

A conceptual architecture for Digital Twins in critical infrastructure: Trust and predictive resilience

Daniela TIMISICĂ^{1,2*}, Paul NICULESCU-MIZIL GHEORGHE¹

¹ National Institute for Research and Development in Informatics – ICI Bucharest, Romania

² National University of Science and Technology POLITEHNICA Bucharest, Romania

daniela.timisica@ici.ro (*Corresponding author), paul.gheorghe@ici.ro

Abstract: Critical infrastructure increasingly integrates operational technology, information technology, Internet of Things (IoT) devices and data-driven decision support systems. This convergence creates opportunities for monitoring and prediction while increasing dependence on data quality, cybersecurity, model credibility and institutional accountability. This paper proposes a Digital Twin reference architecture, structured into seven layers, covering physical sensing and actuation, edge validation, operational data management, digital representation, Artificial Intelligence (AI), decision support, and institutional governance and human oversight. A transversal blockchain-based trust layer asynchronously anchors selected hashes and metadata outside the time-critical control path. The study uses a design-oriented conceptual synthesis of scientific publications, technical guidance and regulatory sources. The architecture is examined through comparison with representative frameworks, a requirement-based coverage assessment and a threat-based walkthrough informed by external sources. A hypothetical water-distribution scenario illustrates the incident-to-decision workflow using synthetic measurements and an illustrative score of 0.72, which is not a calibrated probability of water loss. The proposed architecture links data, model and decision provenance while separating operational control, analytical support and evidence preservation under explicit human authority. These design provisions offer a conceptual basis for predictive resilience in critical infrastructure; their operational benefits remain to be established through implementation and empirical evaluation.

Keywords: Digital Twins, Critical infrastructure, Artificial Intelligence, Internet of Things, Blockchain.

O arhitectură conceptuală pentru gemeni digitali în infrastructuri critice: încredere și reziliență predictivă

Rezumat: Infrastructurile critice integrează din ce în ce mai mult tehnologia operațională, tehnologia informației, dispozitivele internetului obiectelor (IoT) și sisteme de asistare a deciziilor bazate pe date. Această convergență creează oportunități pentru monitorizare și predicție, sporind totodată dependența de calitatea datelor, securitatea cibernetică, credibilitatea modelelor și responsabilitatea instituțională. Lucrarea propune o arhitectură de referință pentru gemeni digitali, structurată pe șapte straturi, care acoperă achiziția datelor prin senzori și acționarea asupra sistemului fizic, validarea la marginea rețelei, gestionarea datelor operaționale, reprezentarea digitală, inteligența artificială (AI), asistarea deciziilor, precum și guvernanța instituțională și supravegherea umană. Un strat transversal de încredere bazat pe blockchain ancurează asincron anumite valori hash și metadata, în afara fluxului de control cu cerințe temporale critice. Studiul utilizează o sinteză conceptuală orientată spre proiectare, bazată pe publicații științifice, ghiduri tehnice și surse de reglementare. Arhitectura este examinată prin comparație cu arhitecturi reprezentative prin evaluarea gradului de acoperire a cerințelor și prin analiza unor scenarii de amenințare, pe baza surselor externe. Un scenariu ipotetic de distribuție a apei ilustrează fluxul de la incident la decizie, utilizând măsurători sintetice și un scor ilustrativ de 0,72 care nu reprezintă o probabilitate calibrată a pierderii de apă. Arhitectura propusă corelează proveniența datelor, a modelelor și a deciziilor, separând controlul operațional, suportul analitic și păstrarea dovezilor, sub autoritate umană explicit definită. Aceste mecanisme oferă o bază conceptuală pentru reziliența predictivă a infrastructurilor critice; beneficiile lor operaționale urmează să fie stabilite prin implementare și evaluare empirică.

Cuvinte-cheie: gemeni digitali, infrastructuri critice, inteligență artificială, internetul obiectelor, blockchain.

1. Introduction

Critical infrastructure underpins essential social and economic functions across energy, transport, finance, digital infrastructure, drinking water and wastewater, information and communication technology (ICT) service management, health, space and public administration. Its

digitalisation integrates operational technology (OT), information technology (IT), Internet of Things (IoT) devices and data-driven decision-support systems, improving situational awareness, simulation and maintenance planning while expanding the attack surface and the number of technical dependencies (Stoleriu et al., 2025). In the European Union, drinking water and wastewater are sectors of high criticality under the NIS2 Directive, while the Critical Entities Resilience Directive provides a framework for strengthening the resilience of entities delivering essential services (European Parliament and Council of the European Union, 2022a, 2022b).

The need for resilience is particularly evident in water infrastructure. ENISA reports comparatively low cybersecurity maturity in the drinking water and wastewater sectors, constrained by limited resources, sectoral heterogeneity, legacy systems and largely reactive practices. These sectors nevertheless increasingly depend on integrated OT, IoT, Industrial Internet of Things (IIoT) and IT environments for monitoring, control, data collection and predictive maintenance (ENISA, 2026).

Digital Twins link observations of physical assets and processes with contextual data, engineering models and simulation capabilities. Their value extends beyond static digital models when the representation remains synchronised with the physical system and supports monitoring, prediction and controlled intervention (Kritzinger et al., 2018; Fuller et al., 2020). However, a Digital Twin does not inherently establish trust. Faulty sensors, outdated state information, incomplete topology, or compromised analytical components may produce coherent but operationally incorrect representations. Security, credibility and trust must therefore be addressed throughout the Digital Twin lifecycle (Voas et al., 2025).

Artificial Intelligence (AI), blockchain and IoT address different aspects of this problem. IoT supplies measurements and device context; AI supports pattern recognition, risk estimation and prediction; the blockchain ecosystem can provide tamper-evident references for selected evidence. Their responsibilities must remain distinct. In particular, blockchain-based evidence mechanisms should remain outside time-critical control loops, where transaction and validation latency could impair operational responsiveness. The central problem is therefore how to integrate these capabilities while preserving data validity, model provenance, human decision authority, auditability and operational latency.

Previous studies have examined AI-supported Digital Twins, IoT-based water management, blockchain-enabled Digital Twins and integrated smart-water architectures. Mohammed et al. (2024), for example, combine Digital Twin technology, a hybrid CNN-LSTM model and blockchain-based validation within an Industrial Internet of Water Things architecture. The present study adopts a different approach by explicitly separating time-critical operational processing, AI-supported analysis, blockchain-supported evidence management and institutionally authorised decision-making.

Predictive resilience is defined here as the capacity to anticipate credible degradation scenarios, assess alternative responses and preserve essential services through controlled, transparent and accountable adaptation. It requires more than predictive accuracy: the quality and provenance of input data, the applicable model and uncertainty, the responsible decision authority and the resulting action must remain identifiable and traceable. Accordingly, the architecture addresses operational visibility, measurement validation, prediction, scenario evaluation, defence-in-depth protection, traceability and human-in-the-loop authority over consequential interventions.

The principal contribution is a seven-layer reference architecture with a transversal blockchain-based trust layer and an explicit separation of operational, analytical, evidentiary and decision-making flows. It is examined through an illustrative water-distribution use case linking anomaly validation, Digital Twin localisation, AI-supported risk assessment, scenario comparison, human authorisation, physical action and audit evidence. The study follows a design-oriented conceptual synthesis of peer-reviewed research, technical guidance and European regulatory sources. The resulting requirements are assessed through the scenario and a qualitative traceability matrix. No deployed prototype, operational dataset, or algorithmic performance is claimed, and all numerical scenario values are illustrative.

The remainder of the paper is organised as follows. Section 2 describes the methodological approach and design procedure. Section 3 presents the technological foundations and positions the proposed architecture against representative prior frameworks. Section 4 introduces the integrated architecture and its requirements. Section 5 applies the architecture to an illustrative water-distribution scenario. Section 6 examines cybersecurity, trust, human oversight and representative threat conditions. Section 7 provides the requirement-based qualitative assessment. Section 8 discusses limitations and broader design trade-offs, while Section 9 concludes the paper and outlines future research.

2. Methodological approach

This study develops a conceptual reference architecture through a design-oriented synthesis of scientific publications, technical guidance and regulatory sources concerning Digital Twins, IoT, AI, cybersecurity, provenance and critical-infrastructure governance. The cited sources were checked for accuracy and relevance and used to identify architectural concerns and inform the proposed allocation of responsibilities. The study does not constitute a systematic literature review or an empirical performance evaluation.

Seven design requirements organise the synthesis and represent the authors' architectural interpretation of concerns identified in the cited sources. Operational visibility, predictive capability and pre-intervention scenario evaluation draw on the monitoring, simulation and anticipated-state functions discussed by Voas et al. (2025). Measurement validity and consistency respond to their concerns about synchronisation, temporal accuracy and the propagation of erroneous data. Defence-in-depth cyber resilience translates their discussion of protection across instrumentation, communications, data and digital twin components into a cross-layer requirement (Voas et al., 2025). Traceability and human accountability are informed by the documentation, clearly assigned responsibilities and human oversight provisions of the AI Risk Management Framework (NIST, 2023). In the proposed architecture, these concerns motivate linked records of observations, model versions and decisions, together with explicit authority for consequential actions. Each requirement is assigned to the layers responsible for its principal mechanisms, while dependencies on other layers and organisational procedures remain explicit.

The conceptual assessment combines three complementary analyses. First, the architecture is compared with representative prior frameworks to clarify its design emphasis. Second, a requirement matrix examines whether the design specifies a responsible component, a mechanism, an inspectable output and a failure or degraded-response path. Coverage is classified as direct when all four elements are explicitly specified, partial when relevant provisions are incomplete or rely on unspecified external controls, and failed when no adequate mechanism or response is defined. These classifications concern architectural specification, not demonstrated operational effectiveness.

Third, a threat-based walkthrough considers stale or replayed observations, coordinated false data, compromised or drifting analytical models and trust-layer unavailability. For each condition, the analysis identifies the affected functions, proposed response and residual limitation. This complements the requirement matrix but does not constitute independent expert or experimental validation. The water-distribution scenario illustrates the resulting information and decision flows using synthetic values; it provides no evidence of detection accuracy or resilience improvement.

3. Enabling technologies for Digital Twins in critical infrastructure

3.1. Internet of Things

Internet of Things (IoT) connects physical infrastructure to its digital representation through sensing, embedded processing and communication. In water-distribution systems, measurements may include pressure, flow, reservoir levels, equipment status and water quality. However, device identity does not establish measurement accuracy: authenticated sensors may transmit faulty, stale

or manipulated observations. Device security and management should therefore be complemented by freshness, plausibility and cross-sensor consistency checks before measurements are used to update the Digital Twin (Fagan et al., 2021).

3.2. Artificial Intelligence

Artificial Intelligence (AI) supports anomaly detection, forecasting, risk estimation and decision prioritisation, contributing to proactive risk management in critical infrastructure (Botezatu & Ciupercă, 2025). Within the proposed architecture, analytical outputs are associated with their input evidence, model version and relevant uncertainty information. These provisions support review and accountability, consistent with the trustworthiness considerations described in the AI Risk Management Framework (NIST, 2023). AI-generated recommendations inform decision-making, while consequential interventions remain subject to authorisation by designated operators.

3.3. Blockchain technology

Blockchain provides a distributed, tamper-evident record that can support integrity verification and provenance in Digital Twin environments (Suhail et al., 2022a). The proposed architecture retains operational datasets off-chain and asynchronously anchors selected hashes and metadata, including references to observations, model versions and authorised decisions. Hash verification can reveal changes to the referenced records after anchoring, but cannot establish the physical accuracy of the original measurements. Ledger operations therefore complement validation and cybersecurity controls and remain outside the time-critical control path.

3.4. Digital Twin technology

Digital Twins combine observations of physical systems with contextual information and models to support monitoring and analysis. Kritzinger et al. (2018) distinguish digital models, digital shadows and Digital Twins by their degree and direction of automated data exchange, while Fuller et al. (2020) discuss their enabling technologies and implementation challenges. In the proposed architecture, the Digital Twin maintains contextual system state and supports the comparison of potential interventions. Its usefulness depends on data freshness, topology accuracy, model assumptions and cybersecurity throughout its lifecycle (Voas et al., 2025).

3.5. Technological convergence

The architecture assigns complementary responsibilities to these technologies: IoT supplies observations; edge and data services qualify and organise them; the Digital Twin represents system state and supports simulation; AI provides analytical outputs; and authorised operators decide on consequential actions. Blockchain preserves selected evidence asynchronously. This allocation provides the basis for the architecture in Section 4.

3.6. Positioning in relation to representative prior frameworks

The proposed architecture does not claim novelty from the individual use of IoT, AI, blockchain or Digital Twin technologies. Previous research has already combined subsets of these capabilities. The architectural contribution considered here concerns how operational, analytical, evidentiary and institutional responsibilities are separated and subsequently reconnected through traceable decision cases. Table 1 therefore compares the proposed model with representative frameworks that overlap substantially with its technological scope.

Table 1. Positioning of the proposed architecture in relation to representative prior frameworks

Framework	Primary focus	Explicit operational / analytical / evidentiary separation	Institutional decision rights	Linked data–model–decision provenance	Asynchronous evidence / control-path non-interference
Suhail, Hussain et al. (2022)	Trustworthy IIoT data and blockchain-supported Digital Twins	No explicit three-path separation	Not explicitly addressed	Partial – primarily data/source trust	Not explicit
Suhail, Malik et al. (2022), TTS-CPS	Blockchain-supported security and situational awareness for CPS Digital Twins	Partial	Not explicitly addressed	Partial – source integrity and security rules	Partial
Mohammed et al. (2024)	IIoWT interoperability, data standardisation, CNN-LSTM processing and blockchain validation	No explicit separation	Not explicitly addressed	Partial – primarily data validation	Not explicit
Homaie et al. (2026), MOSAIC	Adaptive multi-layer cybersecurity for water-distribution Digital Twins	Partial – adaptive security pipelines	No explicit institutional layer	Partial – security/tamper evidence	Partial – latency-aware rather than explicit asynchronous provenance
Proposed architecture	Governed critical-infrastructure monitoring and decision support	Yes	Yes	Yes – data, model and decision provenance	Yes – evidence layer explicitly outside the time-critical control path

Table 1 shows that the novelty of the proposed architecture lies in the simultaneous adoption of four design commitments: explicit separation of operational, analytical and evidentiary flows; asynchronous provenance mechanisms that do not gate time-critical control; traceable linkage among data provenance, model provenance and decision provenance; and explicit institutional decision rights and human authorisation for consequential actions. These characteristics are individually related to mechanisms present in prior work, but they are treated here as a combined organising principle for critical-infrastructure decision support. The comparison concerns architectural organisation rather than performance superiority.

4. Proposed integrated architecture

4.1. Architectural requirements

The architecture addresses seven requirements: continuous operational visibility; measurement validity and consistency; predictive capability; pre-intervention scenario evaluation; traceability of data, models and decisions; defence-in-depth cyber resilience; and human accountability for consequential actions. These requirements address incomplete or stale system state, unreliable measurements, delayed recognition of degradation, uncertain intervention consequences, incomplete provenance, cyber compromise and ambiguous decision authority. Each requirement is mapped to the layers responsible for its principal mechanisms, with cross-layer dependencies retained explicitly. Section 7 assesses their conceptual coverage and residual limitations.

4.2. Seven-layer reference architecture

Figure 1 presents seven functional layers. The Physical and IoT Layer provides sensing and actuation; the Edge and Gateway Layer checks device identity, freshness and measurement plausibility; and the Operational Data Layer stores validated telemetry and contextual information off-chain. The Digital Twin Layer maintains contextual system state and supports simulation, while

the AI Layer provides detection, forecasting and risk estimation. The Decision-Support Layer presents evidence, uncertainty and response alternatives. The Institutional Governance and Human Oversight Layer assigns decision rights and accountability, requiring human authorisation for consequential interventions.

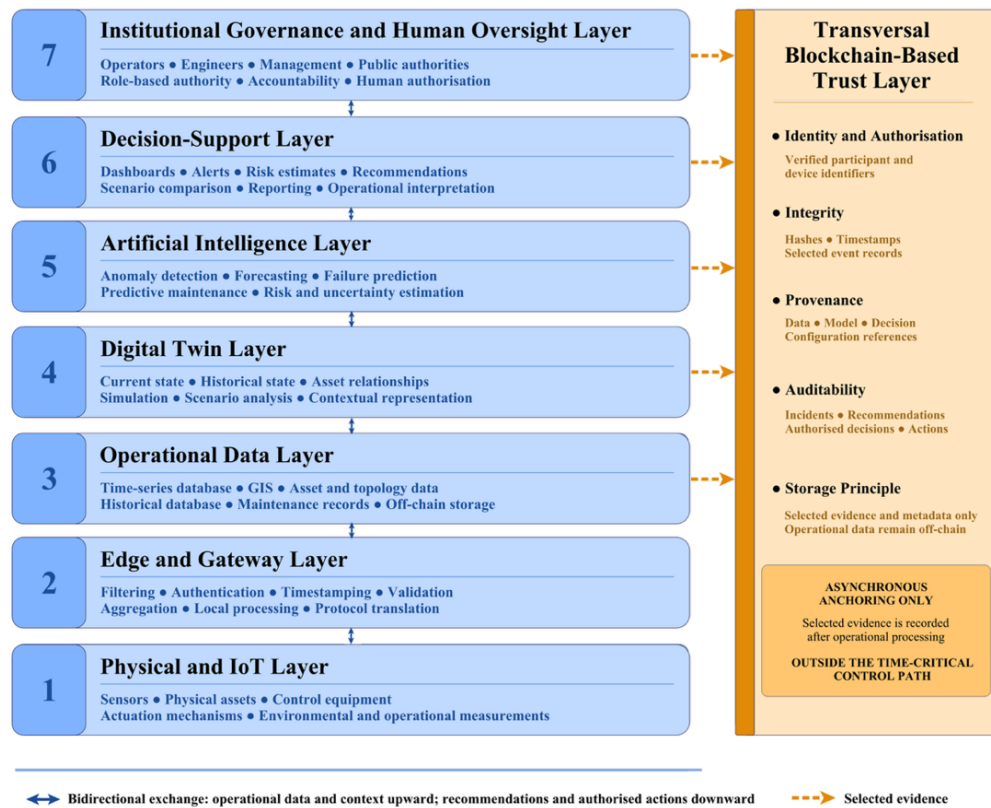


Figure 1. Proposed seven-layer Digital Twin reference architecture for critical infrastructure, including IoT, AI and a transversal blockchain-based trust layer. (Source: Author's own elaboration)

Information exchanged between layers retains source identifiers, timestamps, quality indicators, Digital Twin configuration references and AI model versions. Stale information or unavailable analytical services must be visible to operators, while existing local control and safety mechanisms support degraded operation.

4.3. Blockchain as a transversal trust layer

The blockchain-based trust layer asynchronously anchors selected evidence without delaying time-critical control. Operational and emergency commands must not depend on ledger availability or consensus completion. Raw telemetry and operational records remain in protected off-chain repositories; anchored metadata link event and asset identifiers, timestamps, dataset hashes, Digital Twin configurations, AI model versions and authorised decisions. Hash verification can reveal subsequent changes to referenced records, but cannot establish the physical accuracy of the original observations.

The architecture links data provenance, identifying the observations used; model provenance, identifying the configuration and analytical version; and decision provenance, recording who accepted, rejected or modified a recommendation. Blockchain-supported trust mechanisms in Digital Twin environments have also been examined by Suhail et al. (2022b). Where evidence is shared across identified organisations, a permissioned network may define participation, submission, endorsement and access rights.

Evidence may be signed and batched by incident or time window, retaining the hash algorithm, schema version and off-chain retention reference. Corrections should use linked, append-only records. Continuity procedures must address key loss, participant revocation, ledger

outages and subsequent reconciliation of pending evidence. These provisions complement device authentication, measurement validation, access control, monitoring, backup and recovery.

4.4. Separation of operational, analytical and evidentiary flows

The operational path carries observations, alarms and authorised commands under sector-specific latency constraints. The analytical path supplies validated data to Digital Twin and AI services for detection, prediction and scenario evaluation. The evidentiary path asynchronously anchors selected hashes and metadata without gating operational processing. Shared identifiers link each observation window to the analytical case, Digital Twin configuration, model output, simulated alternatives, authorisation and execution feedback. Incident reconstruction therefore requires consistent timestamps, controlled access to off-chain records and aligned retention periods, rather than duplication of operational datasets on the ledger. Section 5 illustrates how these paths interact while preserving the distinction between recommendations and control authority.

5. Critical water-distribution infrastructure use case

5.1. Motivation for the use case

Water distribution illustrates the interaction between hydraulic dependencies, service continuity and connected monitoring. Digital Twin applications in this sector address leakage, demand, water quality and operational optimisation, with capabilities dependent on data availability and integration (Ghorbani Bam et al., 2025). Their security requires controls across field devices, communications, data and models (Homaie et al., 2026).

Figure 2 illustrates the assessment process and response options rather than the complete network topology. The hypothetical network comprises Reservoir A, Pump P1, Valve V12, pressure sensors P16–P18, flow meter F16 and pipe segment PS-44. Operational context includes demand estimates, pump settings, valve positions, topology and maintenance records. The scenario is conceptual and does not a laboratory experiment, a deployed utility or a hydraulic simulation result.

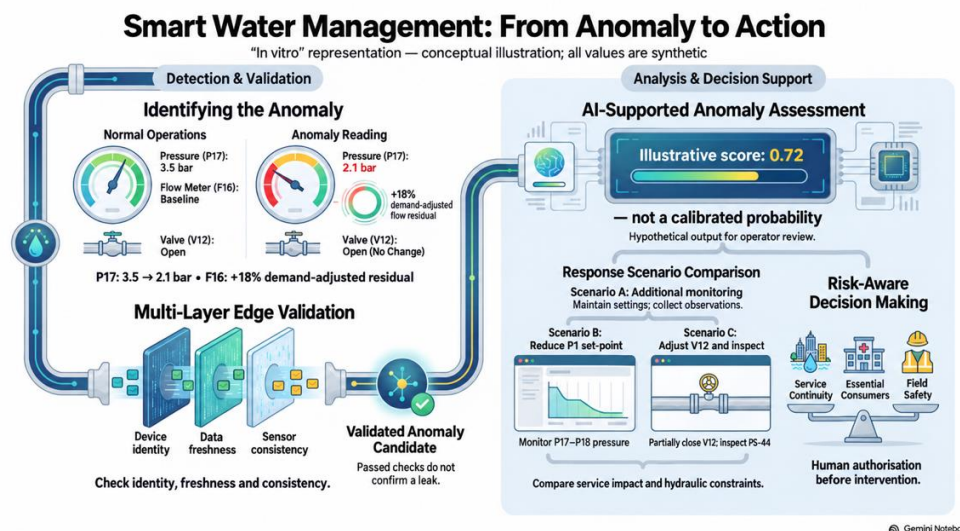


Figure 2. Smart water management: conceptual representation of anomaly assessment and response options. Illustration generated using Google NotebookLM. All numerical values are synthetic

All numerical values are synthetic. Baseline pressures of 3.4–3.5 bar and reduced values of 2.1–2.3 bar illustrate a substantial local deviation, accompanied by a +18% demand-adjusted flow residual. Their interpretation assumes comparable observation times, unchanged pump settings and valve positions, and a demand estimate used as the flow reference. They do not establish hydraulic consistency for a specified network. The value 0.72 is an illustrative score on a notional 0–1 scale, not a calibrated leak probability, confidence estimate or performance measure.

5.2. Event detection and edge validation

During normal operation, P16, P17 and P18 report 3.4, 3.5 and 3.4 bar, respectively. At 10:34, P16 remains unchanged, while P17 and P18 fall to 2.1 and 2.3 bar, respectively, and F16 records a +18% demand-adjusted flow residual. V12 remains open and the P1 set-point is unchanged. The edge gateway checks device identity, message authentication, timestamps, freshness, sensor-health flags, measurement ranges, rates of change and cross-sensor consistency. Observations passing these checks are forwarded as anomaly candidates, without confirming a leak.

Failed authentication or stale messages trigger quarantine. An isolated pressure drop without corroborating observations may instead prompt an instrument-fault investigation. Agreement among pressure and flow measurements supports further assessment of a network disturbance, but does not exclude coordinated false data or shared power and communication failures. Raw measurements, quality flags and rejected observations are retained according to policy for subsequent review.

5.3. Digital Twin synchronisation and AI-based analysis

The Digital Twin maps the validated deviations to the zone around PS-44 and updates its contextual state using available observations, topology and demand estimates. The analysis considers competing explanations, including leakage, unusual demand, valve-state inconsistency, pump disturbance and sensor or communication faults. The AI service examines pressure differences, the flow residual, actuator states and spatial agreement. In this hypothetical scenario, its output is represented by the illustrative score of 0.72, accompanied by model and feature-set versions, contributing observations and an uncertainty notice. The score supports operator review rather than confirming a diagnosis.

The analytical case retains the state-estimation time, available pump and valve observations and the demand profile used. Incomplete topology, inferred actuator states or limited sensor coverage are identified as uncertainty sources. Operators assess the output against alternative explanations and may request additional observations or field inspection before considering intervention.

5.4. Scenario comparison and decision support

A hydraulic model such as EPANET could support comparison of potential interventions before authorisation (Rossman et al., 2020). Table 2 outlines three illustrative options: additional monitoring, reduction of the P1 set-point and staged adjustment of V12 with field inspection. Their expected effects are qualitative assumptions, not simulation results. Decision support should present the supporting observations, alternative explanations and potential consequences for pressure, service continuity, essential consumers and field safety.

In an implementation, scenario outputs should identify the model version, boundary conditions, uncertainty and validity horizon. Changed network conditions or expired outputs require reassessment before action. Each response should define success indicators and rollback or escalation conditions, including pressure behaviour at P17–P18, the F16 flow residual and service conditions in dependent zones.

Table 2. Illustrative response scenarios for the PS-44 anomaly

Scenario	Intervention	Expected effect	Principal constraint
A	Maintain settings briefly while collecting additional observations.	Avoids immediate disruption; increases diagnostic evidence.	Permissible only within a defined risk and time bound.
B	Reduce P1 set-point and monitor P17–P18.	May limit losses and pressure transients.	Can reduce service pressure in dependent zones.
C	Partially close V12 and dispatch a field team.	Localises the suspected segment and supports inspection.	Requires authority, hydraulic constraints and staged execution.

5.5. Human authorisation, operational feedback and blockchain-supported audit

In the hypothetical scenario, an engineer recommends Scenario C after reviewing the evidence, alternatives and model limitations. The authorised operator checks service and safety constraints, records the rationale and approves a staged V12 adjustment and field inspection of PS-44. Subsequent observations support assessment of the intervention and updates to the Digital Twin. Unexpected pressure or flow behaviour triggers rollback or escalation under the applicable procedure.

Observations, simulation files where available, recommendations, decisions and execution feedback remain in protected off-chain repositories. The evidentiary path asynchronously anchors selected hashes and metadata linking these records to the relevant Digital Twin configuration, analytical model and authorised participants. Figure 3 summarises the incident-to-decision workflow.

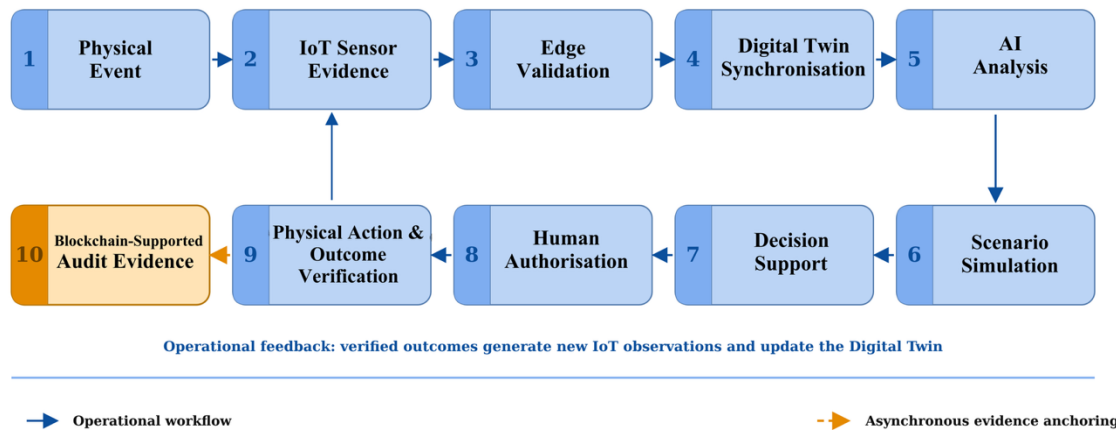


Figure 3. Incident-to-decision workflow within the proposed architecture (Source: Author' own elaboration)

Although audit evidence appears at the end of the diagram, anchoring may occur at multiple stages without delaying operational processing. Failed validation redirects observations for review, uncertain analysis may prompt further investigation, and rejected recommendations produce no physical command. The scenario illustrates the allocation of responsibilities and evidence links; it does not demonstrate improvements in detection accuracy, hydraulic efficiency or service resilience.

6. Security, trust and human oversight

6.1. The role and limitations of blockchain in cybersecurity

Blockchain does not secure sensors, firmware, communication links, credentials, analytical models or industrial controllers. False measurements from a compromised device may be correctly signed and preserved without revealing their physical inaccuracy. The architecture therefore requires defence-in-depth controls, including secure device configuration, authenticated and encrypted communication, network segmentation, least privilege, multi-factor authentication where applicable, monitoring and tested backup and recovery procedures. Ledger-based verification complements these controls rather than replacing them (Voas et al., 2025).

Availability requires defined degraded modes, including local control independent of cloud and ledger services, access to authorised operating procedures and a route to manual operation. Recovery should restore trusted configurations, clock synchronisation, model baselines and the correspondence between physical assets and digital identifiers, as well as restore service availability.

6.2. Representative threats

A threat-based walkthrough complements the requirement matrix by examining representative conditions informed by Digital Twin and critical-infrastructure security literature (Voas et al., 2025; ENISA, 2026; Homaei et al., 2026). For each condition, the analysis considers the affected functions, specified response and residual limitation. It assesses conceptual handling paths rather than measured security performance and does not constitute independent validation.

Stale or replayed observations are addressed through authentication, freshness checks and replay protection at the edge, with failed checks triggering rejection or quarantine before Digital Twin ingestion. Freshness checks alone cannot exclude replays within an accepted time window; implementation therefore requires message identifiers, sequence checks or equivalent controls. Coordinated false data remain partially addressed: authenticated sensors may produce mutually consistent but incorrect observations, requiring additional physical verification when analytical checks are inconclusive.

Model drift or poisoning may be identified or contained through version control, performance monitoring, model-integrity checks and human review, but these mechanisms cannot guarantee detection. During a trust-layer outage, operational and safety functions remain independent of ledger availability, while evidence awaits anchoring in protected off-chain storage. Provenance continuity depends on recovery and reconciliation of pending records. These cases identify proposed responses and unresolved dependencies without demonstrating attack-detection effectiveness or operational non-interference.

6.3. Trustworthy AI-supported recommendations

AI-supported recommendations should identify contributing observations, alternative explanations, model versions, uncertainty, simulation assumptions where applicable and expected service impacts. This information enables authorised operators to accept, modify, reject or defer a recommendation and record their rationale. Lifecycle procedures should define performance monitoring, drift detection, approval of retraining and threshold changes, and safe fallback when analytical services are unavailable. Explainability therefore supports review of a specific decision and its limitations, consistent with AI risk-management principles (NIST, 2023).

6.4. Governance and responsibility allocation

Governance assigns responsibility for data quality, OT configuration, Digital Twin topology, AI model lifecycle and intervention authorisation. Changes to device identities, topology, analytical thresholds and decision rights require approval and versioning. Ledger administration remains separate from authority to approve evidence records or operational actions. Emergency overrides must record the responsible person, rationale, duration and subsequent review. Where one organisation provides data, develops models and operates the infrastructure, an additional authorised role should review high-impact interventions.

7. Requirement-based qualitative evaluation

Table 3 assesses the conceptual coverage of the seven requirements defined in Section 4.1 using the criteria described in Section 2. Direct coverage requires an explicitly identified responsible component, mechanism, inspectable output and failure or degraded-response path. Partial coverage indicates incomplete provisions or dependence on unspecified external controls; failed coverage indicates the absence of an adequate mechanism or response. Because the requirements informed the architecture, the matrix assesses internal consistency rather than independently validating the design. The comparison in Table 1 and the threat-based walkthrough in Section 6.2 provide complementary perspectives, without demonstrating operational effectiveness.

Table 3. Qualitative coverage of the architectural requirements

Requirement	Coverage	Primary layers	Architectural evidence	Residual dependency or risk
Operational visibility	Direct	IoT; edge; data; twin	Timestamped observations and contextual state; stale or unavailable information is flagged, with local control retained	Sensor coverage, communication availability and site-specific freshness limits
Data quality	Partial	Edge; data; twin	Authentication, plausibility and consistency checks produce quality flags; suspect observations are quarantined or referred for review	Calibration error and coordinated false data may remain undetected
Predictive capability	Partial	AI; twin	Versioned analytical outputs retain evidence and uncertainty; suspect outputs prompt review or fallback	Training data, calibration, drift detection and model-specific fallback rules
Scenario evaluation	Partial	Twin; decision support	Response options identify constraints and expected effects; stale or uncertain assessments require reassessment	Hydraulic model fidelity, topology completeness and validated operating limits
Traceability	Direct	Data; trust layer; decision support; governance	Linked observation, model and decision records support reconstruction; ledger outages defer anchoring, followed by reconciliation	Off-chain retention, key management and successful recovery of pending evidence
Cyber resilience	Partial	All layers; transversal trust layer	Segmentation, access control, monitoring support protection; degraded operation and recovery procedures address failures	Legacy systems, control configuration and untested recovery performance
Human accountability	Direct	Decision; governance	Named authority, recorded rationale and approved action establish responsibility; rejection, deferral and emergency overrides remain traceable	Operator competence, workload and enforcement of organisational procedures

The assessment distinguishes specified architectural provisions from outcomes requiring implementation evidence. Direct coverage does not imply that a requirement has been achieved in operation or that residual risks have been eliminated. Each identified dependency should inform future acceptance tests, including data freshness, detection performance, simulation fidelity, recovery behaviour, evidence reconstruction and operator comprehension. No aggregate numerical score is assigned, as the conceptual assessment cannot establish improvements in predictive resilience.

8. Discussion and limitations

In relation to the frameworks discussed by Mohammed et al. (2024), Suhail, Hussain et al. (2022) and Suhail, Malik et al. (2022), the proposed architecture emphasises institutional decision rights, separation of operational and evidentiary processing, and links between data, model and decision provenance. Its contribution concerns the organisation of these responsibilities rather than technological novelty or demonstrated performance superiority. The hypothetical water-distribution scenario illustrates their interaction but cannot establish improvements in detection accuracy, service continuity, recovery time or operator workload.

Implementation depends on legacy-system interoperability, reliable asset identifiers, sensor coverage and calibration, accurate topology and suitable physical models. AI introduces dependencies on representative data, calibration, drift monitoring and explanation quality. Human oversight requires usable interfaces, adequate staffing, training and enforceable procedures. A permissioned ledger adds key management, participant governance, retention and recovery requirements; hashes neither establish measurement accuracy nor automatically anonymise

sensitive information. Utilities with limited cybersecurity maturity may need to prioritise foundational security and recovery capabilities before introducing further technological dependencies (ENISA, 2026).

Deployment should progress from read-only monitoring and advisory analysis to bounded, human-authorised pilots. Evaluation should compare conventional monitoring, Digital Twin and AI support without ledger anchoring, and the complete architecture under sensor faults, false data, communication loss, model drift and operator override. Transfer to other infrastructure sectors requires different physical models, safety constraints and authority rules. The proposed separation of responsibilities provides a design basis for such adaptation, while its practical benefits and costs remain to be tested.

9. Conclusions and future research

This paper proposed a seven-layer Digital Twin architecture for critical infrastructure, complemented by a transversal blockchain-based trust layer. Its contribution lies in the explicit separation of operational processing, analytical support, evidence preservation and institutional decision authority. The architecture links observations, Digital Twin configurations, AI model versions, recommendations and authorised actions, while keeping blockchain anchoring outside the time-critical control path. Human authorisation remains a prerequisite for consequential interventions.

The hypothetical water-distribution scenario illustrated how these responsibilities interact during anomaly assessment, response comparison, intervention and outcome verification. Comparison with representative frameworks, requirement-based assessment and a threat-based walkthrough clarified the design provisions and residual dependencies. These analyses support examination of conceptual plausibility and internal consistency, but do not establish improvements in detection accuracy, service continuity or predictive resilience. Practical effectiveness remains dependent on data quality, model fidelity, cybersecurity controls and organisational procedures.

Future research should implement the architecture in a representative testbed and compare conventional monitoring, Digital Twin and AI support without ledger anchoring and the complete proposed architecture. Evaluation should measure latency, detection and false-alarm performance, model calibration, simulation fidelity, recovery behaviour and operator usability. Particular attention should be given to operational responsiveness and incident reconstruction during sensor faults, communication loss, model drift and ledger outages. These investigations would establish the benefits, costs and conditions under which the proposed architecture could support predictive resilience in operational infrastructure.

Author contributions

Conceptualization: D.T. and P.N.-M.G.; Data curation: D.T. and P.N.-M.G.; Project administration: D.T. and P.N.-M.G.; Supervision: D.T. and P.N.-M.G.; Validation: D.T. and P.N.-M.G.; Writing—original draft: D.T. and P.N.-M.G.; Writing—review and editing: D.T. and P.N.-M.G. All authors have read and approved the version of the manuscript intended for publication.

Submission received: 25 August 2026; Revised: 09 September 2026; Accepted: 14 September 2026; Published: 30 September 2026.

REFERENCES

Botezatu, U. E. & Ciupercă, E.M. (2025) AI-Driven Space Security: Future Trends and Strategic Imperatives for Critical Infrastructures. *Romanian Journal of Information Technology and Automatic Control*. 35(1), 7-22. <https://doi.org/10.33436/v35i1y202501>.

European Parliament and Council of the European Union (2022a) Directive (EU) 2022/2555 on Measures for a High Common Level of Cybersecurity Across the Union (NIS2). *Official Journal of the European Union*. L 333, 80–152.

European Parliament and Council of the European Union (2022b) Directive (EU) 2022/2557 On the Resilience of Critical Entities. *Official Journal of the European Union*. L 333, 164–198.

European Union Agency for Cybersecurity (ENISA) (2026) *ENISA NIS360 2026*. Athens: European Union Agency for Cybersecurity.

Fagan, M., Marron, J., Brady, K.G. et al. (2021) IoT Device Cybersecurity Guidance for the Federal Government: Establishing IoT Device Cybersecurity Requirements. *NIST Special Publication*. 800-213. <https://doi.org/10.6028/nist.sp.800-213>.

Fuller, A., Fan, Z., Day, C. & Barlow, C. (2020) Digital Twin: Enabling Technologies, Challenges and Open Research. *IEEE Access*. 8, 108952–108971. <https://doi.org/10.1109/access.2020.2998358>.

Ghorbani Bam, P., Rezaei, N., Roubanis, A. et al. (2025) Digital Twin Applications in the Water Sector: A Review. *Water*. 17(20), 2957. <https://doi.org/10.3390/w17202957>.

Homaei, M., Di Bartolo, A., Mogollón-Gutiérrez, et al. (2026) Integrated Multi-Layer Security Framework for Water Distribution Digital Twins. *IEEE Access*. 14, 51646–51666. <https://doi.org/10.1109/access.2026.3680021>.

Kritzinger, W., Karner, M., Traar, G. et al. (2018) Digital Twin in Manufacturing: A Categorical Literature Review and Classification. *IFAC-PapersOnline*. 51(11), 1016–1022. <https://doi.org/10.1016/j.ifacol.2018.08.474>.

Mohammed, M.A., Lakhan, A.R., Abdulkareem, K.H. et al. (2024) Industrial Internet of Water Things Architecture for Data Standarization Based on Blockchain and Digital Twin Technology. *Journal of Advanced Research*. 66, 1–14. <https://doi.org/10.1016/j.jare.2023.10.005>.

National Institute of Standards and Technology (2023) *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1. <https://doi.org/10.6028/nist.ai.100-1>.

Rossman, L.A., Woo, H., Tryby, M. et al. (2020) *EPANET 2.2 User Manual*. U.S. Environmental Protection Agency, EPA/600/R-20/133.

Stoleriu, R., Petre, I. & Pop, F. (2025) Cybersecurity Governance in Large-scale Infrastructures. *Romanian Journal of Information Technology and Automatic Control*. 35(1), 51-66. <https://doi.org/10.33436/v35i1y202504>.

Suhail, S., Hussain, R., Jurdak, R. & Hong, C.S. (2022) Trustworthy Digital Twins in the Industrial Internet of Things with Blockchain. *IEEE Internet Computing*. 26(3), 58–67. <https://doi.org/10.1109/mic.2021.3059320>.

Suhail, S., Malik, S.U.R., Jurdak, R. et al. (2022) Towards Situational Aware Cyber-Physical Systems: A Security-Enhancing Use Case of Blockchain-Based Digital Twins. *Computers in Industry*. 141, 103699. <https://doi.org/10.1016/j.compind.2022.103699>.

Voas, J., Mell, P., Laplante, P. & Piroumian, V. (2025) *Security and Trust Considerations for Digital Twin Technology*. NIST IR 8356. <https://doi.org/10.6028/nist.ir.8356>.



Daniela TIMISICĂ is a PhD student at the Doctoral School of Automatic Control and Computer Science of the National University of Science and Technology POLITEHNICA Bucharest and a Scientific Researcher in the Software and Complex Systems Engineering Department at the National Institute for Research and Development in Informatics – ICI Bucharest. Her research interests and activities include the Internet of Things (IoT), edge computing, cloud computing, artificial intelligence, machine learning and cybersecurity. Her professional activity also includes web development, with a focus on designing and implementing web applications and digital platforms. She is also interested in robotics, blockchain technology and Digital Twins, particularly in exploring the potential integration of these technologies into emerging digital ecosystems.

Daniela TIMISICĂ este doctorandă la Școala Doctorală de Automatică și Calculatoare a Universității Naționale de Știință și Tehnologie POLITEHNICA București și Cercetător Științific în cadrul Departamentului Ingineria Software și a Sistemelor Complexe al Institutului Național de Cercetare-Dezvoltare în Informatică – ICI București. Domeniile sale de interes și activitățile de cercetare includ internetul obiectelor (IoT), edge computing, cloud computing, inteligența artificială, învățarea automată și securitatea cibernetică. Activitatea sa profesională cuprinde, de asemenea, dezvoltarea web, cu accent pe proiectarea și implementarea aplicațiilor web și a platformelor digitale. Este interesată și de studiul roboticii, al tehnologiei blockchain și al gemenilor digitali, urmărind în special potențialul integrării acestor tehnologii în dezvoltarea unor ecosisteme digitale emergente.



Paul NICULESCU-MIZIL GHEORGHE is a Scientific Researcher III at the National Institute for Research and Development in Informatics – ICI Bucharest, where he coordinates the Technology Transfer Compartment, innovation and research valorisation. His work focuses on emerging technologies, system architectures and collaboration among research organisations, industry, universities and public institutions. His main areas of interest include blockchain technology, Web3, artificial intelligence, the Internet of Things (IoT), cybersecurity, critical infrastructures and complex systems. At ICI Bucharest, he coordinates blockchain and digital asset projects, including ICI D|SERVICES. He holds a PhD in Industrial Engineering from the University POLITEHNICA of Bucharest and has international experience in research, consultancy, management and partnership development.

Paul NICULESCU-MIZIL GHEORGHE este Cercetător Științific gradul III în cadrul Institutului Național de Cercetare-Dezvoltare în Informatică – ICI București, unde coordonează Compartimentul de Transfer Tehnologic, activitățile de inovare și valorificare a cercetării. Activitatea sa profesională se concentrează pe tehnologiile emergente, dezvoltarea arhitecturilor de sisteme și colaborarea dintre mediul de cercetare, industrie, universități și instituții publice. Domeniile sale principale de interes includ tehnologia blockchain, Web3, inteligența artificială, internetul obiectelor (IoT), securitatea cibernetică, infrastructurile critice și sistemele complexe. În cadrul ICI București, coordonează proiecte în domeniul tehnologiilor blockchain și al activelor digitale, inclusiv ICI D|SERVICES. A obținut titlul de doctor în Inginerie Industrială la Universitatea POLITEHNICA din București și are experiență internațională în cercetare, consultanță, management și dezvoltarea de parteneriate.



This is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.