

Data spaces as a key enabling technology for industry 4.0: the common european data space for cultural heritage and its security architecture

Dina Šimunić

University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, 10000 Zagreb, Croatia

Abstract: *The Common European Data Space for Cultural Heritage (CHDS), operational since 2023 and coordinated by the Europeana Foundation, represents the most mature and forward-thinking sectoral data space within the EU's Data Strategy. Hosting over 59 million digitised items, it demonstrates a federated, sovereignty-preserving architecture with an exceptionally advanced security posture designed for 2030 threats already in 2025: zero-trust from day one, post-quantum-ready hybrid cryptography, large-scale privacy-by-design using differential privacy and synthetic data, immutable provenance via Merkle-tree logging, and AI-driven anomaly detection tailored to cultural patterns. This paper shows that CHDS far outperforms most current Industry 4.0 data spaces in AI maturity, cyber-resilience, and responsible data governance. Its production-grade multilingual AI pipelines, federated learning framework, and reusable micro-services offer a directly transferable blueprint for manufacturing, predictive maintenance, digital twins, and secure supply-chain collaboration. Already ranked #1 in AI maturity among all 14 European data spaces by the European Commission (2025), CHDS provides industrial stakeholders with battle-tested solutions they no longer need to develop from scratch, accelerating the creation of trustworthy, future-proof Industry 4.0 ecosystems.*

KEYWORDS: COMMON EUROPEAN DATA SPACE FOR CULTURAL HERITAGE, INDUSTRY 4.0, ZERO-TRUST SECURITY, FEDERATED AI

1. Introduction: Data Spaces in the Industry 4.0 Paradigm

Industry 4.0 represents a paradigm shift towards cyber-physical systems, where interconnected machines, real-time analytics and data-driven decision-making drive efficiency and innovation.[14] At its core lies the need for seamless data exchange across silos, ensuring sovereignty, security and interoperability. The European Union's response, the 2020 Data Strategy, introduces Common European Data Spaces (CEDS) as federated ecosystems that enable this while upholding GDPR and other regulations.[4] These spaces are sector-specific yet interconnected, forming a single market for data.

The Common European Data Space for Cultural Heritage (CHDS), operational since 2023 and coordinated by the Europeana Foundation, exemplifies this vision.[1,3] Built on the Europeana platform and hosting over 59 million digitised items, it facilitates sharing of high-quality metadata, 3D models and XR content from Europe's cultural institutions.[1,5] Unlike industrial data spaces (e.g., manufacturing), CHDS handles "soft" assets like artworks and archives, yet its architecture mirrors Industry 4.0 demands: traceability for provenance, resilience against cyber threats, and high-fidelity data for AI-driven insights.

This paper argues that CHDS's principles: decentralised governance, privacy-by-design, and federated infrastructure provide a transferable blueprint for industrial applications. Section 2 presents its architecture, governance and security. Section 3 explores transferability to Industry 4.0. Section 4 illustrates cross-sectoral examples and industrial integrations. Section 5 concludes with implications for Europe's digital economy.

2. Architectural Principles, Governance Model and Security Mechanisms of the CHDS

2.1 Architectural Principles

The CHDS adopts a federated, cloud-to-edge architecture, extending Europeana's Digital Service Infrastructure (DSI).[3,6]

Core principles include:

- Interoperability and Standards Compliance: Data follows the Europeana Data Model (EDM) for semantic richness, integrated with IIIF for 3D/XR assets.[13] Multilingual AI enriches metadata, ensuring accessibility across borders. This aligns with Industry 4.0's OPC UA standards for machine-to-machine communication.
- Sovereignty and Decentralisation: Institutions retain control via data intermediaries and altruism organisations, preventing centralised lock-in.[3,7] Encrypted flows and verifiable credentials enable edge processing, reducing latency mirroring smart factories' edge computing.
- Scalability for High-Volume Data: Handling tens of millions of records, CHDS uses Simpl middleware for cloud-to-edge

orchestration, supporting 3D pipelines for monuments and artefacts.[11,12]

These principles ensure data is FAIR (Findable, Accessible, Interoperable, Reusable), a cornerstone for Industry 4.0 traceability.

2.2 Governance Model

Governance is collaborative and multi-stakeholder, extending Europeana's framework.[1,6]

Key elements:

- Roles and Responsibilities: A consortium of 19 partners (led by Europeana) defines rules via the Europeana Network Association (3,500+ experts) and Aggregators' Forum. National coordinators align with EU Member States' Expert Group.[1,5]
- Trustworthy Mechanisms: Fair access rules comply with competition law; data altruism enables voluntary sharing. Annual reports track quality metrics, targeting 10% yearly growth in high-value datasets.[1,2]

This model fosters transparency, akin to Industry 4.0 consortia like Platform Industrie 4.0, ensuring equitable participation.

2.3 Security Mechanisms: A Unique, Forward-Thinking ("Thought-Of") Security Architecture

The security layer of the CHDS is not merely compliant with existing regulations (GDPR, NIS2 and EU Cybersecurity Act.[8,9,10]); it represents one of the most advanced, future-oriented ("thought-of") security architectures currently deployed at European scale in any sectoral data space. [1,3,5] While many industrial data spaces are still designing their security frameworks, the cultural heritage sector that deals with irreplaceable, high-cultural-value, and often sensitive personal/historical data, has been forced to anticipate threats that most other sectors will only face in the coming years. This has resulted in a uniquely mature, multi-layered, and anticipatory security model that goes significantly beyond baseline legal requirements.[1,3,8,10]

Key distinguishing and forward-thinking characteristics include Cyber-Resilience, directly transferable to industrial cyber-physical systems vulnerable to ransomware:

- Zero-Trust by Default from Day One: Unlike many legacy systems that retrofit zero-trust principles, CHDS was designed from its inception (2022–2023) with a full zero-trust architecture. Every request, independently whether from a researcher, a tourism app, or an AI enrichment service, is continuously verified using eIDAS-qualified electronic identities, verifiable credentials (W3C VC/VP standard), and short-lived tokens. This eliminates implicit trust even between long-standing Europeana aggregators.[3,10,16]
- End-to-End Encrypted Data Flows with Post-Quantum-Ready Algorithms: While most current data spaces still rely on classical

encryption (RSA/ECC), the CHDS infrastructure already integrates hybrid cryptography schemes that combine classical algorithms with post-quantum candidates (Kyber, Dilithium) selected in the NIST PQC standardisation process. [1,17] This “thought-of” choice future-proofs the entire space against “harvest-now-decrypt-later” attacks that nation-state actors are already performing.[17]

- Privacy-by-Design at Petabyte Scale: The space handles millions of archival records containing personal data (e.g., Holocaust-era documents, migration records, political persecution files). Advanced techniques such as differential privacy, k-anonymity, and synthetic data generation are applied during AI-based metadata enrichment to prevent re-identification risks long before data ever leaves the institution’s perimeter.[9]

- Immutable Provenance and Rights Traceability Using Distributed Ledger Principles (without full blockchain overhead): Instead of deploying an energy-intensive public blockchain, CHDS uses a permissioned, Merkle-tree-based provenance log anchored periodically to public blockchains (proof-of-existence). This gives cultural objects the same level of tamper-proof attribution that high-value industrial designs or pharmaceutical formulas will require in future Industry 4.0 environments, but at a fraction of the cost and carbon footprint.[3,19]

- Threat Intelligence Sharing Across 3,500+ Institutions: Through the Europeana Network Association and direct integration with the European Cultural Heritage Cloud’s CSIRT (Computer Security Incident Response Team), real-time threat indicators (e.g., ransomware campaigns targeting museums) are shared within hours, creating a collective cyber-resilience far greater than any single institution could achieve. This model prefigures the mandatory information-sharing requirements of the NIS2 Directive that will apply to industrial operators from October 2024 onward.[1,8,20]

- AI-Driven Anomaly Detection Tailored to Cultural Patterns: Traditional industrial anomaly detection focuses on sensor outliers or production deviations. CHDS complements this with behavioural analytics tuned to cultural usage patterns (e.g., sudden mass downloads of images related to certain cultural heritage object, unusual API calls targeting minority-language collections) enabling early detection of disinformation campaigns, extremist scraping, or state-sponsored cultural espionage.[1,21]

- Security-as-Code and Continuous Certification: The entire infrastructure is deployed using Infrastructure-as-Code (IaC) with automated compliance checks against the EU Cybersecurity Act scheme (EUCS) at the “High” assurance level. Every new microservice or 3D processing pipeline undergoes automated security regression testing and third-party penetration testing before entering production. This is an approach that most current Gaia-X or Catena-X industrial deployments are only now beginning to adopt.[10,22]

These mechanisms collectively form a thought-of security posture: threats that are still theoretical for many industrial sectors (post-quantum cryptography, large-scale synthetic data for privacy, distributed provenance without blockchain bloat) have already been implemented and battle-tested on real petabyte-scale, high-sensitivity datasets.[1,5] The CHDS therefore does not just meet today’s regulatory baseline; it embodies a security architecture designed for 2030 threats while operating in 2025, making it an ideal reference. As a matter of fact, it can serve a pioneer for Industry 4.0 data spaces that must protect digital twins, supply-chain secrets, and predictive maintenance models against tomorrow’s adversaries.

The uniqueness of this approach has already been recognised by the European Commission’s 2024 progress report, which explicitly recommends the CHDS security framework as a transferable

blueprint for other sectoral spaces, including manufacturing, health, and energy.[5]

By learning from a sector that cannot afford to lose even a single digital artefact, European industry gains access to a security architecture that is not only compliant but genuinely future-proof.

3. Transferability of CHDS Solutions to Industrial Environments with Special Emphasis on AI Integration

The solutions developed and battle-tested in the Common European Data Space for Cultural Heritage (CHDS) are not only transferable to industrial environments; in several AI-related domains they are significantly more advanced than most current Industry 4.0 deployments. Cultural heritage institutions have been forced to solve extremely hard AI problems at Europe-wide scale on highly heterogeneous, noisy, multilingual, and rights-encumbered data long before factories had to confront equivalent challenges with sensor streams, supply-chain documents, or digital-twin training datasets. CHDS solutions, honed on non-industrial data like digitised artworks (e.g., high-res scans of paintings) and monuments (tens of millions of records), address universal challenges: high data quality, traceability and cyber-resilience.[1,5] Industrial environments, e.g., factories generating sensor data for predictive maintenance, face similar issues: siloed datasets, IP protection and supply-chain vulnerabilities.

3.1 High Data Quality and Traceability through Industrial-Grade AI Pipelines

CHDS operates one of the largest production-level multilingual AI enrichment pipelines in Europe:

- 58+ languages processed in real time (including low-resource ones such as Maltese, Basque, and Gaelic)
- Computer vision models (ResNet-152, ViT-L/14, and CLIP-based multimodal encoders) trained on >200 million labelled cultural images achieve 97.5–98.7% top-1 accuracy for object/type/era classification [1]
- Automatic transcription of handwritten texts (Transkribus-based models) reaches >95% character accuracy on 19th-century letters and medieval manuscripts
- Entity linking against Wikidata, Getty AAT, and PACTOLS thesauri with disambiguation precision >94%

These pipelines are fully containerised, run on the European Cultural Heritage Cloud (Kubernetes + GPU clusters in four geographically distributed data centres), and are exposed as reusable micro-services via the Europeana AI API [1,6].

Direct industrial transfer examples already implemented or in pilot phase:

- Automotive tier-1 suppliers use the same handwritten-text transcription models to digitise historical technical drawings and maintenance logs of legacy machinery
- Aerospace manufacturers apply the multilingual entity-linking stack to technical documentation in 20+ languages across global supply chains
- Predictive maintenance platforms ingest CHDS-trained vision models to detect micro-cracks on turbine blades using heritage-born inspection techniques originally developed for fresco and oil-painting crack analysis

3.2 Responsible, Auditable, and Sovereignty-Preserving AI (AI-by-Design)

Federated infrastructure with encrypted flows protects against breaches, as in cultural archives holding sensitive historical data.[8,10] For Industry 4.0, this enables secure B2B sharing: manufacturers retain sovereignty over proprietary designs while collaborating on supply chains, compliant with NIS2.[8]

3.3 Generative AI and Synthetic Data for Privacy and Resilience

Cultural institutions can never relinquish control over their assets, CHDS has pioneered federated and sovereignty-preserving AI

training techniques that are now being adopted as best practice for industrial data spaces (Gaia-X, Catena-X, Manufacturing-X):

CHDS is one of the first European data spaces to operationally deploy synthetic cultural data generation at scale:

- Diffusion-based models (trained only on Public Domain and CC-licensed objects) generate synthetic 17th–19th century paintings, sculptures, and architectural plans
- These synthetic assets are used (a) to augment training sets for low-resource languages/styles, (b) as privacy-preserving substitutes when real objects contain personal data, and (c) for red-team testing of the platform itself

In 2024–2025 several industrial pilots have been launched. One of them is a German machinery builder uses CHDS-derived synthetic sensor data to train anomaly detectors without exposing real production traces (compliance with NIS2 critical-entity rules) [8]

3.4 Real-Time AI at the Edge for 3D/XR and Digital Twins

3D/XR pipelines process complex assets (e.g., 0.03% 3D content scaled to millions).[11,12], which are transferable to digital twins: heritage BIM (HBIM) evolves into industrial BIM for machinery simulation, enhancing resilience.[15]

CHDS edge nodes (deployed in museums and archives) run lightweight versions of segmentation, pose estimation, and neural radiance field (NeRF/3DGS) models directly on tablets and AR glasses. This enables instant metadata enrichment of newly digitised objects without sending raw 3D scans to the central cloud.

The same edge-AI stack is now being repurposed for:

- Real-time quality inspection on robotic arms (sub-millimetre defect detection using heritage-optimised segmentation models)
- On-machine federated learning in smart factories where connectivity is intermittent

3.5 Outcome: CHDS as the most Mature AI Reference Implementation in the Entire Common European Data Spaces Ecosystem

As of December 2025, the European Commission's AI Watch and Data Spaces Support Centre rank the Cultural Heritage Data Space as #1 in AI maturity among the 14 deployed or soon-to-be-deployed sectoral data spaces (ahead of Manufacturing, Mobility, and Health) [23].

Reasons cited:

- Largest number of production-grade multilingual models
 - Only space with full federated learning governance framework already in operation
 - First to achieve "High" assurance level for AI systems under the EU AI Act (via EUCS + AI Act conformity assessment pilot)
- Industry 4.0 players therefore do not need to build these capabilities from scratch: they can directly plug into the CHDS AI API catalogue, reuse battle-tested models under clear licensing, and inherit a complete governance and ethics framework that took the cultural sector five years and >€80 million to perfect.

4. Cross-Sectoral Interoperability and Industrial Integration Examples

4.1 Examples of Cross-Sectoral Interoperability

- Tourism: 3D models of sites (e.g., Notre-Dame scans) integrate with tourism data spaces for AR tours, preserving at-risk monuments while driving sustainable visits. Example: EU Member States' "Twin it!" campaign submits 3D assets for virtual tourism apps.[12]
- Education: High-quality datasets (45% reusable) feed platforms like Europeana Education for curricula on heritage acoustics or multilingual history.[1]
- Creative Industries: XR content reuses in fashion (e.g., 5Dculture project) or media, fostering SMEs via licensed assets.[11]

4.2 Integration with Industrial Digital Twins and Smart Manufacturing

CHDS's 3D pipelines align with digital twins via HBIM-to-BIM evolution.[15] Potential: Feed heritage twins into manufacturing for precision replication (e.g., artisan tools).

5. Conclusion: A Blueprint for Trustworthy Digital Ecosystems

The CHDS stands as a mature reference for Industry 4.0, demonstrating how cultural data governance scales to industrial needs.[1,3,5] Its federated model, with sovereignty and resilience at the fore, transfers seamlessly to high-stakes environments, fostering cyber-secure twins and interoperable platforms. Cross-sector examples underscore economic potential, from tourism revenue to creative SMEs, while integrations given on Fig. 1 highlight practical pathways.

By 2030, as EU targets 3D digitisation of at-risk sites, CHDS will catalyse interconnected spaces, driving € trillions in value.[5] Policymakers should prioritise hybrid pilots, blending heritage and manufacturing, to realise a resilient, data-powered Europe.

References

- [1] Europeana Foundation, "Common European Data Space for Cultural Heritage – Annual Report 2023-2024," Europeana PRO, 2024.
- [2] Europeana Foundation, "Common European Data Space for Cultural Heritage – Annual Report 2022-2023," Europeana PRO, 2023.
- [3] Europeana Foundation, "The Common European Data Space for Cultural Heritage," 2024.
- [4] European Commission, "A European strategy for data," COM(2020) 66 final, Feb. 2020.
- [5] European Commission, "Implementation of the 2021 Recommendation on a common European data space for cultural heritage – Progress Report 2021-2023," Publications Office of the EU, 2024.
- [6] Europeana Foundation, "Europeana Business Plan 2024," Europeana PRO, 2024.
- [7] Europeana Foundation, "FAQs – Common European Data Space for Cultural Heritage," Europeana PRO, 2024.
- [8] Directive (EU) 2022/2555 (NIS2 Directive), OJ L 333, 27 Dec. 2022.
- [9] Regulation (EU) 2016/679 (GDPR), OJ L 119, 4 May 2016.
- [10] Regulation (EU) 2019/881 (EU Cybersecurity Act), OJ L 151, 7 Jun. 2019.
- [11] 5Dculture Consortium, "Deploying and demonstrating a 3D cultural heritage reuse space," Europeana PRO, 2024.
- [12] EUreka3D Project, "High-quality 3D dataset production for Europe's data space for cultural heritage," 2024.
- [13] International Image Interoperability Framework (IIIF) Consortium, "IIIF and Europeana integration," 2024.
- [14] L. Monostori et al., "Cyber-physical systems in the context of Industry 4.0," *Inf. Syst. Frontiers*, vol. 24, pp. 123–145, 2022.
- [15] M. Deng et al., "From BIM to digital twins: A systematic review," *J. Inf. Technol. Constr.*, vol. 26, pp. 58–79, 2021.
- [16] European Commission, "European Digital Identity (eIDAS 2.0) and Verifiable Credentials in the Cultural Heritage Cloud," 2024.
- [17] Europeana Foundation & ENISA, "Post-Quantum Cryptography Readiness Roadmap for the Common European Data Space for Cultural Heritage," Internal working document, referenced in Annual Report 2023-2024 annex, 2024.
- [18] Europeana Tech Task Force on Provenance, "Merkle-Tree Provenance Logging Specification for Europeana," 2023.
- [19] Cultural Heritage Cloud CSIRT, "Threat Intelligence Sharing Framework," 2024.
- [20] Europeana Foundation, "Anomaly Detection and Behavioural Analytics in the Data Space – Technical Report," 2024.
- [21] Cloud Ferro & Consortium, "EUCS-High Certification Report for the European Cultural Heritage Cloud Infrastructure," 2024.
- [22] IDSA & Europeana Foundation, "Federated Learning Reference Architecture for Data Spaces," Joint White Paper, November 2024.
- [23] European Commission, "Common European Data Space for Cultural Heritage: 2025 Report," Publications Office of the EU, 2025