:

The CEP engine [8] analyzes and correlates data streams from IoT devices and other systems to detect *patterns, anomalies, and events of interest*.

Function:

It enables proactive decision-making by identifying deviations, disruptions, or opportunities for optimization in real time.

FDIF Interaction:

- CEP integrates into FDIF as an *event-driven intelligence layer*.

- It subscribes to IoT, ERP, and production data streams and publishes *derived event entities* such as alerts, warnings, and predictions.

- These events are structured as Traceability.Event or DisruptionEvent instances within the FDIF ontology.

- The FDIF harmonizes and forwards these to the IMC for further reasoning and response coordination.

Distinct Role:

CEP represents the *cognitive and reactive component* of the federated data mesh, bridging data analysis with operational response, and providing the temporal intelligence needed for real-time resilience and sustainability assessment.

Table-5 shows the FDIF data layers, their functions, the role of FDIF and relevant IPC-2591 mappings.

Layer	System	Function	FDIF Role	Output / Mapping
Enterprise Layer	ERP	Production planning, resource allocation	Extracts structured business logic	IPC-2591 Production.WorkOrder
Product Layer	BOM/MBOM	Product & process definition	Maps to semantic product structures	IPC-2591 ProductData.Definition
Operational Layer	IoT	Real-time data capture	Enriches with contextual metadata	IPC-2591 ProcessData.SensorReading
Cognitive Layer	CEP	Event correlation & pattern detection	Generates derived events	IPC-2591 Traceability.Event
Federation Layer	FDIF	Semantic harmonization & interoperability	Integrates and aligns all domains	Feeds Blueprint Frames & IMC

TABLE 6 FDIF DATA LAYERS, THEIR FUNCTIONS, THE ROLE OF FDIF AND RELEVANT IPC-2591 MAPPINGS.

7. METADATA TAGGING MECHANISM

FDIF uses a *metadata tagging schema* (implemented via JSON-LD, XML, or RDF) that assigns *standardized tags* to each data element. Tags are defined by referencing ontology vocabularies and open manufacturing standards, enabling interoperability.

Understanding how the Federated Data Integration Framework (FDIF) harmonizes heterogeneous data sources in NARRATE is a crucial part of D3.6. Below we unpack in clear detail *how JSON-LD metadata tagging enables semantic interoperability* between ERP, BOM/MBOM, IoT/CEP, and other manufacturing data domains.

7.1. THE ROLE OF JSON-LD IN FDIF

When FDIF receives data from ERP, BOM, MBOM, IoT, or CEP sources, it:

- **Ingests and normalizes** the data (syntactic level).
- **Annotates** each entity with JSON-LD context definitions drawn from domain ontologies (semantic level).

- **Maps** these annotations to recognized standards — such as:
- **IPC-2591 (CFX)** for electronics manufacturing interoperability.
- **AAS (Asset Administration Shell)** for Industry 4.0 asset models.
- **Publishes** the harmonized data for use by the NARRATE Blueprint Frames, enabling analytics, traceability, and event-driven control.

JSON-LD (JavaScript Object Notation for Linked Data) is a lightweight semantic web format that allows machine-readable metadata to be embedded directly into data records.

In the FDIF Harmonization Layer, JSON-LD tags provide *semantic context* and *ontology references*, enabling diverse data (from ERP, BOM, IoT, etc.) to be interpreted consistently across the Smart Manufacturing Network (SMN).

FDIF uses JSON-LD for:

- **Data Categorization** — identifying what kind of manufacturing data it is (e.g., production order, bill of materials, sensor event).
- **Functional Role Mapping** — showing what function the data serves (e.g., scheduling, traceability, quality control).
- **Context Linking** — linking data entities to manufacturing ontologies (e.g., IPC-2591).
- **Relationship Encoding** — expressing how one dataset relates to another (e.g., "this production order consumes this material lot").

Below are simplified JSON-LD fragments showing how FDIF annotates various manufacturing data types and can apply metadata tagging to different types of manufacturing data (ERP, BOM, MBOM, and CEP). Each example uses semantic annotations to capture *data category, context, functional role,* and *relationships,* making them interoperable and mappable to standards like IPC-2591 or Asset Administration Shell (AAS) models.

7.2. ERP DATA — *PRODUCTION ORDER*

Purpose: Represents an ERP-originated production order annotated for use in FDIF.
FDIF Role: Business and scheduling data harmonized with shop-floor execution and IoT data.

```
{
 "@context": "https://schema.narrate.ai/fdif",
 "@UseCaseId": "MEDWOOD Pilot",
 "@type": "ERPProductionOrder",
 "orderId": "PO-2025-1178",
 "category": "ProductionOrder",
 "functionalRole": "Scheduling",
 "productCode": "COT-DELUXE-01",
 "quantity": 150,
 "status": "Planned",
 "dueDate": "2025-11-15",
 "metadata": {
  "sourceSystem": "SAP-ERP",
  "createdBy": "Planner#42",
```

```
  "plant": "Plant-A",
  "timestamp": "2025-10-20T10:15:00Z",
  "unit": "pcs",
  "version": "1.2"
 },
 "relatedBOM": "bom:BOM-1458",
 "tags": ["ERP", "WorkOrder", "Scheduling", "Production"]
}
```

Semantic mapping:

- @type identifies the object as an *ERP Production Order*.

- Context links (erp, schema) connect it to a vocabulary recognized by FDIF.

- erp:relatedBOM allows semantic linking to a Bill of Materials, enabling AAS mapping to the *Production Order* submodel.

IPC-2591 mapping: Production.WorkOrder message.

Meaning of "tags"

In the FDIF context, *tags* are lightweight semantic labels that describe the *context, origin, and functional intent* of a data entity. They serve several complementary purposes:

Contextualization:

Tags help classify data instances (like this production order) in terms of their domain and role within the manufacturing ecosystem.

- For example, "ERP" indicates the data's source system,

- "WorkOrder" links it to a process or operational activity,

- "Scheduling" identifies its functional purpose, and

- "Production" ties it to the manufacturing domain.

Discovery and Filtering:

Within the FDIF, tags allow other components (e.g., the Blueprint Management System) to *search, filter, and subscribe* to specific data types without hard-coding system dependencies. For instance, a data consumer might query all entities tagged as ["WorkOrder", "Production"] to retrieve only production-relevant orders, regardless of their origin system.

Semantic Federation Aid:

Tags contribute to the *semantic alignment* across distributed systems. They provide a simple yet consistent classification mechanism that complements the deeper semantic annotations (e.g., JSON-LD context and type definitions). This makes it easier for FDIF's harmonization and orchestration layers to reason about and match data entities originating from different schemas or platforms.

Who Creates the Tags

Tags are typically assigned automatically by the FDIF ingestion or harmonization layer, though they may also be defined or overridden by data engineers or ontology designers during the configuration phase. Specifically:

- Initial tagging may occur at the data ingestion level: When FDIF receives input from ERP, BOM, IoT, or CEP systems, predefined mapping rules associate system fields with a set of standard tags. *(e.g., all ERP-originated work orders are automatically tagged ["ERP", "WorkOrder", "Scheduling"]).*

- Users or integrators may extend or modify tag sets to support specific pilots, use cases, or analytics functions, ensuring they remain interoperable with the federated data model.

7.3. BOM DATA — *PRODUCT DEFINITION AND STRUCTURE*

Purpose: Encodes a Bill of Materials with material composition and hierarchical structure.
FDIF Role: Provides the *product configuration backbone* for Digital Twins and traceability.

```
{
  "@context": "https://schema.narrate.ai/fdif",
  "@type": "BOMStructure",
  "bomId": "BOM-COT-DELUXE-01",
  "category": "ProductComponent",
  "functionalRole": "AssemblyStructure",
  "product": {
    "name": "Deluxe Baby Cot",
    "revision": "3.1",
    "material": "MixedWood-Aluminium"
  },
  "components": [
    {
      "componentId": "COT-FRAME-001",
      "quantity": 1,
      "unit": "set",
      "material": "Aluminium",
      "supplierId": "SUP-AL-007"
    },
    {
      "componentId": "COT-RAIL-002",
      "quantity": 2,
      "unit": "pcs",
      "material": "BirchWood",
      "supplierId": "SUP-WOOD-012"
    }
  ],
  "metadata": {
    "sourceSystem": "SiemensTeamcenter",
    "createdBy": "DesignEngineer#14",
    "timestamp": "2025-10-18T14:25:00Z"
  },
  "tags": ["BOM", "ProductDefinition", "Traceability", "Material"]
}
```

IPC-2591 mapping: ProductData.Definition message.

7.4.   MBOM DATA — *MANUFACTURING PROCESS DEFINITION*

Purpose: Represents a manufacturing Bill of Materials linking components to specific operations.
FDIF Role: Integrates product structure with process and resource data for execution planning.

```
{
 "@context": "https://schema.narrate.ai/fdif",
 "@type": "MBOMProcessPlan",
 "mbomId": "MBOM-COT-DELUXE-01",
 "category": "ManufacturingProcess",
 "functionalRole": "ProcessPlanStep",
 "operations": [
  {
    "operationId": "OP-001",
    "description": "Laser Cutting Frame Components",
    "workcenter": "LaserCutter#2",
    "cycleTime": "180s",
    "inputComponents": ["COT-FRAME-001"],
    "output": "CutFrameSubassembly"
  },
  {
    "operationId": "OP-002",
    "description": "Surface Finishing and Assembly",
    "workcenter": "AssemblyLine#5",
    "inputComponents": ["CutFrameSubassembly", "COT-RAIL-002"]
  }
 ],
 "metadata": {
  "sourceSystem": "MES-FDIFConnector",
  "plant": "Athens-Plant-A",
  "timestamp": "2025-10-20T09:45:00Z"
 },
 "tags": ["MBOM", "ProcessDefinition", "Execution", "Manufacturing"]
}
```

IPC-2591 mapping: ProcessData.Operation message.

7.5.   CEP DATA — *REAL-TIME EVENT*

Purpose: Represents a detected complex event (pattern) from real-time IoT and sensor data.
FDIF Role: Supports predictive analytics and adaptive control in the Smart Manufacturing Network.

```json
{
  "@context": "https://schema.narrate.ai/fdif",
  "@type": "ComplexEvent",
  "eventId": "EVT-OVERHEAT-03",
  "category": "ProcessEvent",
  "functionalRole": "PredictiveAlert",
  "description": "Overheating trend detected in Laser Cutter #2",
  "origin": "CEP-Engine@FDIF",
  "eventPattern": "Temperature > 80°C sustained for >60s",
  "relatedSensor": "TEMP_SENSOR_02",
  "severity": "High",
  "confidence": 0.91,
  "timestamp": "2025-10-20T10:22:00Z",
  "metadata": {
    "correlationId": "PO-2025-1178",
    "affectedOperation": "OP-001",
    "source": "FDIF-CEP",
    "responseRequired": true
  },
  "tags": ["CEP", "EventDriven", "PredictiveMaintenance", "IoT"]
}
```

IPC-2591 mapping: Traceability.Event message or custom event schema.

Once these JSON-LD tagged records are federated, FDIF can:

- **Align semantics** across systems using shared vocabularies and ontologies.

- **Auto-generate AAS submodels** for assets, processes, or events.

- **Integrate with IPC-2591 CFX messages** for equipment-level interoperability.

- **Feed harmonized data into the IMC analytics engine** for predictive, event-driven decision-making.

8. PROOF OF CONCEPT IMPLEMENTATION

The Federated Data Integration Framework Proof-of-Concept (PoC) demonstrates the practical application of the FDIF Prototype Implementation Framework through a *FastAPI-based microservice implementation* [9]. The goal of this PoC is to validate how heterogeneous industrial data—particularly from IoT sources—can be *harmonized, semantically enriched, validated, and mapped* into a standardized Digital Twin Blueprint model.

The FDIF builds upon the IPC-2591 (Connected Factory Exchange – CFX) standard and semantic web technologies (RDF/OWL, JSON-LD, SHACL) to ensure interoperability, data traceability, and digital continuity across the manufacturing and engineering lifecycle.
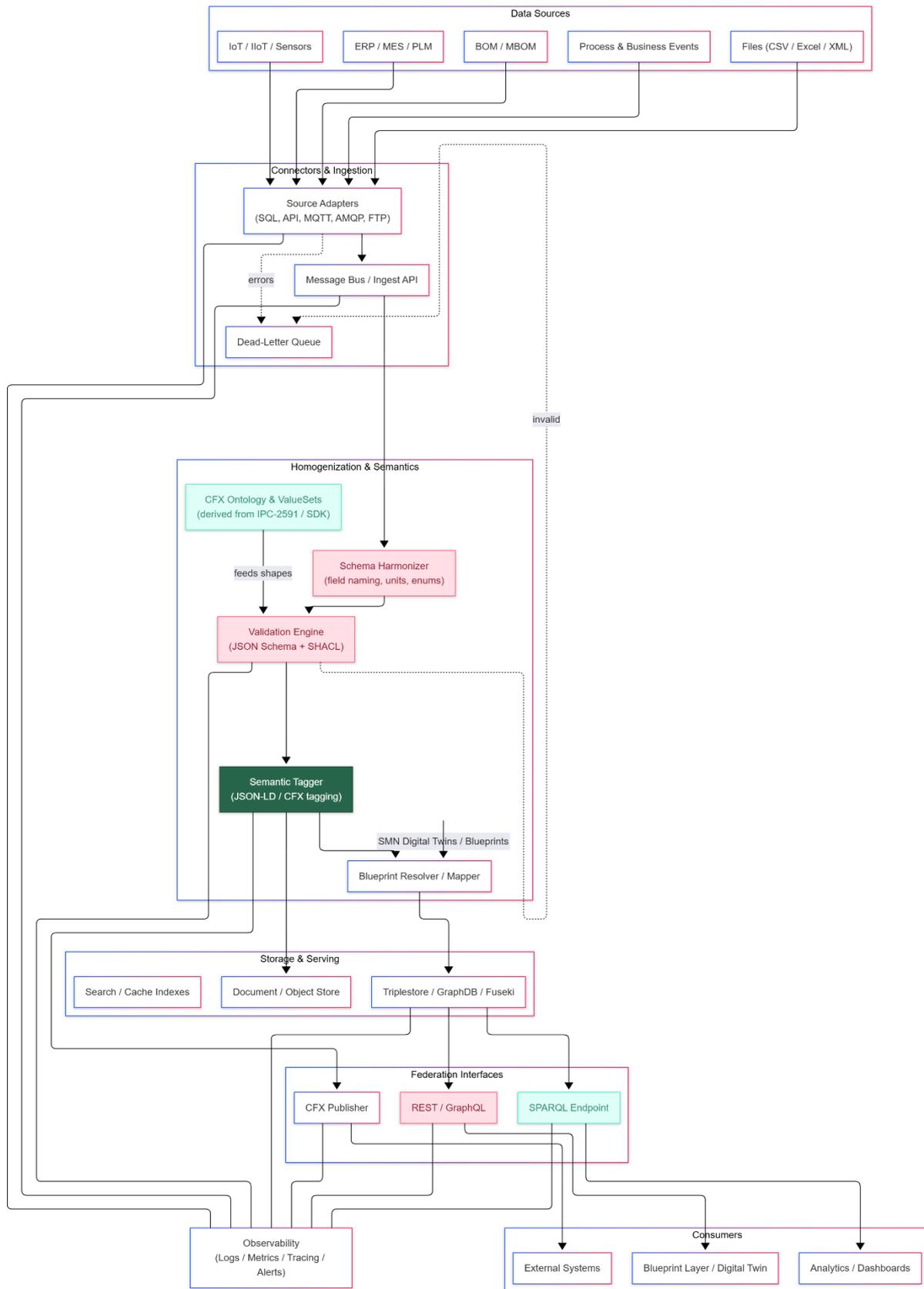
FIGURE 5 PROTOTYPE IMPLEMENTATION

8.1. PROTOTYPE IMPLEMENTATION FRAMEWORK

The prototype implementation framework is composed of *seven conceptual layers*, each responsible for a distinct phase in the data lifecycle—from acquisition to harmonization, validation, semantic enrichment, persistence, and consumption. These layers are depicted in Figure-5 and described below.

8.1.1 Data Sources Layer

The Data Sources layer defines all external systems that act as primary data providers. It captures the heterogeneity of industrial and enterprise environments, where data originates from structured systems, event streams, and unstructured documents as shown in Figure-5.

- **ERP / MES / PLM Systems:** Business and manufacturing systems containing production orders, maintenance logs, and product lifecycle information.

- **BOM / MBOM Repositories:** Define product composition and manufacturing hierarchies essential for engineering traceability.

- **IoT / IIoT / Sensors:** Generate continuous operational data reflecting machine status, process parameters, and environmental metrics.

- **Process & Business Events:** Capture transactional and workflow activities within enterprise systems.

- **Files (CSV, Excel, XML):** Represent legacy or manual data exchanges that complement digital pipelines.

This layer acts as the **data origin**, continuously feeding the ingestion layer through adapters and connectors.

8.1.2 Connectors & Ingestion Layer

The Connectors & Ingestion layer manages data acquisition from heterogeneous sources.
It provides protocol adapters, message handling, and basic reliability mechanisms to ensure consistent and lossless ingestion.
- **Source Adapters:** Interfaces supporting various protocols (SQL, API, MQTT, AMQP, FTP) to connect with source systems.

- **Message Bus / Ingest API:** A communication backbone for streaming and batch data ingestion, standardizing transport and buffering flows.

- **Dead-Letter Queue (DLQ):** Captures malformed or un-processable records for later inspection and reprocessing.

This layer enables data continuity and resilience, ensuring that all incoming data is received, logged, and traceable.

8.1.3 Homogenization & Semantics Layer

The Homogenization & Semantics Layer forms the core intelligence of the FDIF. Its purpose is to transform raw, ingested data into harmonized, semantically enriched entities that adhere to shared ontologies and standardized schemas.

- **Schema Harmonizer:** Normalizes data structures, field names, units of measurement, and enumerations across all data sources to ensure structural consistency.

- **Validation Engine:** Executes a two-tier validation process:

  - Structural validation (via JSON Schema) to verify completeness, data types, and structural correctness.

  - Semantic validation (via SHACL) to ensure conformance with ontology-driven constraints.

- **Ontologies & Value Sets:** Leverages IPC-2591 (CFX) classes and extended SKOS vocabularies to guarantee that every attribute is associated with a clearly defined semantic domain.

- **Semantic Tagger:** Converts validated data into JSON-LD format, embedding semantic context and linking attributes to their corresponding ontology URIs.

- **Blueprint Resolver / Mapper:** Aligns harmonized data with domain blueprints or Digital Twin templates from the SMN Digital Twins/Blueprints repository, enabling traceable and semantically consistent relationships across the enterprise.

In essence, this layer ensures that all data entering the federation is clean, intelligible, semantically coherent, and fully interoperable.

How Semantic Validation Is Performed
1. SHACL Shape Application (Core Mechanism)
Each entity type (e.g., Product, Operation, Machine, Material Lot) has an associated SHACL shape that defines:
- Required properties

- Data type expectations

- Cardinality (e.g., must have exactly one serial number)

- Relationship constraints (e.g., an Operation must reference a valid ProcessStep class)

- Controlled vocabulary bindings (e.g., acceptable state values)

Semantic validation applies these SHACL shapes to the incoming RDF graph (derived from JSON input) to verify that the data conforms to all semantically defined constraints.
Outcome:
The system identifies:
- Missing required attributes

- Incorrect value domains

- Invalid relationships

- Violations of ontology rules

---

2. Ontology Binding & Class Compliance
Incoming data elements are checked against the IPC-2591 (CFX) reference ontology and any extended Simple Knowledge Organization System (SKOS) vocabularies to ensure:
- Entity types map to valid ontology classes

- Attributes correspond to defined object or data properties

- Units, states, and value sets match the allowed semantic domains

Examples:

- A machine state must belong to the SKOS "MachineState" concept scheme.

- A process step identifier must map to a CFX ProcessStep class.

- A temperature property must use a unit compatible with the ontology definition.

Outcome:

Ensures that data "means what it is supposed to mean" according to the standardized domain model.

Why SKOS Is Used in the FDIF Context

SKOS is ideal for modeling ValueSets and enumerations such as:

- A unit of meaning (e.g., MachineState, DefectType, MaterialCategory).

- Machine Mode (STARTUP, RUNNING, MAINTENANCE, ERROR)

- Defect Categories

- Process Step Types

- Unit of Measure Lists

- Quality Status Codes (GOOD, REWORK, SCRAP)

3. Cross-Entity Relationship Validation

Semantic rules also ensure that **references across entities** follow valid ontology-based relationships.

Examples:

- A MaterialLot *must* be produced by an Operation.

- A ProductInstance *must* belong to a specific ProductDefinition.

- A SensorObservation *must* reference a defined Sensor instance with a compatible capability.

This prevents structurally valid but semantically nonsensical data from entering the system.

4. Value Set & Enumeration Compliance

The layer verifies that all coded values match the **ValueSets** derived from the ontologies.

Examples:

- "UnitOfMeasure" must be from the UoM SKOS scheme.

- "DefectType" must match a concept in the DefectTaxonomy.

- Enumerations such as "GOOD / FAIL / REWORK" must match the prescribed CFX classes.

Outcome:

Eliminates semantic drift caused by inconsistent coding conventions.

5. Contextual and Conditional Constraints

Advanced SHACL rules (SHACL-AF) enable **conditional logic**, such as:

- If operationType = "soldering", then a temperature property is required.

- If machineMode = "maintenance", then productionRate must not be provided.

6. End Result

After semantic validation:

- The data is **logically coherent**, not just structurally correct.

- All attributes and relationships comply with the shared semantic model.

- Violations are reported early, preventing bad data from contaminating the federation.

- The data can safely be converted into **JSON-LD** and merged with Digital Twins and Blueprint frames.

8.1.4 Storage & Serving Layer

The Storage & Serving layer provides federated persistence and optimized access for different types of data. It supports both **semantic querying** and **document-level retrieval**.

- **Triplestore / GraphDB / Fuseki:** Stores ontology-based triples for reasoning and SPARQL-based queries.

- **Document or Object Store:** Maintains raw or enriched JSON data, binary files, and sensor payloads.

- **Search / Cache Indexes:** Support fast retrieval, keyword search, and real-time analytics acceleration.

This layer ensures data durability, scalability, and semantic accessibility for upper layers.

8.1.5 Federation Interfaces Layer

The Federation Interfaces layer exposes harmonized data to external consumers and partner systems. It acts as the unified access layer that abstracts the internal data complexity.

- **REST / GraphQL APIs:** Allow developers and systems to query harmonized data models in a structured, human-readable way.

- **SPARQL Endpoint:** Enables semantic queries directly over the ontology, supporting advanced reasoning and inferencing.

- **CFX Publisher:** Disseminates standardized manufacturing data (e.g., IPC-CFX) to Digital Twin environments or downstream integration platforms.

Through these interfaces, the FDIF supports both data federation and semantic interoperability, ensuring that data consumers can access a unified view without centralization.

8.1.6 Consumers Layer

The **Consumers** layer represents the end applications and stakeholders benefiting from the federated data ecosystem.

- **Blueprint Layer / Digital Twin:** Constructs dynamic, model-based representations of physical assets or processes using the harmonized semantic data.

- **Analytics / Dashboards:** Power BI, Grafana, or other tools that consume REST or SPARQL APIs for insights, trend analysis, and KPI monitoring.

- **External Systems:** Include partner systems, suppliers, and external regulatory or enterprise services integrating with federated APIs.

This layer drives *data-driven decision making*, leveraging the federation's harmonized foundation.

8.1.7 Observability Layer

The Observability layer is a cross-cutting concern integrated across all components to ensure transparency, traceability, and reliability.

- **Logs:** Capture every operation and event across adapters, validators, and APIs.

- **Metrics:** Quantify data flow, latency, and system health.

- **Tracing:** Tracks transactions and dependencies across distributed services.

- **Alerts:** Proactively notifies operators about anomalies, ingestion failures, or data quality issues.

- **DLQ Management**: Faulty data records are isolated and reprocessed without interrupting the main pipeline.

This layer guarantees operational continuity across the entire federated data lifecycle.

8.2. ENDPOINT DESCRIPTIONS AND FUNCTIONALITY

The PoC establishes a clear, modular pipeline in which each logical stage of data processing is exposed as an independent *REST API endpoint*. This modularity ensures flexibility for *testing, chaining, validation, and reuse* of individual services, while maintaining interoperability with other enterprise and manufacturing systems (e.g., ERP, BOM, and IoT platforms).

At its core, the implementation operationalizes FDIF's principles of:

- **Semantic Interoperability:** achieved through ontology-based data annotation (JSON-LD, CFX-aligned context).

- **Data Federation:** by integrating diverse, distributed data into a unified, blueprint-ready representation.

- **Reusability and Extensibility:** through open REST endpoints that can be chained into broader workflows or integrated into the Intelligent Manufacturing Custodian (IMC).

The FDIF PoC pipeline is composed of **five sequential endpoints**, each responsible for a distinct transformation stage in the data lifecycle as show in Table-6.

Endpoint	Purpose	Transformation
POST /FDIF/IOT/HARMONIZE	Normalizes raw IoT data fields, units, and structures.	Raw IoT → Harmonized Data

POST /FDIF/IOT/VALIDATE	Enforces schema and ontology constraints; performs unit conversion.	Harmonized → Validated Data
POST /FDIF/IOT/TAG	Adds JSON-LD semantic annotations aligned with CFX ontology.	Validated → JSON-LD Data
POST /FDIF/IOT/MAP	Maps semantic data into a Digital Twin Blueprint representation.	JSON-LD → Blueprint Model
POST /FDIF/IOT/RUN	Executes the complete pipeline (steps 1–4) in one integrated call.	Raw IoT → Final Blueprint

TABLE 7 FIVE SEQUENTIAL ENDPOINTS OF THE FDIF POC PIPELINE.

Each endpoint can be called independently for partial testing or chained for full end-to-end execution. This architecture enables modular deployment, scalability, and adaptability to multiple industrial data domains beyond IoT (e.g., ERP or MES data streams).

/fdif/iot/harmonize – Harmonization of Raw IoT Data

Purpose:
Standardizes field names, data structures, and measurement units from heterogeneous IoT sources to ensure consistent downstream processing.

Example Input:

```
{
 "deviceId": "M-101",
 "TMP": "296",
 "unit": "K",
 "timestamp": "2025-10-24T12:34:56Z",
 "status": "ok",
 "meta": { "locationName": "Plant A", "line": 3 }
}
```

Output (Harmonized):

```
{
 "harmonized": {
  "deviceId": "M-101",
  "temperature": 296.0,
  "temperature_unit": "K",
  "timestamp": "2025-10-24T12:34:56Z",
  "status": "ok",
```

```
    "location": { "site": "Plant A", "line": 3 }
  }
}
```

Key   **Function:**
Creates consistent data naming conventions and structure. Converts irregular raw fields (e.g., TMP) to meaningful standardized attributes (e.g., temperature).

/fdif/iot/validate – Structural and Semantic Validation

Purpose:
Validates harmonized data against the FDIF ontology and enforces structural integrity and type constraints. It also performs unit conversions (e.g., Kelvin → Celsius) for interoperability.

Output Example:

```
{
  "validated": {
    "deviceId": "M-101",
    "temperature": 296,
    "temperature_unit": "K",
    "temperature_celsius": 22.85,
    "timestamp": "2025-10-24T12:34:56Z",
    "status": "ok",
    "location": { "site": "Plant A", "line": 3 }
  }
}
```

Error   **Handling:**
Invalid or missing fields trigger HTTP 422 responses with detailed messages describing the validation errors.

Key   **Function:**
Ensures compliance with expected data semantics, accuracy, and completeness before proceeding to semantic enrichment.

/fdif/iot/tag – Semantic Annotation

Purpose:
Adds semantic metadata using the JSON-LD (Linked Data) format, aligning field meanings with the Connected Factory Exchange (CFX) and FDIF ontologies. This enables machine-readable data interoperability across systems.

Output Example:

```
{
  "jsonld": {
    "@context": {
      "ex": "http://example.org/ontology#",
```

```
    "deviceId": "ex:Machine/id",

    "temperature_celsius": "ex:SensorReading/value",

    "timestamp": "ex:Observation/time",

    "location": "ex:Machine/locatedAt"

  },

  "@type": "ex:Observation",

  "deviceId": "M-101",

  "temperature_celsius": 22.85,

  "timestamp": "2025-10-24T12:34:56Z",

  "location": { "site": "Plant A", "line": 3 }

 }

}
```

Key   Function:
Transforms syntactic data into semantically linked entities, allowing cross-system reasoning, querying, and mapping into Digital Twins.

/fdif/iot/map – Mapping to Digital Twin Blueprint

Purpose:
Translates the semantically annotated data into a Digital Twin Blueprint structure—a unified, twin-ready representation compatible with the Intelligent Manufacturing Custodian (IMC) and Blueprint Management System (BMS).

Output Example:

```
{
  "blueprint": {
    "machine": {
      "id": "M-101",
      "metadata": { "location": { "site": "Plant A", "line": 3 } },
      "sensors": {
        "temperature": {
          "currentValue": 22.85,
          "unit": "C",
          "lastUpdated": "2025-10-24T12:34:56Z"
        }
      }
    }
  }
}
```

Key   **Function:**

Creates a standardized Blueprint Frame, linking sensor readings, metadata, and contextual information into a Digital Twin model ready for analytics, simulation, and real-time monitoring.

/fdif/iot/run – End-to-End Pipeline Execution

Purpose:

Executes the full FDIF transformation chain in a single call—ideal for rapid validation, demonstration, or batch data processing.

Output Example:

```
{
  "step1_raw": {...},
  "step2_harmonized": {...},
  "step3_validated": {...},
  "step4_jsonld": {...},
  "step5_blueprint": {...}
}
```

Key   **Function:**

Provides a comprehensive trace of all transformation steps, ensuring transparency and debuggability across the data processing workflow.

8.3. SUMMARY

The FDIF FastAPI Proof-of-Concept provides a robust validation of the FDIF architecture by demonstrating:

- End-to-end data interoperability from raw IoT input to semantically rich, twin-ready output.

- Modular, API-driven design, enabling seamless integration with external enterprise systems and IMC components.

- Scalability and reusability, as each endpoint represents a reusable service that can evolve independently or be repurposed for other data domains.

This prototype establishes a practical, operational baseline for future extensions of the FDIF within the NARRATE architecture, supporting the harmonization and federation of real-time industrial data streams into Smart Manufacturing Networks.

9. CONCLUSIONS AND NEXT STEPS

Within NARRATE's architecture, the Federated Data Integration Framework serves as a semantic and syntactic data harmonization layer that federates heterogeneous manufacturing information — from ERP, MES, PLM, IoT, and CEP sources. CFX provides the standard message format for FDIF to publish this data in a way that can be universally understood, reused and exchanged.

This deliverable marks a significant step in the evolution of the Federated Data Integration Framework from conceptual design to tangible implementation within the NARRATE architecture. Building on the foundational principles established in D3.1a, the present work has focused on operationalizing the FDIF's federated and semantically enriched data integration capabilities to support seamless interoperability across the Smart Manufacturing Network (SMN).

The implemented framework demonstrates the feasibility of harmonizing heterogeneous data sources — including ERP, BOM/MBOM, IoT, and CEP systems — into a unified semantic model capable of supporting advanced analytics, intelligent orchestration, and Digital Twin synchronization. By incorporating standardized exchange mechanisms such as IPC-2591 (CFX) and adopting JSON-LD-based semantic annotation, FDIF provides a scalable and interoperable foundation for data-driven decision-making and resilience management.

Furthermore, the practical adaptations and design simplifications introduced during this phase have ensured closer alignment with pilot requirements and prototype development efforts. These refinements not only enhanced system performance and adaptability but also reinforced the FDIF's role as a central enabler of interoperability, traceability, and transparency within the NARRATE ecosystem.

9.1 OUTLOOK

The next phase of work will concentrate on extending the FDIF implementation towards full integration with the Blueprint Management System (BMS) and the Intelligent Manufacturing Custodian. This integration will enable real-time data exchange across distributed manufacturing entities, support predictive analytics for disruption detection and response, and facilitate sustainability and circularity monitoring through the Sustainability & Circularity Stress Testing (SCST) Tool. The outcomes of these activities will further consolidate the FDIF's position as a cornerstone of the NARRATE architecture, supporting its overarching goals of resilience, interoperability, and intelligent manufacturing coordination.

REFERENCES

[1] L. Guerreiro Azevedo, E. Figueiredo de Souza Soares, R. Souza and M. Ferreira Moreno (2020), Modern Federated Database Systems: An Overview. Proceedings of the 22nd International Conference on Enterprise Information Systems (ICEIS 2020) – Vol. 1, pp. 276-283

[2] Kaj Burchardi, Dwaipayan Nath and Jiaqi Lan Boston Federated Data Governance Model (2024), Consulting Group, June 2024, https://media-

publications.bcg.com/Federated-Data-Governance-Model.pdf.

[3] Z. Dehghani (2022), Data Mesh: Delivering Data-Driven Value at Scale. O'Reilly 2022.

[4] Global Electronics Association, IPC-2591, Connected Factory Exchange (CFX). https://www.electronics.org/ipc-2591-connected-factory-exchange-cfx.

[5] Platform Industrie 4.0, Specification: Details of the Asset Administration Shell, https://industrialdigitaltwin.org/wp-content/uploads/2022/06/DetailsOfTheAssetAdministrationShell_Part1_V3.0RC02_Final1.pdf

[6] JSON for Linking Data, https://json-ld.org/.

[7] Papazoglou M., Krämer B., Raheem M., & Elgammal A. (2025). Medical Digital Twins for Personalized Chronic Care, 2025, 10.21203/rs.3.rs-5826330/v1 and https://www.researchsquare.com/article/rs-5826330/v1

[8] S. Suhothayan et. al. (2011), Siddhi: a second look at complex event processing architectures. GCE '11: Proceedings of the 2011 ACM workshop on Gateway computing environments, pp. 43-50.

[9] Sildatke, M. et al. (2023) A Distributed Microservice Architecture Pattern for the Automated Generation of Information Extraction Pipelines. SN COMPUT. SCI. 4, 833 (2023). https://doi.org/10.1007/s42979-023-02256-4.