Data Ecosystem Analysis »»»
MANUFACTURING 4.0

A Text-Mining and Topic Modeling Study of Eight European Dataspaces

GROUP 5

Giulio D'Amico - 817781 Daniele Giovanardi - 815601

Filippo Nannucci - 821281 Edoardo Riva - 816511 Gianfranco Votta - 807861

MANUFACTURING & INDUSTRY 4.0

DATA SPACES

Evidence & Insights from European initiatives

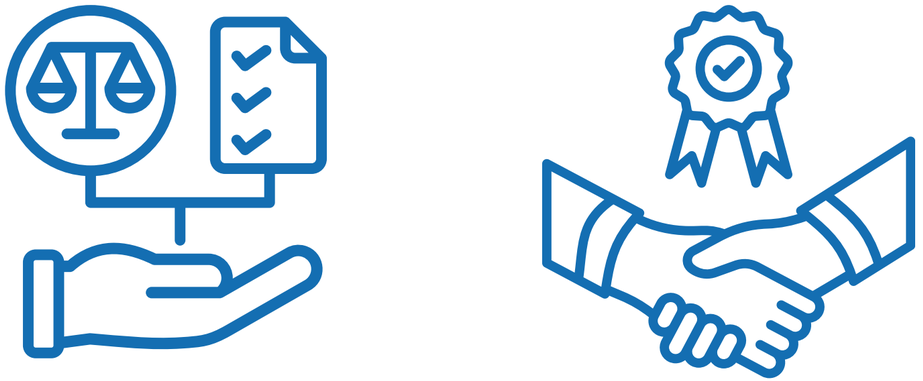
- Focus on how data spaces support Industry 4.0 transformations through **interoperability**, **trust** and **cross-company data exchange**.

- **Analysis of 8 European dataspaces** covering manufacturing, aerospace, occupational health, regional platforms and printing.

- From concrete use cases to architectural patterns and ecosystem-level **recommendations** for **scalable digital manufacturing**.

WHY DATA SPACES MATTER IN MANUFACTURING?

Industrial data is still dispersed across OEMs, suppliers, integrators and service providers. Without a common framework, many high-value scenarios remain constrained: digital twins require **cross-factory consistency**, predictive maintenance depends on **harmonised operational data**, and ESG reporting demands **reliable**, **multi-source inputs**.



Data spaces address this fragmentation by introducing **shared standards**, **trustworthy identity**, **governance models** and **technical connectors** that enable controlled, interoperable data flows across organisational boundaries.

OUR LENS AND METHOD

We analysed **eight dataspaces** using material sourced from project websites, architectural descriptions, governance summaries and actor–role mappings

Selection criteria:

- Relevance to Manufacturing 4.0
- Domain diversity
- Enough English documentation

Manufacturing-X

Smart Factory Web

Sm4rtenance

Pressious

Post Platform

Cooperants

Occupational

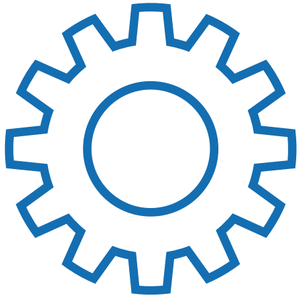
Datahub Tirol

TEXT MINING PIPELINE

SOURCES (Dataspaces)

Project websites, technical descriptions, governance summaries, actor-role matrix

PREPROCESSING

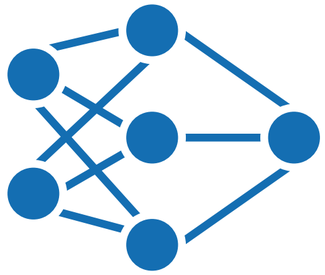


- Lowercasing
- Remove punctuation % numbers
- English stop-words
- Domain-specific stop-words (study, figure, results, example, summary...)
- Tokenisation

VECTORISATION (Bag-of-Words)

CountVectorizer
unigrams + bigrams
min_df / max_df filters
document-term matrix

TOPIC MODELLING (LDA)



Separate LDA per dataspace
3-6 topics
Coherence score
Topic labels

INSIGHTS & MACRO-TOPICS

Federated trust & self-descriptions, technical standards (OPC UA, AAS, IIoT, digital twins) supply/value-chain integration, predictive maintenance & analytics policy & legal frameworks, platform governance & usage control

USE CASE 1: CO_2 TRACKING ACROSS THE AUTOMOTIVE SUPPLY CHAIN (CATENA-X)

Problem

An automotive product is made of thousands of parts from hundreds of suppliers. **CO_2 data** (Product Carbon Footprint - PCF) are scattered across ERP/SAP/Excel systems, **rarely shared and hard to aggregate**. OEMs cannot see end-to-end emissions, certify “low-carbon” products or prove ESG / Green Deal compliance.

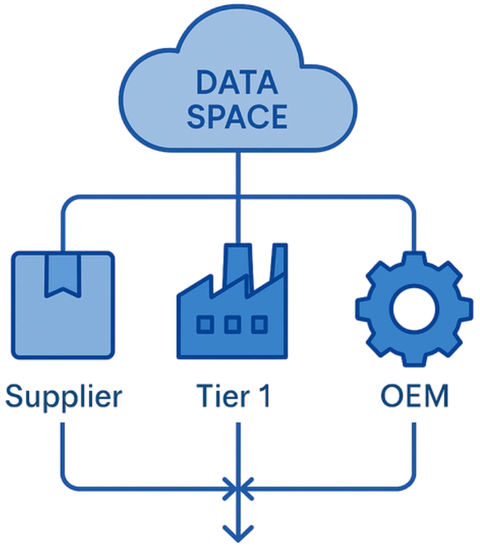
Dataspace solution

Catena-X defines a **standardised way to exchange PCF data across the whole supply chain**. Each company (OEM or supplier) computes its own PCF locally, then publishes it via the data space using **Eclipse Dataspace Connector** (EDC) and machine-readable usage policies.

Self-Sovereign Identity (SSI), Participant Registry and Gaia-X / IDSA frameworks ensure that only authorised partners can access the data and that every exchange is traceable and auditable.

Ecosystem value

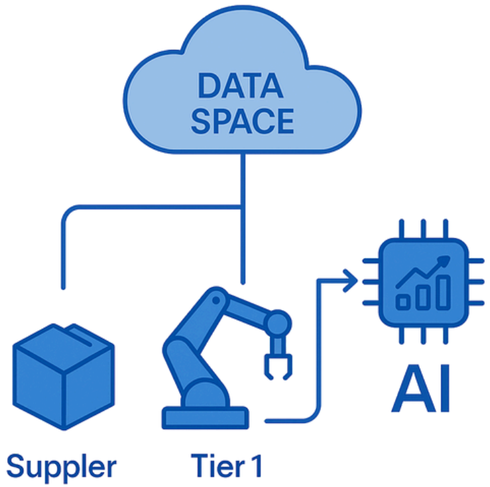
OEMs get an **end-to-end view of CO_2 emissions** and can steer low-carbon sourcing strategies. Suppliers and SMEs can prove transparency and gain a competitive advantage in “green” tenders. **Policy makers and EU gain access to more reliable aggregated data**, reducing greenwashing and supporting Green Deal monitoring.



MAIN TOPICS:

“PCF”, “CO_2”, “EDC”, “SSI”, “supply chain”, standards, federated trust and value-chain integration

USE CASE 2: FEDERATED PREDICTIVE MAINTENANCE (SM4RTENANCE)



MAIN TOPICS:

Reusable blueprint for federated AI services: interoperable standards + trust layer + domain-specific analytics

Problem

Industrial assets generate large volumes of sensor and operational data, but these **data remain siloed in OEM, operators and service providers**. Single-factory datasets **limit the accuracy of predictive models** and make it hard to offer reliable maintenance-as-a-service.

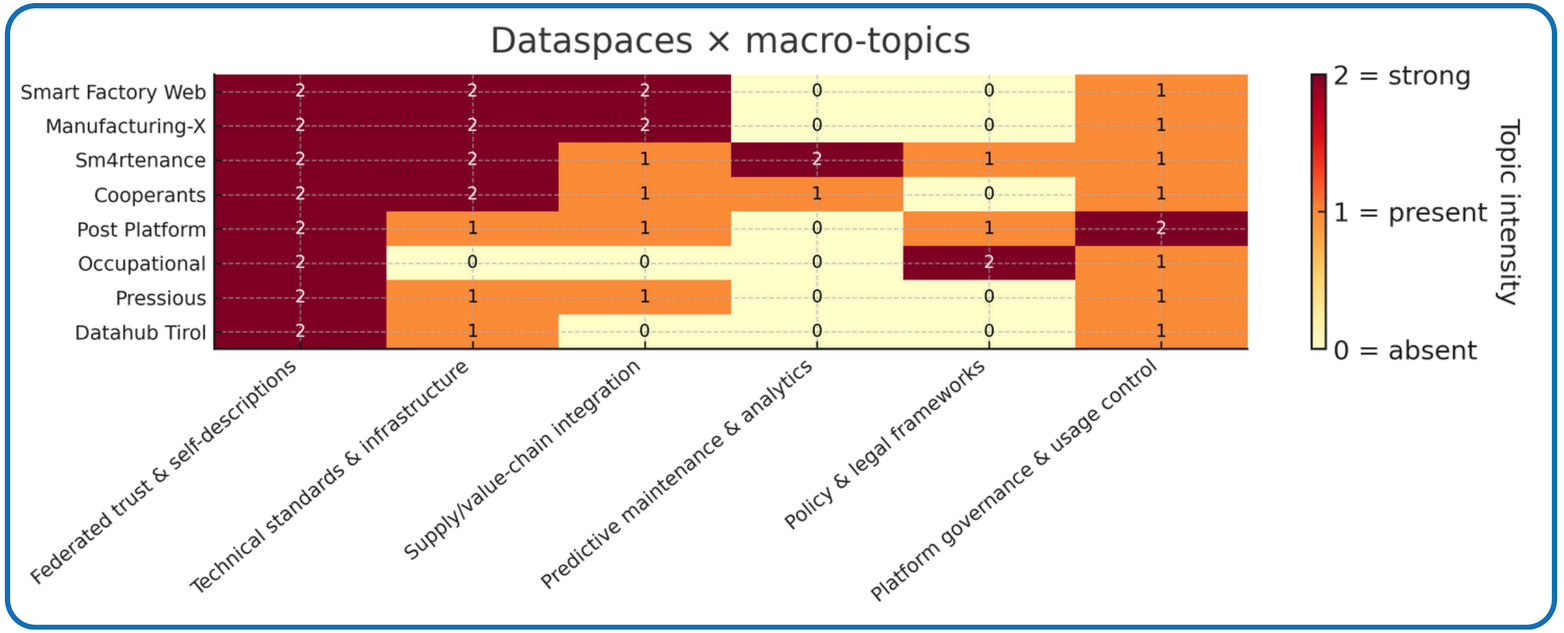
Dataspace solution

SM4RTENANCE builds a **federated data space for maintenance**, using open standards such as AAS, OPC UA and Eclipse Dataspace Connector (EDC). Companies **share selected, policy-governed data** (vibrations, temperatures, maintenance logs, operating conditions) **under Gaia-X / IDSA usage control**, without centralising raw data. A shared AI & analytics layer trains predictive models (RUL, failure risk) across fleets and domains, while Self-Sovereign Identity and trust framework handle identities and access rights.

Ecosystem value

OEMs and operators **reduce unplanned downtime and optimise maintenance schedules**. Maintenance service providers and SMEs can offer **new predictive services built on federated data**. The EU and research partners validate **interoperable architectures** that can be replicated in other Industry 4.0 domains.

Key findings from topic modeling (I)



Across initiatives, **technical standards and federated trust form the most consistent foundation**. Supply- and value-chain topics are strongest in Smart Factory Web and Manufacturing-X, while predictive maintenance and simulation cluster around Sm4rtenance and Cooperants. Occupational is dominated by policy and legal themes, and Post Platform concentrates governance and usage-control topics, with Pressious remaining comparatively under-specified.

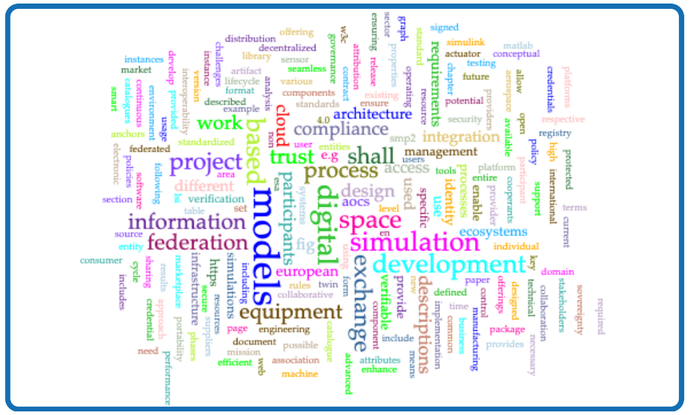
Key findings from topic modeling (II)

Common vs. distinctive vocabulary

- **Common language**: terms such as “digital”, “European”, “data space” and “services” appear across almost all documents. This confirms that the **dataspaces share a common narrative of EU-level digital transformation** rather than isolated projects.

- **Manufacturing-specific**: words like “manufacturing”, “Industry 4.0”, “factory” and “supply” are concentrated in Smart Factory Web, Manufacturing-X and Sm4rtenance, showing a clear industrial focus.

- **Domain-specific clusters**: Occupational is characterised by “health”, “EHDS”, “GDPR”, while Cooperants shows aerospace terms such as “spacecraft”, “orbit”, “AOCS”. These distinctive vocabularies reflect the specialisation of each dataspace.

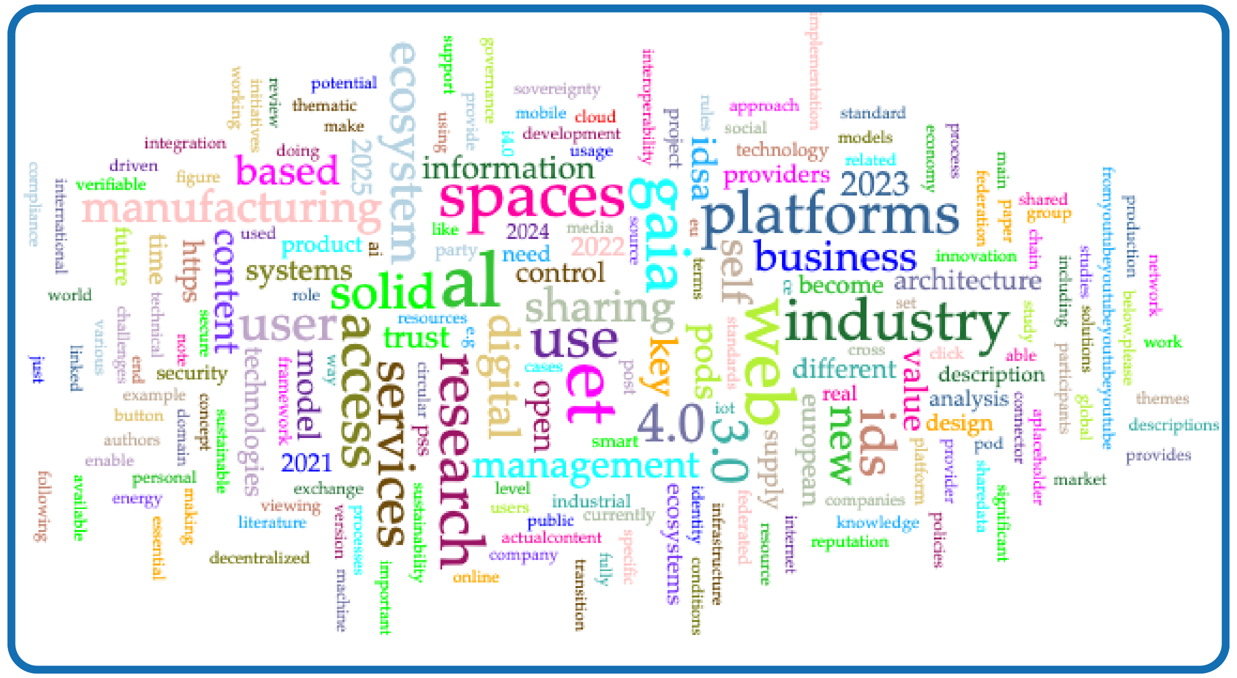


COOPERANTS

MANUFACTURING

OCCUPATIONAL

Key findings from topic modeling (III)



POST PLATFORM

PRESSIOUS

SMART FACTORY

SMARTENANCE

Data ecosystem dynamics

Common building blocks (trust, identity, self-descriptions and usage-control mechanisms) reappear across otherwise diverse sectors. Manufacturing, health, mobility and space ecosystems often share the same architectural base.

On top of this shared foundation, each dataspace introduces its own domain-specific specialisation: occupational health regulations (EHDS, GDPR), aerospace simulation (AOCS), predictive maintenance, regional integration, printing/offset workflows, and more.

Strategic recommendations

SCALING UP

Manufacturing dataspaces advance **more effectively when organisations share connectors**, identity services and semantic assets.

Testbeds and sandboxes help validate **interoperability**, and **clear onboarding paths** allow SMEs to participate without excessive technical burden.

MITIGATING RISKS

Clear role allocation (between orchestrators, operators, participants and regulators) **reduces uncertainty**. Alignment with GDPR, EHDS and sector-specific rules is essential. Operational trust grows when usage-control and audit mechanisms are implemented concretely, not abstractly.

DATA GOVERNANCE VISION

A **layered architecture** offers the most robust path: a technical foundation, a federated trust layer, and domain-specific services on top.

Modularity enables reuse across dataspaces while supporting specialised, high-value capabilities.

THANK YOU

GROUP 5