Street Art & Urban Creativity | Vol. 11, No. 7, 2025 | ISSN 2183-9956
Street Art & Urban Creativity. Scientific Journal
https://doi.org/10.62161/sauc.v11.5997

SURPLUS: STRATEGIC URBAN RESOURCE PLANNING USING LDTS AND UNIFIED KQR SUSTAINABILITY METRICS

A METHODOLOGY TO ALIGN LOCAL DIGITAL TWINS WITH MEASURABLE OUTCOMES

ANTONIO JARA[1], PATRICIA TAMARIT[2], JORDI ORTUÑO[3], ERNESTO FAUBEL[4]
[1]LIBELIUM, ESPAÑA
[2]NUNSYS, ESPAÑA
[3]AYUNTAMIENTO DE BARCELONA, ESPAÑA
[4]AYUNTAMIENTO DE VALÈNCIA, ESPAA

KEYWORDS
EU LDT Toolbox
LORDIMAS
KQR
Smart Cities
Sustainability
Urban Planning
AI for Cities
FIWARE
EDIC

ABSTRACT
An AI-enabled framework has been designed and validated across multiple city use cases to transform business problems into measurable, sustainable outcomes. Grounded in the Key Business Questions (KBQ) paradigm, it integrates governance-ready data unification (ETSI NGSI-LD and MiMs), digital twin models, and economic evaluation (ROI, IRR, NPV) to maximize fiscal and societal surplus. Operationalization combines the EU Local Digital Twin Toolbox for modular twins, sensor-to-decision pipelines, and integration with LORDIMAS for comparative evaluation. Evidence from deployments in energy, mobility, waste, and water demonstrates positive NPVs, rapid payback, and a structured pathway to prioritize and scale investments into urban programs.

RECEIVED: 05/ 09 / 2025
ACCEPTED: 06 / 10 / 2025

1. Introduction

Today, Cities do not suffer from a lack of technology; they suffer from a lack of a methodology that helps them to explore among all the different interventions, an exhaustive analysis of the advantages, disadvantages, side effects, dependencies, real deployment costs, maintenance costs, and return of investment. This kind of tool, or even toolbox, and methodologies is what the recent advances of AI applied for urban planning as the EU LDT Toolbox and GenAI as a supportive tool would support in the decision-making process.

The difference between an attractive innovation pilot and an investable program is the presence of a rigorous, testable economic thesis. Following the new capabilities of the Digital Twin Simulators for advanced urban efficiency, quality of life, and resiliency, we address the design of a value-first, measurement-forward practice. In that context, we propose SURPLUS: a methodology to convert good business problems into portfolios of actions evaluated under a unified surplus lens, where budgetary (investment) and societal (externalities) gains are fully auditable through digital twins.

As discussed, SURPLUS methodology is built on two complementary capabilities that have recently matured: Local Digital Twins (LDTs) together with testbeds where cities can safely simulate alternative policies before spending, and Generative AI (GenAI) as an integrative analyst that can assemble, contrast, and compare evidence across heterogeneous sources with unprecedented speed.

SURPLUS is driven by Key Business Questions (KBQs), that express the decision logic that urban planners, ICT departments in cities, as in this paper represented by the cities of Barcelona and Valencia, they actually need answered and anchor the entire analytics lifecycle (Troyanos, 2020). From there, interoperable context models and open smart data models establish a stable substrate for AI and simulation build over the Livingin.eu MiMs as the best practices to reach interoperability. Digital twins enable counterfactual experiments to forecast the surplus of alternative interventions before committing capital. We then apply standard economic reasoning (ROI, IRR, NPV) to a set of well-documented municipal cases and demonstrate how disciplined framing, measurement, and automation can compound value.

Until now, Cities make repeated, consequential decisions under uncertainty: whether to retime a corridor, replace luminaires, re-segment a water network or rationalize waste routes. Traditional analytics can estimate correlations, but public finance demands counterfactual answers: what will happen if we change tariffs, control policies, or maintenance schedules compared with a baseline, what is the real impact or consequence of a Low Emission Zone?

LDTs answer that question at operational fidelity. Far beyond static dashboards, the European Union's Local Digital Twin Toolbox provides modular building blocks that cover from acquisition with highly relevant datasets and databases accessed via Data Spaces, and initiatives as Destination Earth, the European Building Database, with a highly relevant base of knowledge that is seamless integrated thanks to semantic interoperability, simulation build on validated AI Models, visualization, and orchestration among different use-case and scenarios. These capabilities allow administrations to compose city scale twins capable of running policy relevant "what if" experiments and projecting fiscal and welfare consequences before committing capital. The toolbox is explicitly designed around openness and interoperability so that models, data, and results remain portable across districts and vendors, as defined by Interoperable Europe ambition.

If LDTs supply the controlled environment for testing and simulation, GenAI supplies the integrative layer that public programs chronically miss. Foundation model research shows that large models can generalize across modalities and tasks; applied carefully, they can act as comparative analysts that assemble the decision canvas for each choice, checking provenance, reconciling units, and producing standardized memos with confidence intervals and sensitivity bands. None of this obviates human judgment, human criteria for validation of the proposed ideas, and the most important human legal responsibility and liability. LDTs with GenAI interest is to transform data to evidence and raises the transparency bar.

As the Stanford report on foundation models observes, these systems are adaptable across domains yet must be governed for bias, safety, and auditability, considerations that municipal use magnifies. (Stanford CRFM, Freeman Spogli Institute) In the public sector, early surveys report measurable benefits where guardrails are explicit and workflows are redesigned rather than merely automated, but they also caution that quality, provenance, and evaluation remain the hard work. (McKinsey & Company, ONE MP)

SURPLUS enables this criteria of evaluation build over the capabilities of Key Business Questions (KBQs), a managerial discipline of phrasing what must be true, in quantitative terms, before a decision is taken (e.g., “What is the 10 year NPV of adaptive signal control on Corridor A at a 7% discount rate, under a measured value of time and bus OPEX capture?”). KBQs are not slogans; they are decision bound questions that define counterfactuals, acceptance thresholds, and measurement plans in advance. This framing has deep roots in analytics management and is a practical antidote to scope creep in data projects.

Once a KBQ is explicit, the twin can be configured to run the relevant counterfactuals, and results can be evaluated with standard public finance metrics such as net present value (NPV), internal rate of return (IRR), and Return of Investment (ROI) concepts that remain the lingua franca of budget offices and auditors. It is remarkable that based on the highly recommended book about why data sciences projects fail (Gray and Shellshear, 2024), the key reason behind several data projects fail is due to the lack of answering to the right business need and business case, in that domain platforms as Palantir (Palantir Technologies, 2025) has demonstrated in the market how a solution focused on finding the right business cases, and make solutions focused on impact have reached a major market impact and market acceptance.

Once we have the right KBQs, i.e. properly identified the right business case, and opportunity we need data to validate Europe’s Minimal Interoperability Mechanisms (MIMs Plus) provide the minimum agreements cities need so that systems, data, and services interoperate across departments and jurisdictions. In practice, MIMs are operationalized via the ETSI NGSI LD standard for context information management for the MIM1, an API and graph based information model that allows assets, events, and relationships to be described consistently, and via the Smart Data Models initiative (MIM2), which offers open, community vetted schemas across domains. These are not abstract standards; they are what allows a “LampPost” or “Intersection” to mean the same thing across a portfolio, and for a GenAI agent to reason safely across domains with consistent identities, units, and provenance (living-in.eu MiMs Plus).

Data sharing is catalyzed by the European Data Spaces, a policy and technical framework that lets participants exchange data under explicit governance while preserving sovereignty. The Data Spaces Support Centre (DSSC) provides a blueprint, building blocks, and guidance that municipal programs can reuse rather than reinvent,, adopted for Smart Cities in DS4SSCC.

When an LDT consumes data via a data space, e.g. energy tariffs, transit telemetry, or environmental measures, purpose binding, traceability, and revocation are not afterthoughts but built in properties of the exchange. This matters when results determine payments, enforcement, or equity analysis. (European Commission, Digital Strategy, dssc.eu).

To make all these tools accessible, sandboxes and the digital infrastructure needs to be made available to create a legal and funding scaffold for Multi Country Projects challenges, in that context the European Digital Infrastructure Consortium (EDIC) for Smart Cities focused on Local Digital Twins, Citiverse and the Smart Cities node for data spaces, that made all the described technologies accessible to cities, simplifying the governance, procurement, and measurement artifacts while scaling what works across borders. For that reason, in this context a SURPLUS methodology is highly relevant for benchmarking surplus across initiatives, enables capabilities to compare, share experiences, evaluate potentials interventions, and beyond technology to promote faster institutional learning with common metrics and auditable counterfactuals. (European Commission, EuroAccess)

EDIC starts with the capacity that AI that touches streets and operators must be validated before deployment, Europe’s network of Testing and Experimentation Facilities (TEFs)—notably CitCom.ai for smart, sustainable cities and communities, which offers large scale testbeds where AI systems are evaluated for safety, performance, and compliance.

2. Related work and theoretical framing

The KBQ discipline from Harvard Business Review reorients analytics toward decision-critical questions, reducing the risk of modeling without actionability (Troyanos, 2020). In parallel, research and policy communities position Local Digital Twins (LDTs) as core instruments to govern and optimize living conditions across urban, peri-urban, and rural areas (European Commission, 2025). ETSI's NGSI-LD specification and guidance documents describe how graph-based context models support digital twins as system-of-systems that integrate cross-domain semantics (ETSI, 2024).

On the platform layer, the EU LDT Toolbox describes governed access to multiple foundation models, LLM-driven agents, evaluation suites, and enterprise automations that connect AI to operations, that evolves the current data platforms into an IoT-to-twin pipelines such as FIWARE-based platforms are enabling (e.g. Iris360 at Cartagena, VLCi at Valencia city). Key opportunities that need to evolve with the benefit of the data spaces is to emphasize data quality, data lifecycle, and alignment with AI Act and Data Act about sovereignty and reuse. These ingredients motivate SURPLUS: a structured pathway from KBQs to surplus capture that is transparent, interoperable, and replicable.

What does this look like in practice? Representative municipal cases show the approach is not limited to a single domain. Streetlighting retrofits modeled with conservative tariffs and failure rates deliver positive NPV even before counting safety co benefits; adaptive signal control produces strong fiscal and societal returns when value of time and bus OPEX are accounted for; waste route optimization and leak detection often pay back in the first year where telemetry is adequate and measurement charters are co signed with finance. A portfolio view across these cases diversifies risk and compounds value, with early wins in one domain financing instrumentation in another. Crucially, each case is paired with a Measurement Charter that specifies formulas, lineage, and acceptance thresholds so that avoided costs and welfare gains move from slides into ledgers.

The societal ledger matters as much as the fiscal one. When twins quantify travel time reliability, air quality improvements, or reduced water losses, they make visible benefits that residents feel but budgets cannot fully capture. Frameworks such as U4SSC Key Performance Indicators help align local outcome tracking with internationally recognized sustainability metrics, while WHO guidelines provide scientifically grounded thresholds for environmental health.

GenAI's role in this discourse is pragmatic, not theatrical. The most useful agents are those that do less but do it reliably: assemble the KBQ decision canvas by auto linking tariffs, inventories, and historical performance; check unit consistency and provenance against a published measurement charter; generate simulation experiment plans and model cards with stated assumptions; produce comparison memos with confidence intervals and sensitivity bands that auditors recognize. Foundation model scholarship reminds us that these systems must be evaluated and governed—particularly for bias and robustness—before influencing operations. In LDT centric programs, that governance is concrete: agents operate in sandboxes, actions are logged with roll back, and performance is tested in TEFs.

3. Methodology

3.1. SURPLUS Methodology

The SURPLUS methodology translates good business problems into investable, auditable programs by chaining seven disciplined steps—Specify, Unify, Run, Prioritize, Land, Uphold, Sustain—over an interoperable Local Digital Twin (LDT) stack. Each step has a clear deliverable, a minimal set of quality gates, and a handoff to the next. The method is intentionally pragmatic: it privileges testable questions, counterfactual simulation, and economic evaluation (NPV, IRR, ROI, payback) over generic "smart city" narratives. The section below expands each step with concrete practices, twin/AI considerations, and governance patterns you can lift into tenders, measurement charters, and operating playbooks.

3.1.1. S: Specify KBQs and value logic

Start by expressing the decision as a testable Key Business Question (KBQ) with an explicit value equation. A well-formed KBQ states (i) the decision and horizon, (ii) the beneficiaries and channels of value (budgetary vs. societal), (iii) the counterfactual, (iv) acceptance thresholds (e.g., NPV > 0 at 7% and payback < 5 years), and (v) how evidence will be produced and captured in the ledger.

Two parallel ledgers prevent confusion: the fiscal ledger (direct flows that hit the city's budget—energy, maintenance, staff hours, tariffs, contracted services) and the societal ledger (monetized externalities—time saved, health, safety, emissions). KBQs should specify which ledger(s) are in scope and with what valuation parameters (e.g., value-of-time, VSL, emission cost). By writing paired KBQs—one fiscal, one societal—you resist the temptation to over-claim cash benefits and still reflect public value comprehensively.

Concretize KBQs with unit economics on the assets or locations where the intervention lands (per luminaire, per intersection, per pumping station, per schoolyard). This forces clarity on the variables that matter (wattage, runtime, tariff code; headway variance and SPaT; kWh/m^3 and tariff periods; shaded m^2 and exposure). Each KBQ concludes with a Decision Canvas: the decision boundary, inputs and assumptions, measurement plan, and the minimum viable twin you will build to test it. Tying the KBQ to a budget line (who pays, who saves, when) prevents "pilot purgatory".

Institutionalize acceptance gates with finance up front. Write down the discount rate (real vs. nominal), valuation parameters, and what counts as proof (e.g., weather-normalized kWh; difference-in-differences for travel-time; excess-mortality model for heat-health). This makes success or failure recognizable—and therefore governable.

3.1.2. U: Unify data and context with MiMs

Harmonize data via NGSI-LD/Smart Data Models and establish a trustworthy data space. Define lineage, licenses, and privacy. A twin is only as credible as its context model. Harmonize data into NGSI-LD entities and Smart Data Models so assets, events, and relationships are described consistently across departments and vendors. In practice, this means: define the entity catalog (e.g., LampPost, Intersection, SignalGroup, EBAR/Pump, Tree, ShadeStructure, HospitalAdmission), map identities across systems, and attach provenance and units as first-class attributes. NGSI-LD's graph semantics make cross-domain reasoning tractable (e.g., "this intersection contains these signal groups and abuts these bus stops").

The EU LDT Toolbox provides the composable tools (ingestion, semantic mediation, simulation, visualization), so unification is not a bespoke integration project each time. The MIMs Plus discipline and open model catalogues keep artifacts portable across districts and vendors.

3.1.3. R: Run simulations on your Local Digital Twin

Run causal models & simulations. Build interpretable AI with digital twins to forecast counterfactuals, including risk and uncertainty. With the question and context fixed, build the minimal twin that can run counterfactuals credibly. Favor causal structure over raw correlation: physics-aware energy/flow models; mesoscopic traffic with platoon dispersion; micro-climate exposure with social vulnerability overlays. Where feasible, anchor identification with natural experiments (phased rollouts, thresholds) and evaluate with difference-in-differences or synthetic control to corroborate simulated effects.

Quantify uncertainty: propagate measurement error and parameter uncertainty via scenario ensembles or Monte Carlo; report confidence intervals and sensitivity bands in Evaluation Memos. For GenAI, keep the role pragmatic and governed: use it to assemble Decision Canvases from heterogeneous sources, check unit consistency and provenance, autogenerate Model Cards and experiment plans, and draft comparison memos. Do not allow opaque agents to change operational states without human-in-the-loop and an audit trail. Validate automation in a sandbox first (e.g., TEF/CitCom.ai) before touching streets or plants.

The outcome of this step is not just a plot—it is a policy-ready counterfactual forecast with assumptions, boundary conditions, and interpretable drivers that finance and auditors can read. That's the bar for "decision-grade" modeling in public administration.

3.1.4. P: Prioritize different options and initiatives

Rank initiatives by NPV, IRR, payback, time-to-value, and risk. Optimize the budget across interdependent investments.

Economic foundation. Let CF_t denote net cashflow at year t and r the discount rate. We compute NPV=Σ $CF_t/(1+r)^t$, IRR such that NPV=0, ROI=(Σ benefits − Σ costs)/Σ costs, and payback as the first year where cumulative CF≥0 (interpolated within the year). We evaluate over a 10-year horizon at 7% unless otherwise stated.

$$\mathrm{NPV} = \sum_{t=0}^{T} \frac{CF_t}{(1+r)^t}, \quad \mathrm{IRR\text{: solve NPV}} = 0, \quad \mathrm{ROI} = \frac{\sum \mathrm{benefits} - \sum \mathrm{costs}}{\sum \mathrm{costs}}.$$

Treat candidate actions as a portfolio under constraints (capital, crews, political attention, data readiness). Rank them by NPV, IRR, payback, time-to-value, and risk, but do not stop at a sorted list—optimize the selection subject to budget and policy constraints (e.g., equity: cap delay variance by district; resilience: ensure coverage in flood-prone tracts). Interdependencies matter: lighting upgrades free poles and crews for sensors; pump scheduling reduces peaks that otherwise drive wastewater OPEX; shade networks interact with heat-health operations.

$$\max_{x_i \in \{0,1\}} \sum_i x_i \, \mathrm{NPV}_i \quad \text{s.t.} \quad \sum_i x_i \, \mathrm{CAPEX}_{i,y} \le \mathrm{Budget}_y \ \forall y, \quad \text{and policy constraints.}$$

3.1.5. L: Land it into operations

Translate model recommendations into standard operating procedures, automations, and budget adjustments. "Landing" means the twin stops being a slide and starts changing Monday-morning work. Translate model recommendations into standard operating procedures (SOPs), automations, and budget adjustments that existing teams recognize. Examples: tele-managed dimming policies with change control; signal-timing governance with degraded modes when detector health dips; pump schedules bound to tariff calendars and wet-weather exceptions; heat-health playbooks tied to alert levels and shelter hours. Connect to work-order/ERP systems and tariff/contract mechanics so savings show up in the ledger (e.g., tariff class updates; line items reduced in maintenance contracts).

Codify RACI (who decides, who executes), change-control and rollback, and the weekly digital operations review where operators, analysts, and finance inspect KPIs, discuss anomalies, and approve policy adjustments. Where AI agents propose actions, insist on policy-enforced scopes, attribution, and rollback. Landing is where value is either captured—or lost to institutional friction.

3.1.6. U: Uphold measurement and check KPIs, and verify returns

Instrument KPIs and verification rules; implement outcome-based payments when appropriate, without measurement that survives audit, there is no surplus—only claims. The Measurement Charter is the contract between operations and finance: it names the canonical KPIs, formulas, lineage, and evaluation designs; sets the publication cadence; and defines roles for dispute resolution. Compute KPIs from raw telemetry, not hand-entered figures. For causal evaluation, prefer designs you can defend in public (stepped-wedge deployments, matched controls, synthetic control). Publish Evaluation Memos with methods, assumptions, and confidence bands; expose raw aggregates for external scrutiny when permitted.

When vendors are paid on outcomes, the charter is also the payment rail: everyone knows how success is computed before the first invoice. Keep a risk register for data drift, price/tariff changes, lock-in, and equity backlashes, with mitigations (drift monitors and retraining; portability clauses; fairness constraints in the twin; grievance channels). Measurement is not a dashboard; it is an institutional act that binds evidence to money.

3.1.7. S: Sustain and Scale

Productize twin assets, APIs, and templates so they can be reused across districts and cities; monitor drift, and replicate via marketplaces as the EDIC Marketplace.

Make the program durable by turning artifacts into products: schemas, entity definitions, playbooks, evaluators, cost functions, and API contracts become a Reuse Kit. Monitor model drift and data-quality regressions; schedule retraining; define deprecation policies. Use open standards (NGSI-LD, Smart Data Models) and portability clauses so cities retain sovereignty and avoid re-platforming costs with every contract cycle. Package assets for replication across districts and neighboring cities—ideally via European rails such as the LDT Toolbox and EDIC marketplaces—so scale-out becomes parameterization, not bespoke code.

Sustain also means people: invest in operator training, celebrate visible wins (safer crossings, fewer leaks, calmer summer plazas), and budget for maintenance of the twin just like any other asset. Report surplus realized annually—fiscal and societal—so the public can inspect where value came from, where it fell short, and what you will try next. A small, honest portfolio that compounds is better than a sprawling one that cannot be measured.

3.2 Research Methodology

This section specifies the technical–scientific procedures used to (i) select municipal cases, (ii) collect and prepare data, (iii) estimate counterfactual impacts and translate them into economic metrics, and (iv) validate, reproduce, and audit results.
We adopt a multi-case, quasi-experimental design. Each intervention (e.g., adaptive signals, shade program, LEZ, route optimization, building plant controls, public lighting) is treated as a stand-alone study following a shared Pre-Analysis Plan (PAP) that locks:

1. Decision & KBQ. Decision boundary, horizon, beneficiaries, and acceptance thresholds (e.g., *NPV > 0 at 7% real; payback < 5 years*).
2. Outcomes & units. Primary/secondary outcomes, unit of analysis (asset, corridor, zone), and aggregation rules.
3. Identification strategy. *Difference-in-Differences* (DiD) with two-way fixed effects and event-study diagnostics; *stepped-wedge* adoption; *synthetic control* for zone-level outcomes; *regression discontinuity* (RD) for threshold policies; or *instrumental variables* (IV) where valid instruments exist.
4. Data & lineage. Sources, access, licenses, privacy constraints, quality gates, and versioning.
5. Uncertainty & robustness. Planned bootstraps, Monte Carlo scenarios, falsification tests.
6. Economics. Dual ledger (fiscal vs. societal), valuation parameters (value of time, VSL, emission costs), and horizon/discount conventions.

Where Local Digital Twins (LDTs) are used, the PAP requires pre-period calibration, hold-out validation, and reconciliation of simulated deltas with empirical estimators before economic evaluation.

3.2.2. Case Selection: Inclusion/Exclusion Criteria and Scoring Rubric

To ensure maximum variation (domains, governance, maturity) while preserving measurement feasibility and policy relevance with a procedure based on the (i) Call for candidate cases with a KBQ fiche from the previous step; (ii) document screening (data, permissions, risks); (iii) scoring by rubric; (iv) ethics/governance review; (v) methodological committee approval.

Table 1. Case-selection rubric (0–5 scale; weights sum to 1.00)

Criterion (weight)	Inclusion signal (4–5)	Exclusion signal (0–1)
KBQ clarity (0.15)	Decision boundary, horizon, acceptance thresholds defined ex-ante	Vague goals; no measurable thresholds
Data sufficiency (0.20)	≥12 mo baseline; ≥6 mo post; stable IDs & units	<3 mo baseline; fragmented logs
Measurement feasibility (0.10)	Charterable KPIs; auditable lineage	Proxy-only; unverifiable
LDT/Testbed readiness (0.15)	NGSI-LD/Smart Data Models; toolbox access	No semantic model; ad-hoc CSV
Identification strategy (0.10)	Natural experiment, rollout, or valid control	No plausible counterfactual
Policy relevance (0.10)	Material OPEX/CAPEX or societal stakes	Trivial or symbolic
Representativeness (0.15)	Adds domain/geography diversity	Duplicates existing case
Ethics & compliance (0.05)	DPIA complete; GDPR-conform	Governance constraints unmet

Source(s): Own elaboration, 2025.

These subsections convert SURPLUS from a management framework into a replicable research methodology: cases are selected transparently, data are unified with auditable lineage, effects are identified with explicit designs, economics are computed with a dual-ledger discipline, and all outputs are packaged for verification and reuse, the next sections present how to implement it using the EU LDT Toolbox.

4. EU LDT Toolbox to implement the SURPLUS methodology

SURPLUS requires the EU LDT Toolbox to get (i) a governed, interoperable data-to-model pipeline; and (ii) a repeatable way to package models, experiments, and evidence so finance and operational teams can act with confidence. The European Local Digital Twin (LDT) Toolbox was designed precisely for this: it aggregates modular building blocks for acquisition, modeling, simulation, visualization, interoperability testing, and adoption/evaluation guidance, with the explicit goal of democratizing twin capabilities across communities and maturity levels. This alignment means the Toolbox is not just “another platform,” but a reference implementation to turn SURPLUS from method into machinery.

How the Toolbox maps to SURPLUS (end-to-end)

1. S — Specify KBQs and value logic.
   The *Use Cases & Scenarios* tool provides the formalism to encode Key Business Questions as configurable, versioned “receipts” that bind an operational need (e.g., LEZ design, leak reduction) to required inputs, decision thresholds, and acceptance criteria. These receipts

orchestrate the parameters to be varied (scenarios) and the experiments to be run (what-ifs), preparing a clean decision canvas before any capital is committed. The *City Innovation Planner* complements this by logging strategic goals and maturity assessments, so every KBQ is linked to explicit outcome targets and the city's baseline capacity to deliver.

2. U — Unify data and semantics.
   The *Data Platform* harmonizes heterogeneous sources (sensors, databases, external systems) into NGSI-LD context graphs and Smart Data Models, implementing MIM1/MIM2 so that assets, events, and relations mean the same thing across departments and suppliers. This is where raw telemetry and reference registers are normalized, historized, and exposed through consistent APIs. When data must be shared or acquired under policy, the *Data Space Ready* tool exposes and consumes governed interfaces, enabling identity-, contract-, and policy-aware exchange consistent with European data space patterns. Together, these tools satisfy SURPLUS's requirement that all experiments rest on a stable, auditable semantic substrate.
3. R — Run counterfactuals (simulation + AI).
   The *AI Notebook* packages models (deterministic simulators, ML, causal estimators) behind standard I/O contracts; *Federated Learning* lets partners train where data resides; and the *Data Modeller* generates synthetic series to fill gaps or protect privacy. These are connected at runtime by *Use Cases & Scenarios*, which orchestrates multi-model experiments (e.g., traffic ↔ pollution ↔ weather), records provenance, and produces structured outputs for analysis and audit. This realizes SURPLUS's call for interpretable, scenario-driven inference instead of single-point predictions detached from decision context.
4. P — Prioritize options as a portfolio.
   Results flow to the *City Innovation Planner* for portfolio-level ranking by NPV/IRR/payback and for comparison against policy KPIs (e.g., U4SSC), while *Use Cases & Scenarios* retains experiment lineages needed for sensitivity analysis and risk. This makes the financial and societal "dual ledger" explicit—an essential condition for moving from attractive pilots to investable programs.
5. L — Land into operations.
   Play & Visualise binds model outputs to 2D/3D geospatial contexts and business dashboards so operators can embed recommendations into SOPs and, where appropriate, semi-automate actuation (e.g., signal plans, pumping schedules) under human approval. Identity and access are centrally handled by *Identity Management*. The Toolbox's integration stance—APIs, message buses, and standard payloads—keeps "last-mile" wiring to control systems pragmatic and auditable.
6. U — Uphold measurement and verification.
   Time-series in the *Data Platform* serve as the system of record for post-deployment verification; *Participate* incorporates citizen feedback into the evidence loop and documents how AI was used, supporting fair-AI and GDPR requirements. This closes the loop demanded by SURPLUS: outcomes are measured, counterfactuals are revisited, and policies are adapted.
7. S — Sustain & scale.
   Marketplace and *Data Space Ready* make models and datasets discoverable and contractable across cities; this accelerates replication and transfers evidence along with assets. The same packaging of "receipt + models + KPIs" that enabled the first deployment becomes a shareable template for the next, in line with EDIC-style cross-border scaling.

4.1. Reference architecture and standards alignment

The Toolbox implements a seven-layer reference architecture that extends the familiar five-layer smart-city stacks (sources → acquisition → knowledge/analytics → interoperability → services) with two LDT-specific meta-layers: Smart Orchestration (to compose cross-domain receipts and what-if scenarios) and Visualization & Applications (to contextualize real-time and simulated states in a twin). This extension is what converts a data platform into a decision and experimentation platform, which is precisely what SURPLUS operationalization requires. The architecture is explicitly aligned to MIMs, ETSI NGSI-LD, OGC formats, OAuth/OpenID, and EU data-space building blocks, ensuring that cities can adopt it incrementally and interface it with existing platforms.

The Toolbox offers deployable tools that map cleanly to layers and MIMs: *Data Platform* (MIM1/2 for context + models), *Data Space Ready* (MIM3 for transactions; MIM7 for geospatial), *AI Notebook / Federated Learning / Data Modeller* (MIM5 for fair-AI documentation), *Use Cases & Scenarios* (MIM5/7/9 for AI explainability, geospatial state, and analytics exchange), *Play & Visualise* (MIM7/9), *Identity Management* (MIM4/6 for privacy and security), *City Innovation Planner* (MIM8/10 for KPI and impact/ROI management), *Participate* (MIM4/6 for privacy/security in engagement), and *Marketplace* (MIM3/5/6 for transactions, fair-AI, and secure onboarding). This explicit mapping minimizes ambiguity during procurement, integration, and audit.

5. SURPLUS validation into different use-cases and domains

We examine five implementations spanning energy, mobility, solid waste, and water. For each we document context, assumptions, data, and economics. Where sources provide ranges (e.g., for public time valuation), we report sensitivity intervals and maintain clarity on what is economical vs societal.

We evaluate three Spain-based interventions; Barcelona's "Programa de Sombras" (heat-health shade), València's district-energy optimization at the City of Arts and Sciences, and Cartagena's Low Emission Zone using SURPLUS, and we will also validate it with other three global interventions in the USA, and Australia.

In SURPLUS, KBQs define the decision to be made, the counterfactual, the unit economics, and the acceptance thresholds for investment. Digital twins then simulate counterfactual outcomes under alternative policies so that value is forecast *before* committing capital; measurement charters ensure that value is later *captured* in the ledger. We adopt a ten-year horizon and a 7% real discount rate. Economic metrics are computed as: Net Present Value (NPV) using discounted cash flows; Internal Rate of Return (IRR) as the rate at which NPV=0; and Return on Investment (ROI) as undiscounted 10-year net gain divided by initial CAPEX (we report ROI as a simple ratio, not discounted). This procedure follows the SURPLUS mathematical model and dual-ledger discipline (economical vs. societal) so that budgetary savings and externalities are reported distinctly and can be audited.

We build each use case on an interoperable context model with the EU Local Digital Twin Toolbox and sandboxes as composable blocks for acquisition, semantics, simulation, and visualization; while preserving data provenance. Economic assumptions are conservative and, where relevant, anchored in official Spanish or EU sources (tariffs, hospital costs, and value of statistical life).

5.1. Barcelona (Barcelona, Spain): Heat-health shade program.

The city has launched a shade program to adapt public space to intensifying heat, with nearly 200 shaded areas planned by 2027 and >50,000 m^2 of new shade surfaces. We evaluate whether deploying tensile structures and pergolas—prioritizing schoolyards and playgrounds—delivers a positive fiscal/societal surplus once health benefits are monetized and modest upkeep is accounted for. We model benefits as avoided heat-related mortality and hospitalizations attributable to improved shade coverage in high-exposure micro-sites, with attribution and baselines documented in the measurement charter. Telemetry binds shade assets to micro-climate readings (air and globe temperature), school/playground occupancy, and health outcomes at neighborhood level; expenditures and OPEX are tracked in the fiscal ledger. Public sources confirm scale and budget (≈€13 M by 2027; ≈194 new shade spaces; ≈50,000 m^2). We monetize mortality reductions using Spain's official updates of the Value per Statistical Life (VSL) and hospital costs using the SNS 2022 cost base; both are used only for the *societal* ledger, with a clear separation from fiscal flows.

For Barcelona we treat CAPEX as €13.0 M (t=0) with €0.26 M/y OPEX, and—under a modest coverage-to-risk reduction elasticity—2 lives/year avoided and 50 hospitalizations/year avoided, monetized at VSL≈€1.6 M and average acute-care cost ≈€5.4 k per discharge (10-year horizon; 7% discount).

5.2. Valencia (Valencia, Spain): District energy at the City of Arts and Sciences

The City of Arts and Sciences is upgrading cooling with a new geothermal plant and control modernization. We evaluate the incremental economics of two control levers that a twin can safely test and then automate: raising chilled-water supply temperature (CHWST) when conditions allow and recovering "lost" ΔT to reduce over-pumping. Engineering literature indicates chiller energy can drop ≈1.5–2.5% per 1 °C increase in CHWST (given stable loads), and *low-ΔT* remedies reduce pumping penalties—capabilities codified in Guideline 36 sequences. *Measurement.* The twin ingests BACnet/SCADA data (flows, temperatures, kW), flags low-ΔT patterns, and runs week-over-week counterfactuals to isolate savings. Energy is monetized with EU non-household electricity prices; we keep a conservative base price to avoid overstating benefits. The CACSA geothermal works and the installed smart lighting/controls track record offer local context for achievable gains.

5.3. Cartagena (Murcia, Spain): Low Emission Zones

The City Cartagena approved a Low Emission Zone ordinance that delimits two areas, Casco Histórico (Historical city center) and Ensanche, and emphasizes superblocks (supermanzana) traffic calming, adaptive access management, and real-time monitoring over blanket bans by DGT label. The perimeter and regulatory approach are publicly documented; the Casco Histórico keeps access for residents/garages and logistics while prioritizing pedestrian use, and the Ensanche is structured around superblocks to discourage through-traffic on internal streets. Unlike other cities, Cartagena explicitly states it does not face acute exceedances city-wide and opts for light-touch, reversible restrictions triggered by demand and air-quality episodes. The KBQ we model is: "What is the 10-year societal NPV and IRR of the Cartagena ZBE (Casco Histórico + Ensanche) under a conservative health-benefit scenario, given measured reductions in exposure and the city's modest enforcement posture?".

The city procured air-quality sensors from Libelium and access-control systems, ANPR cameras and environmental sensors, co-financed through Spain's Recovery Plan (PRTR). This data enables a Twin design and measurement. The operational twin combines (i) an exposure module linking local NO_2/PM telemetry to population micro-zones in Casco Histórico/Ensanche, (ii) traffic-assignment variants that reflect superblock layouts (boundary-link flows increase; internal cells calm to 10–30 km/h), and (iii) policy switches for episodic restrictions (pollution exceedances, cruise-arrival crowding windows). Outcomes are computed from raw telemetry, not hand-entered KPIs, following the SURPLUS measurement charter: traceable formulas, explicit value-of-life assumptions, and auditability of avoided episodes/hours. This is aligned with EU LDT Toolbox practices for modular twins and Living-in.EU MIMs for context models.

This brings an example of societal value, as a net social gain. The primary economic signal is health: fewer pollution-related deaths and admissions. Recent evidence (EU and UK) finds LEZs reduce NO_2 and generate measurable health improvements (notably cardiovascular), with effects extending beyond the zone. We therefore monetize avoided mortality using Spain's updated Value of a Statistical Life (VSL) and avoided hospital admissions using SNS average per-discharge costs

5.4. Los Angeles (California, USA): LED streetlighting

The City of Los Angeles completed a comprehensive conversion to LED street lighting, delivering a 64% reduction in energy use and approximately $10 million per year in energy savings, along with significant CO_2 reduction (LA Bureau of Street Lighting). The initial program cost has been reported at roughly $57 million. In SURPLUS terms, this is a budgetary surplus case: savings are realized as cashflow reductions on the energy bill and maintenance. We model −$57M in year 0 and +$10M annually for 10 years. We do not assign residual value to luminaire life extension or tariff renegotiation upside, so our estimates are conservative.

5.5. Bellvue (Sidney, Australia): Adaptive Signals

Bellevue deployed the Sydney Coordinated Adaptive Traffic System (SCATS) to more than 200 intersections, with city documents and trade coverage noting full adaptive operations by the end of 2015 at an approximate deployment cost of $5.5M. Studies report corridor travel-time and delay reductions in the 13–43% range, and one article estimated public time savings at ~$3.2M per year using a $15/hour value of time. In SURPLUS we classify these gains primarily as societal surplus (time; safety externalities). We deduct a 5% OPEX proxy from the benefit to represent lifecycle costs for detection, communications, and software support.

5.6. Virgina Beach (Virginia, USA): Solid waste management

Route optimization reduced routes and improved workload balance. Documented results indicate a $2.25M reduction in equipment inventory and $1.1M in annual operating savings (personnel, maintenance, fuel). A separate vendor case reports $400k/year savings and $1M inventory relief for a subset, illustrating that benefits vary by scope and phase. For the base evaluation we use the larger program outcomes and adopt $400k as a conservative implementation cost representing software and consulting. Inventory relief is modeled as a year-1 one-off cash inflow.

6. Results

6.1. SURPLUS results over the defined use-cases

Economic evaluation (10-year horizon, 7% discount). We compute NPV, IRR, ROI, and payback for each case and for the aggregate portfolio. The aggregate treats cashflows as additive across cases. Societal valuations are marked and kept separate from budgetary ledger impacts.

Table 2. Economic Evaluation (10-year horizon, 7% discount)

Case	NPV (7%)	IRR	ROI (10y)	Payback (yrs)
Los Angeles — LED Streetlighting	11,3 M€ ($13,235,815)	11.8%	75.4%	5.7
Bellevue — Adaptive Signals	12,8 M€ ($15,043,976)	52.4%	431.8%	1.9

Virginia Beach — Route Optimization	8,0 M€ ($9,428,743)	773.1%	3212.5%	0.1
Barcelona — Heat-health shade	9.58 M€	27,9%	102,2%	4
València — District energy	1,45 M€	52,8%	391%	2
Cartagena — LEZ (Casco + Ensanche, superblocks)	0,88 M€	35,2%	110%	2,7

Source(s): Own elaboration, 2025.

Sensitivity and risk, we examine ±25% variation on benefits and CAPEX. LED lighting remains positive NPV unless benefits fall by 25% while CAPEX rises by 25% simultaneously. Adaptive signals remain beneficial in societal terms across a wide range of time-valuation parameters; fiscal effects depend on fuel tax regimes and transit incentives. Route optimization is robust to assumptions given the large inventory relief; Decatur and Oldcastle pilots pay back quickly, making them resilient to estimation error. Key risks include data drift, tariff or price changes, and institutional capacity to sustain measurement.

6.2 LORDIMAS and SURPLUS to connect maturity with measurable objectives

LORDIMAS gives cities a shared language and evidence base for digital maturity; SURPLUS gives them a disciplined way to turn that maturity into auditable economic surplus via KBQs, interoperable twins, and verified measurement. Mapping the two creates a single operating model: assess readiness with LORDIMAS; pick and shape investments with SURPLUS; then re-score maturity as the city lands, verifies, and scales what works.

LORDIMAS is a digital maturity assessment for local, metropolitan and regional governments developed by ESPON with the European Committee of the Regions and partners in the Living-in.EU movement. It structures a self-assessment across seven dimensions—Governance, Service Design, Data Management, Interoperability, Service Delivery, Technology, and Networking—with six maturity levels from "digitally nascent" to "digitally native". The first full year of collection (2024) captured 99 authorities with a median "digitally emerging" score; respondents tended to be stronger in Governance/Service Delivery/Networking, and weaker in Data Management/Interoperability and adoption of new technologies.

SURPLUS, for its part, is a KBQ-centred practice: specify the decision and value logic; unify data with MIMs; run use-cases and scenarios over LDTs to test; prioritize portfolios on NPV/IRR/payback; land changes into operations; uphold measurement with a charter; and sustain and scale through reuse. LORDIMAS names the organisational and technical pre-conditions for digital programmes, This makes the maturity score actionable: it becomes a *gate* or *multiplier* in the economic calculus.

Table 3. LORDIMAS dimensions and SURPLUS steps pairing

LORDIMAS dimension (what's assessed)	Primary SURPLUS steps it enables (what to do)	Why this pairing matters
Governance (strategy, funding, oversight, risk)	S – Specify, U – Uphold, S – Sustain	If governance is "nascent", the KBQ cannot be owned, measured, or defended; acceptance thresholds and the Measurement Charter will drift.
Service Design (co-creation, digital-by-default)	S – Specify, L – Land	Well-formed KBQs name users and services; landing policies require co-designed SOPs that residents and crews accept.
Data Management (governance, quality, lineage)	U – Unify, U – Uphold	SURPLUS relies on entities, provenance, and decision-critical quality gates; weak data management increases evaluation risk.

Interoperability (standards, APIs, MIMs)	U – Unify	Minimal Interoperability Mechanisms (MIMs Plus) and Living-in.EU practices make twins portable and auditable across vendors.
Service Delivery (operational capability)	L – Land	Twin recommendations must change Monday-morning work; delivery maturity predicts whether savings hit the ledger.
Technology (platforms, emerging tech adoption)	R – Run, L – Land	Running counterfactuals and automations demands reliable platforms; low scores suggest a minimal twin first approach.
Networking (ecosystems, peer learning)	S – Sustain, P – Prioritize	Networks reduce risk and time-to-value by reusing patterns and market testbeds (e.g., EU LDT Toolbox & TEFs).

Source(s): Own elaboration, 2025.

6.3. EU LDT Toolbox as the bridge between LORDIMAS and SURPLUS

The EU Local Digital Twin Toolbox provides modular set of tools that translate LORDIMAS readiness into executable SURPLUS steps without lock-in. Cities can start with semantic mediation and visualisation, then add simulation kernels and automations as Technology and Data Management scores rise. OECD and Commission materials describe the toolbox as a reusable, open framework for city twins; Living-in.EU positions it alongside MIMs and data spaces.

1- Assess and pick. Run the LORDIMAS self-assessment; identify two KBQs that matter and that the current maturity can support with minimal enablers (often one utilities and one mobility/realm case). Decision Canvases record the acceptance thresholds and the readiness gates.

2- Bridge gaps as investments. Where LORDIMAS exposes deficits (e.g., Interoperability), fund “U – Unify” enablers first: NGSI-LD context broker, Smart Data Models adoption, quality gates. Treat these as NPV-bearing projects with their own charters.

3- Run and verify. Compose the minimal twin with the LDT Toolbox; run causal counterfactuals; land one outcome-based contract. The Measurement Charter aligns telemetry with the ledger; results feed back to the next LORDIMAS cycle.

4- Scale and re-score. Package the Reuse Kit; replicate by parameter change across districts; update LORDIMAS to record gains in Service Delivery, Interoperability, and Technology. Portfolio selection in SURPLUS now admits higher-ambition KBQs (e.g., agentic control with sandboxing), with LORDIMAS “Networking” guiding peer reuse and TEF participation.

6.3.1. This connection in the previous examples

- Barcelona heat–health & shade. LORDIMAS typically shows Governance/Service Delivery strengths but Data Management variability. SURPLUS responds by making the Measurement Charter (excess-mortality, admissions) the first deliverable, and by scoping a light-weight twin (exposure + micro-climate) that respects current Technology maturity. As telemetry and evaluation stabilize, the city’s next LORDIMAS cycle should improve in Data Management and Interoperability, unlocking broader adaptation portfolios.
- València district-energy controls. Where LORDIMAS shows stronger Technology but middling Interoperability, SURPLUS “U – Unify” requires NGSI-LD/BACnet mappings and MIMs-aligned schemas before automating set-points; the L – Land playbook binds control changes to ASHRAE-style sequences and rollback. The LDT Toolbox keeps the model portable across campuses. I
- Cartagena LEZ. A low–moderate Data Management score triggers a minimal twin with a conservative savings claim; Networking maturity is leveraged to borrow policies and evaluators from peers. The LEZ brings Governance and Service Design scores determine

whether the city can operate *reversible* restrictions with community acceptance; societal-ledger benefits (health) are computed visibly, per SURPLUS.

6.3.2. Governance and reporting, one score board with two views

To avoid "dual truths", publish a single scoreboard where each SURPLUS initiative lists (i) maturity deltas (the LORDIMAS dimensions it improved and by how much) and (ii) economic results (NPV, IRR, ROI, payback; fiscal vs. societal ledgers). LORDIMAS thus becomes both *baseline* and *beneficiary*: investments strengthen the very capacities such as Interoperability, Data Management, Service Delivery that LORDIMAS measures

7. Discussion

SURPLUS reframes Smart City investments as economic instruments designed to maximize measurable surplus. By binding Key Business Questions (KBQs) to interoperable data, causal digital twins, and disciplined economic evaluation, cities can prioritize portfolios that survive budget scrutiny and deliver visible improvements in daily life. This paper has shown that when decisions are expressed as testable KBQs and traced to auditable ledgers, "smart" ceases to be a catalogue of tools and becomes a practice of public value creation. In short: start from the question; let the twin test before economical investment is committed; post results with methods that withstand audit; and scale only those actions that prove their worth.

A consistent theme across the cases is conservatism with clarity. We deliberately used cautious assumptions, separated fiscal from societal benefits, and published the acceptance thresholds—NPV > 0 at a real 7% discount, payback windows aligned to political horizons, and IRR above hurdle rates. Even within that discipline, many urban problems admit strong economics: energy retrofits, route optimization, leak reduction, and operations-grade control changes often return positive NPV with short payback; climate-health and access-management interventions create significant societal surplus that justifies blended finance. Cross-domain portfolios further diversify risk and accelerate learning: when lighting, mobility, water, and public-realm actions are selected and sequenced together, the program is robust to individual under-performance and can compound value through shared assets (e.g., poles, telemetry, crews).

What distinguishes SURPLUS from prior "pilot cultures" is its administrative spine. First, decision-first: every initiative is born with a KBQ that names the decision boundary, counterfactual, beneficiaries, and acceptance thresholds. Second, semantics-tight: data are unified into NGSI-LD entities and Smart Data Models with provenance, units, and quality gates so that models remain portable and results reproducible. Third, experiment-verified: digital twins run the causal "what-ifs," quantify uncertainty, and link assumptions to outcomes. Fourth, auditor-ready: measurement charters and evaluation memos specify formulas, baselines, confidence intervals, and publication cadence so finance can reconcile telemetry with invoices and budgets. These habits transform analytics from narrative to evidence, and from evidence to budget.

The EU Local Digital Twin (LDT) Toolbox satisfies these demands with a MIMs-aligned, seven-layer architecture and a coherent suite of tools that operationalize the SURPLUS steps end-to-end. In practical terms, the toolbox (i) structures KBQs and measurement via reusable canvases and charters, (ii) harmonizes and governs data through semantic mediation and data-space connectors, (iii) orchestrates multi-model simulations to produce decision-grade counterfactuals, (iv) supports portfolio selection on fiscal and societal ledgers, (v) "lands" recommendations into operations and work-order integrations, and (vi) verifies and scales what works by packaging model cards, evaluators, and portability assets as reuse kits. Cities can adopt this stack incrementally—corridor by corridor, district by district—while preserving sovereignty and avoiding vendor lock-in. In short, the toolbox turns the SURPLUS methodology into an operational product cities can govern confidently and replicate efficiently.

The evidence base presented here in key domains as energy, mobility, solid waste, water, and urban health, makes a pragmatic case for why this approach is timely and tractable. Where budget flows dominate (e.g., electricity, fuel, maintenance), the economical aspects alone clears the hurdle. Where welfare effects dominate (e.g., travel time, safety, heat-health, air quality), the societal aspect is counted with standard valuations and defended with causal designs. Crucially, we did not launder societal benefits into fiscal ROI: we kept ledgers separate and used the separation to argue honestly for blended finance where appropriate (grants, value-capture, outcome-based payments). That clarity is not an academic nicety; it is why programs survive audit and why the public can scrutinize and accept trade-offs.

The institutional rails from the European Comission to reach a Digital Single Market and interoperable Europe that can take benefit of a common infrastructure as EDIC matter as much as the models. Minimal Interoperability Mechanisms (MIMs) keep semantics stable; European data-space principles provide traceability, purpose binding, and revocation; Testing and Experimentation Facilities (e.g., CitCom.ai) let teams validate automation before it touches streets or plants; maturity frameworks such as LORDIMAS help sequence portfolios realistically. Thereby, EDIC-style multi-country collaborations enable reuse and benchmarking across jurisdictions. Together, these rails make SURPLUS portable: the same KBQ, context model, evaluator, and playbook can be parameterized for Barcelona, València, Cartagena—or for any peer city willing to measure what it manages.

Looking ahead, three implications stand out. First, procurement must be outcome-literate. Tenders should include the measurement charter, portability clauses, and rights to re-host data and models; vendors should be paid in part for measured results, not only for inputs. Second, governance must be routine. A weekly digital-operations review where operators, analysts, and finance inspect KPIs, evaluate proposed changes, and authorize policy updates is the engine that converts insight into surplus. Third, replication must be productized. Cities should publish reuse kits with schemas, evaluators, cost functions, and SOPs; invest in operator training; and maintain a public surplus report that residents can read and contest. These are not add-ons: they are the institutions that keep a twin from drifting back into slides.

Limitations remain. Economic results depend on local tariffs, baselines, contract terms, and the stubborn realities of implementation. Societal valuations are sensitive to parameters like value-of-time or value-of-life and must be communicated with humility and sensitivity. Not every benefit can be monetized credibly; those that cannot should be reported qualitatively rather than forced into the fiscal ledger. But none of these limitations negate the central claim: with disciplined questions, interoperable semantics, causal twins, and auditable measurement, cities can choose better, prove value earlier, and scale faster.

The path forward is clear: specify the question, unify the data, run the models, prioritize and land the interventions, uphold measurement, and sustain and scale the results. With these habits, Digital Twins become not a fashion but a durable capability for public value creation. SURPLUS demands that urban investments be decision-first, semantics-tight, experiment-verified, and auditor-ready—and the LDT Toolbox provides the operational scaffolding to meet that demand. The invitation to practitioners and researchers is straightforward: pick two KBQs that matter, build the minimal twin that can test them, sign the measurement charter with finance, and run one outcome-based contract. The rest is repetition and learning. That is how pilots become programs, and how programs become surplus-maximizing portfolios that residents can feel and treasuries can defend.

8. Acknowledgements

This paper has been supported by the European Union through Citcom.AI [Grant Agreement No. 101100728], EDIC BestUSE [Grant Agreement. No. 101225793], SENSE Citiverse [Grant Agreement No. 101167948], and LDT4SSCC [Grant Agreement No. 101226211]. The authors are solely responsible for this paper; the European Union and its bodies are not liable for any use that may be made of the information contained herein.

References

Abid, A., Haddar, M., Zediri, H., Jamoussi, S., & Kacem, A. H. (2022). Toward mapping an NGSI-LD context model on RDF graph approaches: A comparison study. *Sensors, 22*(13), 4798. https://doi.org/10.3390/s22134798

Artyushina, A. (2020). Is civic data governance the key to democratic smart cities? The role of the urban data trust in Sidewalk Toronto. *Telematics and Informatics, 55*, 101456. https://doi.org/10.1016/j.tele.2020.101456

Bayat, A., & Kawalek, P. (2023). Digitization and urban governance: The city as a reflection of its data infrastructure. *International Review of Administrative Sciences*, *89*(1). https://doi.org/10.1177/00208523211033205.

Boardman, A. E., Greenberg, D. H., Vining, A. R., & Weimer, D. L. (2018). *Cost–benefit analysis: Concepts and practice* (5.ª ed.). Cambridge University Press.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brundkii, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Durmus, E., (…) Liang, P. (2021). On the opportunities and risks of foundation models. *Stanford HAI/CRFM Working Paper*. https://doi.org/10.48550/arXiv.2108.07258

Bozkurt, Y., Rossmann, A., Pervez, Z., & Ramzan, N. (2025). Development and evaluation of an urban data governance reference model based on design science research. *Government Information Quarterly, 42*(2), 102025. https://doi.org/10.1016/j.giq.2025.102025

Bellini, P., Palesi, L. A. I., Giovannoni, A., & Nesi, P. (2023). Managing complexity of data models and performance in broker-based Internet/Web of Things architectures. *Internet of Things, 23*, 100834. https://doi.org/10.1016/j.iot.2023.100834

Birkin, M., Ballantyne, P., Bullock, S., Heppenstall, A., Kwon, H., Malleson, N., Yao, J., & Zanchetta, A. (2025). Digital twins and AI for healthy and sustainable cities. *Computers, Environment and Urban Systems, 120*, 102305. https://doi.org/10.1016/j.compenvurbsys.2025.102305

CitCom.ai TEF. (n.d.). *About CitCom.ai and activities*. https://citcomtef.eu/

City of Bellevue. (2018). *Eyes on Traffic (SCATS before/after studies)* [diapositivas]. https://bellevuewa.gov/sites/default/files/media/pdf_document/Eyes%20on%20Traffic.pdf

City of Los Angeles Bureau of Street Lighting. (n.d.). *LED street lighting: Energy savings and maintenance*. https://bsl.lacity.org/led.html

Data Spaces Support Centre. (2023–2024). *Data spaces blueprint and building blocks*. https://dssc.eu/

data.europa.eu. (2023, November 24th). *LORDIMAS: A digital maturity assessment tool for regions and cities*. https://data.europa.eu/en/news-events/news/lordimas-digital-maturity-assessment-tool-regions-and-cities

ESPON & European Committee of the Regions. (2024, 19 December). *Digital maturity of local and regional governments, 2024: LORDIMAS—Final report*. https://errin.eu/system/files/2025-02/LORDIMAS%20Report_Digital%20maturity%20of%20local%20and%20regional%20governments%202024.pdf

ESPON & Partners. (2024). *Digital maturity of local and regional governments—LORDIMAS report*. https://www.espon.eu/

ETSI. (2021). *ETSI GS CIM 009 V1.5.1: NGSI-LD API*. https://www.etsi.org/

ETSI. (2022). *ETSI GS CIM 019 V1.1.1: Embedding W3C Data Integrity digital signatures into NGSI-LD entities* [PDF]. https://cdn.standards.iteh.ai/samples/63595/423f0e6b0c2f4f12b6f5071727366f50/ETSI-GS-CIM-019-V1-1-1-2022-08-.pdf

ETSI. (2024). *ETSI GS CIM 009 V1.8.1: Context Information Management (CIM); NGSI-LD API* https://www.etsi.org/deliver/etsi_gs/CIM/001_099/009/01.08.01_60/gs_cim009v010801p.pdf

European Commission. (2025, July 16th). *EU Local Digital Twin (LDT) Toolbox*. Interoperable Europe Portal. https://interoperable-europe.ec.europa.eu/collection/ldttoolbox

European Commission. (2025, July 22nd). *European Digital Infrastructure Consortium (EDIC)*. Shaping Europe's Digital Future. https://digital-strategy.ec.europa.eu/en/policies/edic

European Commission. (n.d.). *Common European data spaces*. Shaping Europe's Digital Future. https://digital-strategy.ec.europa.eu/

European Commission — DG CNECT. (n.d.). *Testing and Experimentation Facilities (TEFs) for AI*. https://digital-strategy.ec.europa.eu/

FIWARE Foundation. (2024). *Smart Data Models initiative*. https://www.fiware.org/smart-data-models/

Gray, D., & Shellshear, E. (2024). *Why data science projects fail: The harsh realities of implementing AI and analytics without the hype*. CRC Press. https://doi.org/10.1201/9781032661360

Gu, Y., Quintana, M., Liang, X., Ito, K., Yap, W., & Biljecki, F. (2025). Designing effective image-based surveys for urban visual perception. *Landscape and Urban Planning, 260*, 105368. https://doi.org/10.1016/j.landurbplan.2025.105368

Interoperable Europe. (2024, April 9th). *Assessing smart cities and communities' digital maturity (LORDIMAS)*. https://interoperable-europe.ec.europa.eu/collection/iopeu-monitoring/news/assessing-smart-cities-and-communities-digital-maturity

Jara, A. J., Genoud, D., & Bocchi, Y. (2014). Big data for smart cities with KNIME: A real experience in the SmartSantander testbed. *Software: Practice and Experience, 45*(8), 1145–1160. https://doi.org/10.1002/spe.2274 (ResearchGate)

Jara, A. J., Bocchi, Y., Fernandez, D., Molina, G., & Gomez, A. (2017). An analysis of context-aware data models for smart cities: Towards FIWARE and ETSI CIM emerging data model. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-4/W3*, 43–50. https://doi.org/10.5194/isprs-archives-XLII-4-W3-43-2017

Jeddoub, I., Nys, G.-A., Hajji, R., & Billen, R. (2023). Digital twins for cities: Analyzing the gap between concepts and current implementations with a specific focus on data integration. *International Journal of Applied Earth Observation and Geoinformation, 122*, 103440. https://doi.org/10.1016/j.jag.2023.103440

Jeddoub, I., Nys, G.-A., Hajji, R., & Billen, R. (2024). Data integration across urban digital twin lifecycle: A comprehensive review of current initiatives. *Annals of GIS*. https://doi.org/10.1080/19475683.2024.2416135

Lehtola, V. V., Koeva, M., Oude Elberink, S., Raposo, P., Virtanen, J.-P., Vahdatikhaki, F., & Borsci, S. (2022). Digital twin of a city: Review of technology serving city needs. *International Journal of Applied Earth Observation and Geoinformation, 114*, 102915. https://doi.org/10.1016/j.jag.2022.102915

Lei, B., Liu, P., Milojevic-Dupont, N., & Biljecki, F. (2024). Predicting building characteristics at urban scale using graph neural networks and street-level context. *Computers, Environment and Urban Systems, 111*, 102129. https://doi.org/10.1016/j.compenvurbsys.2024.102129.

Liu, Y., Wang, Z., Ren, S., Chen, R., Shen, Y., & Biljecki, F. (2025). Physical urban change and its socio-environmental impact: Insights from street view imagery. *Computers, Environment and Urban Systems, 119*, 102284. https://doi.org/10.1016/j.compenvurbsys.2025.102284

Libelium. (2023, November 30th). *IRIS 360: AI cameras & platform for smart security & traffic*. https://www.libelium.com/blog/iris-360-ai-cameras-platform-smart-security-traffic/

Living-in.EU. (2024–2025). *Minimal Interoperability Mechanisms (MIMs Plus)*. https://living-in.eu/

Lnenicka, M., Nikiforova, A., Clarinval, A., Luterek, M., Rudmark, D., Neumaier, S., Ković, K., & Rodríguez Bolívar, M. P. (2024). Sustainable open data ecosystems in smart cities: A platform theory–based analysis of 19 European cities. *Cities, 148*, 104851. https://doi.org/10.1016/j.cities.2024.104851.

McKinsey & Company. (2024). *The state of AI in early 2024: Gen AI adoption spikes and starts to generate value*. https://www.mckinsey.com/

Mazzetto, S. (2024). A review of urban digital twins integration, challenges, and future directions in smart city development. *Sustainability, 16*(19), 8337. https://doi.org/10.3390/su16198337

OECD. (2024). *Artificial intelligence, data and competition* (Background note). https://one.oecd.org/

Oldcastle Infrastructure. (2025, June). *CivilSense™ AI leak detection—MidWest case study* [PDF]. https://oldcastleinfrastructure.com/wp-content/uploads/2025/06/CivilSense-AI-Leak-Detection-MidWest-Case-Study.pdf

Ondiviela García, J. A., Blanco Álvarez, E., Hueso Tapia, A., & Cordero Batanzo, J. (2025). Creativity as a driver to resilience in cities: The more creative the more resilient. *SAUC—Journal of Street Art and Urban Creativity, 11*(5). https://visualcompublications.es/SAUC/index.php/SAUC/issue/view/32

Open & Agile Smart Cities (OASC). (2023). *European Local Digital Twin Toolbox project (September 2023 – September 2025)*. https://oascities.org/european-local-digital-twin-toolbox-project/

Palantir Technologies. (2025). *AIP overview*. https://palantir.com/docs/foundry/aip/overview/

Peng, Z.-R., Lu, K.-F., Liu, Y., & Zhai, W. (2023). The pathway of urban planning AI: From planning support to plan-making. *Journal of Planning Education and Research*. https://doi.org/10.1177/0739456X231180568.

Routeware. (2022). *Success story—City of Virginia Beach: Fewer vehicles, lower labor costs, and enhanced safety through route optimization*. https://routeware.com/

SCS Engineers. (2015, June 9th). *Solid waste collection route optimization can save you money*. https://www.scsengineers.com/solid-waste-collection-route-optimization-can-save-you-money/

Sanchez, T. W., Brenman, M., & Ye, X. (2024). The ethical concerns of artificial intelligence in urban planning. *Journal of the American Planning Association*, *91*(2). https://doi.org/10.1080/01944363.2024.2355305.

Shaharuddin, S., Abdul Maulud, K. N., Syed Abdul Rahman, S. A. F., & Che Ani, A. I. (2022). Digital twin for indoor disaster in smart city: A systematic review. *ISPRS Archives, XLVI-4/W3-2021*, 315–322. https://doi.org/10.5194/isprs-archives-XLVI-4-W3-2021-315-2022

Stanford HAI. (2025). *AI Index Report 2025 (Key findings)*. https://aiindex.stanford.edu/

Sun, Y., Song, H., Jara, A. J., & Bie, R. (2016). Internet of Things and big data analytics for smart and connected communities. *IEEE Access, 4*, 766–773. https://doi.org/10.1109/ACCESS.2016.2529723 (OUCI)

Troyanos, K. (2020). Use data to answer your key business questions. *Harvard Business Review*. https://hbr.org/2020/02/use-data-to-answer-your-key-business-questions

United for Smart Sustainable Cities (U4SSC). (2020–2024). *Key performance indicators for smart sustainable cities (Concept and reports)*. https://u4ssc.itu.int/

Universidad Francisco de Vitoria. (2025, April 7th). *La Universidad Francisco de Vitoria presenta el Citizen-Centric Intelligent Cities Research Institute (CCICRI)*. https://www.ufv.es/la-universidad-francisco-de-vitoria-presenta-el-citizen-centric-intelligent-cities-research-institute-noticias-actualidad/

Weil, C., Birbi, S., Longchamp, R., Golay, F., & Alahi, A. (2023). Urban digital twins: Challenges & perspectives for sustainable smart cities. *Sustainable Cities and Society, 99*, 104862. https://doi.org/10.1016/j.scs.2023.104862

World Health Organization. (2024). *Global air quality guidelines: Health risks and recommended thresholds*. https://www.who.int/

Ye, X., Du, J., Han, Y., Newman, G., Retchless, D., Zou, L., Ham, Y., & Cai, Z. (2023). Developing human-centered urban digital twins for community infrastructure resilience: A research agenda. *Journal of Planning Literature, 38*(2), 187–199. https://doi.org/10.1177/08854122221137861

Zhang, F., Salazar-Miranda, A., Duarte, F., Vale, L., Hack, G., Chen, M., Liu, Y., Batty, M., & Ratti, C. (2024). Urban visual intelligence: Studying cities with artificial intelligence and street-level imagery. *Annals of the American Association of Geographers*, *114*(5). https://doi.org/10.1080/24694452.2024.2313515