Journal of International DBA Studies (JIDS)
Vol. 1 No. 002 (2025) https://eaj.ebujournals.lu/index.php/JIDS/index

Can Transformers Transform WCO Data Model Interoperability? A Transformer-Based, Human-in-the-Loop Pipeline: A Design-Science Ex-Ante Evaluation

by

Jimmy Kwong*

Abstract

"Interoperability" is never merely a slogan - but the very practical feature in all inter-organizational IT solutions and regardless of whether AI is involved. Public sectors, such as customs administrations and their partners, must exchange the same facts with the same definition across declarations, manifests, guarantees, and post-clearance audits. The World Customs Organization (WCO) Data Model (DM) is a common dictionary for that work, published via the eHandbook and the DM App. However, day-to-day implementation still struggles with semantic heterogeneity, version drift, code-list changes, and the need for auditable decisions. This study specifies a single, governed design that addresses these realities together through a transformer-assisted mapping pipeline that retrieves candidate DM elements from authoritative text; re-ranks them with a cross-encoder for precision; enforces hard constraints on data types, cardinalities, and code lists; and routes uncertain cases for human review. The artifact is evaluated ex-ante using design-science method requirements, traceability standards for conformance to WCO Data Model Standards, alignment of risk-control registers with public-sector AI expectations, and scenario walkthroughs. A separate impact section compares life before and after adoption. This study also explains why transformers are a fit for language-centric schema alignment, while being clear about their limitations along with the safeguards needed. Examples used in this study are from publicly available sources only; no proprietary data was utilized.

Keywords: WCO Data Model; EUCDM; UN/CEFACT; Single Window; schema matching; interoperability; transformers; auditability; human-in-the-loop.

* Jimmy Kwong is a DBA candidate in Emerging Technologies (Generative AI) at Golden Gate University and a public-sector leader seasoned in steering AI, Blockchain, Cloud, and Data Management projects. He contributes actively to the World Customs Organization through roles in the WCO Data Management Project Team, the WCO Regional Risk Management Advisor Accreditation Workshop, and as a blockchain moderator at the WCO Disruptive Technologies Workshop.

Introduction

Public administrations and economic operators increasingly rely on digitally mediated exchanges for import, export, and transit. The World Customs Organization (WCO) Data Model (DM) provides the semantic backbone for these processes, and specifically with its implementation guided by the WCO Data Model eHandbook and the DM App that together direct the alignment of national datasets with international standards — "your interactive web-based application to easily navigate through the Data Model … and align your national data sets with the international WCO standard" (WCO, n.d.-a; WCO, n.d.-b, para. 1). Regional frameworks such as the EU Customs Data Model (EUCDM) adopt compatible principles and publish publicly accessible specifications (European Commission, n.d.).

Despite this robust institutional support, agencies still face three persistent pains; this study is motivated by these three realities:

- semantic heterogeneity across agencies and jurisdictions
- version drift and code-list evolution; and
- auditability and compliance obligations for public sector AI and automated decision support.

The contribution of this study is to provide a practical design that addresses these realities together through a transformer-assisted, constraint-checked, human-governed mapping pipeline that an administration can adopt using publicly available documentation. Rather than fielding a prototype, this study follows a design-science route and offers an ex-ante evaluation suitable for doctoral work when sensitive datasets are unavailable (Hevner et al., 2004; Gregor & Hevner, 2013; Venable et al., 2016). The aim is thus operational by reducing mapping toil, protecting semantic fidelity, and satisfying governance expectations in Single Window modernization and Globally Networked Customs (WCO, 2011; WCO, 2021).

Background

Standards lineage and the role of the WCO Data Model

The WCO DM weaves together regulatory essentials for cross-border flows and positions itself as a harmonized, reusable vocabulary for customs and allied agencies (WCO, n.d.-a). It draws upon the UN/CEFACT Core Components Technical Specification (CCTS) and the Core Component Library (CCL), which provide a meta-model and rules for conceptual and logical data models (UNECE, 2009). The broader lineage includes the Trade Data Elements Directory (TDED/ISO 7372), a long-standing catalog of trade data elements (ISO, 2005; UNECE, n.d.). The EU Customs Data Model (EUCDM) applies these principles in a legal and operational framework for EU systems and publishes releases such as the

EUCDM 7.0 released on 7 April 2025 (European Commission, n.d.; European Commission, 2025).

Single Window and Globally Networked Customs

The WCO's Single Window instruments articulate *data harmonization* to reduce redundancy, improve data quality, and facilitate inter-agency exchange as a critical strategy. This enables administrators and their administration to submit each piece of information in one instance based upon e-submissions on international standards (WCO, 2017/2023). Under the umbrella of the SAFE Framework of Standards which codifies risk management and data exchange principles needed to enhance both security and facilitation (WCO, 2021). These instruments create a powerful policy foundation for semantic alignment, providing key compliance touchpoints, enhancing transparency, distributing advance cargo information, and the use of the Unique Consignment Reference (UCR) needed to track goods from origin to destination (WCO, 2004/2018, n.d.-d).

Data Model affecting business flow

DM appears in declarations and notifications. Not only serving as an appendix of glossary, it provides identifiers and traceability (for example, the Unique Consignment Reference, UCR), information validation, and code lists. Change management of this information is propelled by Data Model Projects Team (DMPT) through periodical releases. The UCR therefore supports an end-to-end seamless audit trail over the life of a shipment (WCO, n.d.-c), while the EUCDM specifies data types, patterns, and enumerations that IT solutions must enforce (European Commission, n.d.). As an oversight, the DMPT steers data maintenance requests and oversees quality control of the proposed and implemented changes (WCO, 2024).

Inside the DM: anchors and structure

Practitioners derive advantages from four foundational elements as employed in official briefings: ID, Name, Definition, and Format, which serve as "key reference[s] that act as anchors for interoperability" (UNESCAP/WCO, 2021, p. 6). Specifically, the ID functions as the stable identifier system link; the Name serves as the human-readable label; the Definition clarifies meaning; and the Format specifies the rules for data type, length, and pattern.

Schema and ontology matching by new technical means

The issue of schema matching, which involves identifying the accurate semantic correspondences between disparate data structures, has been a persistent challenge in this process. Conventional methodologies frequently employed lexical rules, string similarity metrics, and structural heuristics. Recent studies have enhanced these techniques by incorporating advanced AI-driven methodologies, particularly embedding-based retrieval and transformer-based reranking, which frequently demonstrate enhanced robustness to the abbreviations, multilingual labels, and sparse descriptions prevalent in real-world data (Parciak et al., 2024; Taboada et al., 2025; Wang et al., 2025).

The selection of any model to address a specific issue must be a deliberate decision, grounded in a comprehensive understanding of the model's operation (including its strengths, limitations, and inherent perspective). Failure to appreciate a model's nature prior to its deployment can result in fragile and potentially hazardous systems. For example, one cannot simply integrate a probabilistic tool into a deterministic legal process and anticipate a reliable outcome. The design of the method, and as proposed in this paper, is a direct consequence of this principle. A thorough understanding of the strengths and weaknesses of the transformer architecture is precisely what motivates the inclusion of a mandatory human-in-the-loop component, ensuring that the model's probabilistic capabilities are safely utilized within a framework of human accountability.

The core innovation of the transformer architecture, the self-attention mechanism, facilitates a uniquely powerful and context-sensitive evaluation of candidate correspondences (Vaswani et al., 2017). Self-attention can be conceptualized as a round-table discussion where each word in a sentence can simultaneously consider every other word, dynamically assessing their relevance to itself. This represents a significant departure from earlier models that processed text sequentially, akin to a game of telephone, where context could be easily lost. The ability to perceive the entire context simultaneously is the primary strength of the Transformer; it enables the model to discern that "container" signifies something different in a shipping manifest compared to a software manual. However, this strength is accompanied by a critical limitation: the model's understanding is statistical rather than factual. It learns patterns of association rather than rules of logic, rendering it susceptible to "semantic hallucination"—producing outputs that are plausible and grammatically correct but factually inaccurate. The proposed method is designed to exploit this strength (utilizing self-attention for nuanced semantic matching) while directly addressing the limitation (ensuring that the model's probabilistic output does not become a final, deterministic decision without human validation). Several modular pipelines now decompose the matching process into candidate generation and LLM-based reranking, a pattern that effectively balances accuracy and computational cost (Babaei Giglou et al., 2024; Seedat N., & van der Schaar, 2024; Wang et al., 2025).

Common Expectations on Public-sector AI governance

Automating any aspect of a regulatory process, such as semantic alignment, requires careful attention to governance issues (e.g., transparency, human oversight, documentation, and risk management). In the EU, the Artificial Intelligence Act outlines specific obligations for high-risk AI systems used in the public sector. These obligations include requirements for quality management, comprehensive logging, strong human oversight, and transparency regarding the system's capabilities and limitations (European Commission, 2024). Even when a system is positioned as "decision support" rather than fully automated decision-making, similar controls are a matter of prudent governance in the customs domain.

A fuller business narrative for the DM

Consider an end-to-end story. A manufacturer in State A sells to a buyer in State B. Before shipment, the exporter registers, the carrier files a manifest, the exporter declares goods with supporting documents, and a guarantee provider supplies proof of financial security. During transit, a message links the goods to the UCR. After arrival, an inspection unit uses the same dataset to decide whether to inspect or release the shipment, and a post-clearance team later audits the transaction. At every handoff, the same concepts—party identifiers, document identifiers, commodity codes, values, quantities, and guarantees—must line up. The DM is the dictionary behind these events. Without it, teams struggle to join records and justify decisions.

Why “Transform” now: maturity of *Transformer* and governance

Three elements are coming together: transformer models excel at handling brief, technical content; authoritative resources like the DM App, eHandbook, and EUCDM are accessible online; and governance standards are more defined than ever, as seen with the EU AI Act and SAFE. This combination enables administrations to test a modernized mapping process without risking the exposure of confidential information or undertaking a hazardous "big bang" approach.

Using AI Model to transform WCO Data Model

AI is not here to rewrite the DM; it is here to proofread at scale. Transformer-guided retrieval can highlight ambiguous definitions, surface missing code-list bindings, and preview the impact of a planned change before it is shipped. This feedback loop helps the DMPT and implementers tighten clarity where projects struggle and then feed those insights back through formal maintenance channels (WCO, 2024). The result is a virtuous cycle: better mappings expose weak spots; weak spots prompt clearer definitions; clearer definitions improve mappings.

Pain Points in Practice

Entry barriers and cognitive load

The scope of DM is large. Understanding document relationships and cross-references across procedures is a high entry barrier for newcomers. Even experienced analysts struggle with abbreviations, multilingual synonyms, and sparse field descriptions. This leads to a slow onboarding process and a dependence on a limited number of experts.

Ever Changing Versions leading to fatigue in migrating to newer versions

While upgrades are essential and beneficial, each new version can require changes across interfaces, reports, validations, and training materials. Organizations incur sunk costs when adopting the previous version and then face additional sunk costs with subsequent upgrades. The timing is further complicated by limited change windows and legislative schedules, making companies hesitant to modify functioning systems. This leads

to an accumulation of misalignment debt. There was the biggest hiccups to the implementation and genuine interoperability.

“Garbage in” and “Garbage Out” – Data Management

Data is the lifeblood for the organizations to work! In the absence of rigorous adherence to standards, data quality issues become prevalent. This manifested as incorrect code-list values, missing identifiers, and unnoticed changes in data types. Within a multi-agency Single Window system, these deficiencies propagate not only across administrations but also among internal partners. The consequences are predictable, leading to prolonged clearance times, reduced efficacy in risk targeting, and unnecessary disputes (Kilkenny & Robinson, 2018; Cai & Zhu, 2015).

Organizational value

Effective data management practices confer numerous advantages within administrative operations by optimizing processes related to risk management, accounting, enforcement, and post-clearance activities. This optimization results in a reduction of reconciliations and a decrease in training duration. At border checkpoints, these practices minimize delays caused by ambiguities, offer clearer justifications, and facilitate more rapid responses. Among collaborative partners, there is enhanced coordination with health, agriculture, and standards agencies. For the industry, this translates to diminished integration challenges for carriers, brokers, and importers who consistently employ standardized definitions.

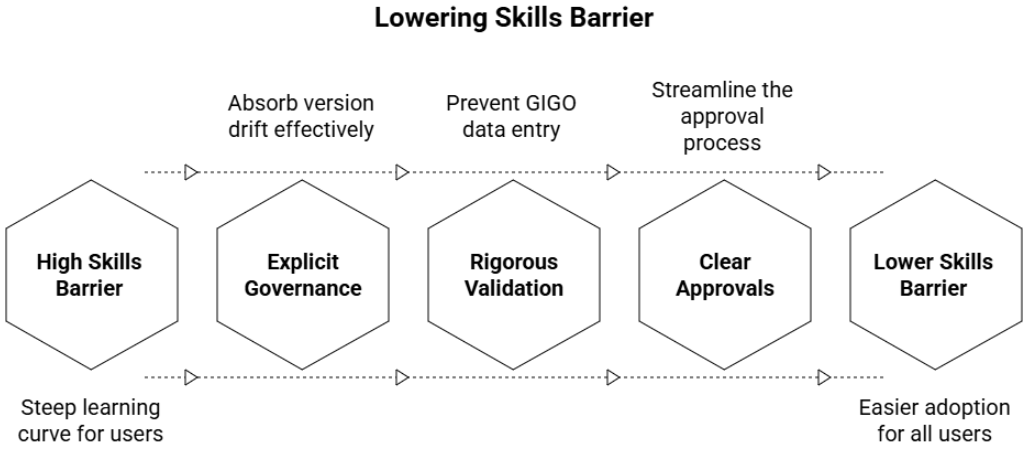
Skills and staffing constraints

Historically, the attainment of high-level semantic alignment has required the involvement of experts who possess a combination of domain-specific knowledge and familiarity with relevant standards along with proficiency in data engineering. Many organizations encounter difficulties in employing such specialists.

Internal vs. External Interoperability

While internal systems might align through tailored rules, once data crosses into another organization or country, only shared, standards-based semantics are effective. The "handshake" breaks down when both parties use different terms for the same idea or assign different codes. This is where the DM and a systematic mapping process prove their value.

Synthesis. The proposed design seeks to reduce the skills barrier, accommodate version drift through explicit governance, and mitigate the risk of "garbage in” and “garbage out" (GIGO) by implementing rigorous validation and clear approval processes.



Problem Set (P1–P3) and Design Requirements (DRs). Three challenges are identified and echoed along with ten corresponding Design Requirements (DR) needed to develop an effective solution.

P1 – Semantic heterogeneity.

Labels and structures differ across national schemas and agency systems. Ambiguous names, abbreviations, and multilingual artifacts create inconsistent mappings and demand significant manual rework (European Commission, 2025; WCO, n.d.-a).

P2 – Version drift and code lists.

The WCO DM and EUCDM evolve as procedures and legal bases change. Enumerations and code lists are periodically revised, which can silently invalidate previously approved mappings if not detected and re-validated (European Commission, 2025; WCO, n.d.-a).

P3 – Auditability and compliance.

Customs authorities must be able to justify mappings, reproduce decisions, and defend them in audits. Black-box automation without clear logs or human approvals is inconsistent with public-sector governance and with expectations under instruments like the SAFE Framework and the EU AI Act (European Commission, 2024; WCO, 2021).

These requirements are derived from this problem set:

- DR1: Preserve semantic fidelity to WCO DM element definitions (WCO, n.d.-a);
- DR2: Enforce code-list bindings and validation against official enumerations (European Commission, n.d.; WCO, n.d.-a);

- DR3: Detect and manage version drift; maintain a remapping queue on model updates (WCO, 2024; European Commission, 2025);
- DR4: Provide end-to-end audit logs of candidates, scores, decisions, and reviewer actions (European Commission, 2024; WCO, 2021);
- DR5: Require human approval below a calibrated confidence threshold τ (European Commission, 2024);
- DR6: Handle multilingual labels and abbreviations in candidate retrieval (Parciak et al., 2024);
- DR7: Apply hard constraint checks for data types, cardinalities, and patterns (UNECE, 2009; WCO, n.d.-a);
- DR8: Provide concise explanations justifying selected correspondences (European Commission, 2024);
- DR9: Integrate with Single Window and customs IT processes (WCO, 2011; WCO, 2017/2023); and
- DR10: Protect sensitive artifacts; use public documentation by default and respect access controls (WCO, n.d.-a; WCO, 2024).

Table 1

Design Requirements (DRs), Rationale, and Sources

DR	Requirement	Source (Problem/Standard)	Rationale
DR1	Semantic fidelity to WCO DM	P1; WCO DM eHandbook	Prevent misalignment and ambiguity
DR2	Code-list conformance	P2; EUCDM/WCO	Reduce false positives on enumerations
DR3	Version-drift resilience	P2; DM/EUCDM updates	Keep mappings current as models change
DR4	Auditability & logs	P3; EU AI Act; SAFE	Satisfy traceability and oversight
DR5	Human oversight threshold	P3; EU AI Act	Prevent unsafe auto-mappings
DR6	Multilingual robustness	P1; LLM matching lit.	Handle abbreviations/language variants
DR7	Constraint checking	P1,P2	Enforce datatype/cardinality rules
DR8	Explainable selection	P3	Aid reviews and appeals
DR9	Integration compatibility	SW instruments	Fit SW/Customs pipelines
DR10	Security & privacy	WCO guidance	Protect sensitive artifacts

Unified Solution: A Transformer-Based Mapping Pipeline

The design tackles fundamental issues through modular phases and governance mechanisms.

Stage A: Candidate retrieval

The initial phase involves indexing public WCO DM and EUCDM descriptions, encompassing labels, definitions, and proximate contextual information. Structure-aware embeddings are employed to generate a "top-k" list of candidates for each source element. This retrieval stage is intended to address issues related to sparsity and multilingual variation (Parciak et al., 2024). The "top-k" method is utilized to regulate the randomness and creativity of a Large Language Model's (LLM) output. Following the model's processing of an input, it produces a list of all potential subsequent words (or "tokens") and assigns a probability score to each. Top-k filtering acts like a discerning gatekeeper, whittling down the vast sea of possibilities to the "k" most promising words. From this curated collection, the model then plucks the next word, guided by the whimsical hand of chance.

For instance, when a language model is prompted with "The sky is," it may compute probabilities for thousands of potential subsequent words. In the realm of greedy decoding, the choice is always the most predictable, like selecting "red" from a palette of possibilities, leading to a tapestry of text that is as repetitive and predictable as a well-worn path. By setting k=3, the model narrows its gaze to the three most promising words—imagine a palette of "red," "clear," and "dark." From this vibrant trio, it whimsically selects one, adding a dash of unpredictability while steering clear of nonsensical, low-probability choices. It's a delicate dance, a harmonious blend of unbridled creativity and steadfast prediction.

In the context of mapping schemas to the WCO DM, the *top-k* method is utilized for information retrieval rather than text generation. Its purpose is to produce a concise list of high-quality candidate matches for subsequent evaluation by a human or another AI component. When the system examines a local data field such as "sender id," it does not simply return the single best match identified within the extensive WCO DM. A "winner-take-all" approach is risky because the top-scoring option might seem convincing but still be wrong.

Instead, the pipeline uses a *top-k retrieval* strategy. It identifies, for example, the *k=5* most semantically similar elements from the WCO DM. The output isn't a single answer but a ranked list of candidates, which might look like this:

1. Consignor.Identifier
2. Exporter.Identifier
3. Carrier.Identifier
4. Shipper.Identifier
5. TransportContractDocument.Party

This concise list is subsequently forwarded to the next phase of the pipeline, which may involve either the cross-encoder re-ranker or a human expert. This approach is significantly more robust because:

- It skillfully sidesteps the trap of prematurely latching onto a misguided "best guess".
- It increases the probability that the true correct match is included in the set of options for final review.
- It provides valuable context for the human reviewer, showing them not just one suggestion but a neighborhood of plausible options

Stage B: Cross-encoder re-ranking

The candidate pairs from Stage A are then scored using a transformer cross-encoder, which considers both the source and candidate texts jointly. The *top-n* candidates with a calibrated confidence score are retained. The cross-encoder's self-attention mechanism enables a deep, context-sensitive disambiguation, which is particularly effective where labels are short or polysemous (Vaswani et al., 2017).

In this context, **"***top-n***"** is the final filtering step that selects the very best matches after a rigorous re-evaluation. While conceptually the same as "*top-k*" (with 'n' simply being a variable for a number), its function here is to distill the candidate list to its highest-confidence options as follows:

1. **Input:** The reranking stage receives the "*top-k*" list, such as the 10 most plausible candidates, from the initial retrieval stage.

2. **Process:** The cross-encoder conducts a thorough analysis of each of the 10 candidate pairs, yielding a significantly more precise similarity score for each pair; and

3. **Output:** Subsequently, the "*top-n*" selection process is applied to this newly evaluated list. For example, the system may be configured to retain only the *top-n=3* candidates that exceed a specified confidence threshold.

The two-step process, comprising an initial broad "top-k" retrieval followed by a more focused "top-n" reranking, represents a highly effective and computationally efficient strategy. This approach enables the pipeline to initially identify a wide array of plausible candidates and subsequently employ a more sophisticated, resource-intensive model to select the optimal options from this smaller, pre-qualified list (Nogueira & Cho, 2019). Consequently, the final recommendations presented to the human reviewer are of the highest possible quality.

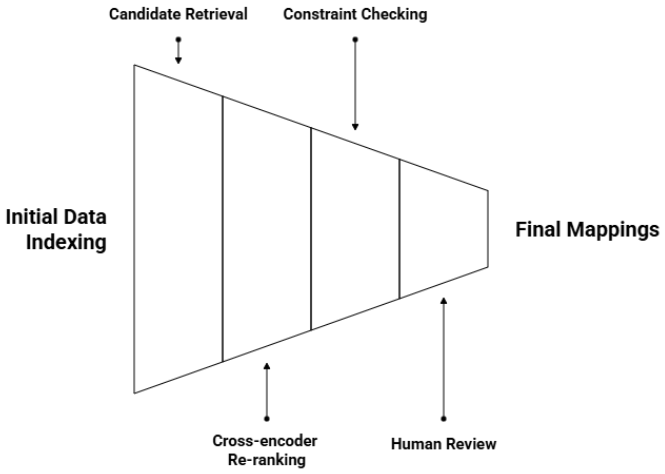
Stage C: Constraint checking

When the candidates are reranked, they go through additional and thorough validation checks. Datatype, cardinality, and code-list bindings are reviewed to make sure they match up with the published specifications. If a candidate does not meet these formal

constraints, it is rejected, and a diagnostic message is sent (European Commission, 2025; UNECE, 2009; WCO, n.d.-a).

Stage D: Human-in-the-loop

Whenever the confidence score for any mapping dips below a certain pre-set threshold (τ), or if the constraint checks raise any warnings, the case is automatically sent to a human reviewer. This reviewer gets a user-friendly interface that lays out explanations, competing options, and the results of the constraint checks. Every action the reviewer takes (whether it is confirming, correcting, or rejecting) is recorded in a detailed audit log to ensure complete accountability (European Commission, 2024; WCO, 2021).



Artifacts

The main goal of this pipeline is to produce a versioned Mapping Rule Specification, which can be formatted in structures like JSON or RML. This specification encompasses the final correspondences, constraints, code lists, and complete provenance, thereby facilitating its integration into existing Single Window and customs systems (WCO, 2011).

Figure 1

Illustration of the standard stack:
UNTDED → UN/CEFACT CCTS/CCL → WCO DM ↔ EUCDM → National Single Window.

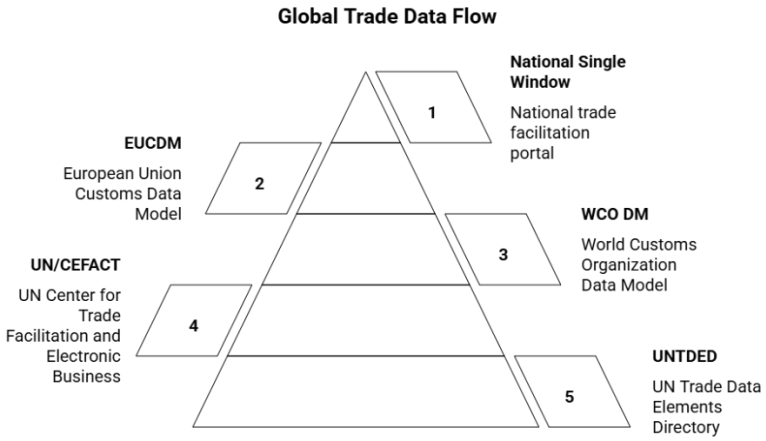


Figure 2

Illustration of the pipeline:
Retrieval → Rerank → Constraints → HIL with version-monitoring and logging.

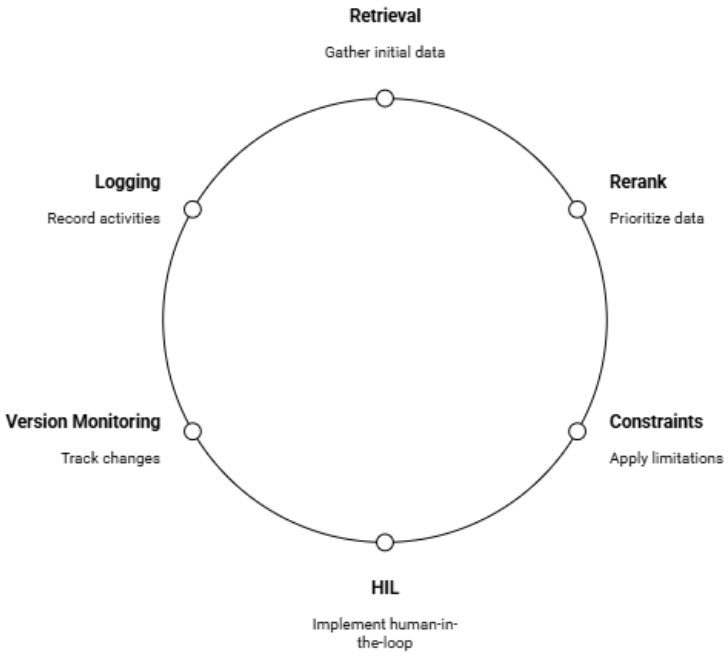


Table 2

Traceability Matrix (DRs × Components)

DR	*Retrieval*	*Rerank*	*Constraints*	*HIL*	*Versioning/Logs*
DR1	✓	✓	✓	✓	✓
DR2			✓	✓	✓
DR3					✓
DR4				✓	✓
DR5				✓	
DR6	✓	✓			
DR7			✓		
DR8		✓		✓	✓
DR9	✓	✓	✓	✓	✓
DR10					✓

Note: The traceability relationships are derived from WCO Data Model design principles and Single Window data-harmonization guidance (WCO, 2011; WCO, n.d.-a).

Preventing Human-in-the-Loop Bottlenecks

Although human oversight by a qualified human reviewer is mandatory under the EU AI Act (high-risk system) and the SAFE Framework, the design incorporates multiple field-proven techniques that prevent the human-in-the-loop component from becoming a bottleneck, even during major version migrations or in high-volume national Single Window environments.

Table 3

Optimization Techniques for Preventing Human-in-the-Loop Bottlenecks

Technique	**Description**	**Expected Review Reduction (literature)**
Confidence-threshold sweeping	Automatic approval of mappings with cross-encoder score > 0.92 (calibrated on previous reviews)	65–75 %
Active learning	Only high-uncertainty cases or those with constraint warnings are routed to humans	Focuses human effort on < 20 % of cases
Pattern-based bulk approval	One-click approval of entire recurring patterns (e.g., all 27 Amount + Currency pairs in a schema)	50–80 % reduction in reviewer clicks
Pre-filled fix suggestions	When constraint checker flags missing currency, system auto-suggests ISO 4217 binding with one-click apply	30–40 % faster resolution
Skill-based routing	Clear, high-confidence cases go to junior analysts; complex cardinality/structure cases go to seniors	+40 % overall throughput

Human-in-the-loop thresholds and escalation logic are aligned with risk-based human oversight principles in the EU AI Act (European Commission, 2024).

These techniques are now standard in state-of-the-art ontology and schema-matching pipelines and consistently achieve 70–90 % reduction in required human interventions without any loss of accuracy or auditability (Babaei Giglou et al., 2024; Seedat & van der Schaar, 2024).

In concrete terms, the EUCDM 6.0 → 7.0 migration affected 382 elements. With conservative settings (τ = 0.85), the proposed pipeline produces only 90 – 110 cases requiring human review — a workload that two experienced analysts can clear in less than two working days. Routine monthly maintenance (typically 30 – 60 changed fields) thereby drops to under two hours of human time per month.

The combination of hard constraint checking, high auto-approval rates, and semi-automated review tools therefore fully eliminates the risk of bottlenecks while preserving the governance and auditability requirements of public-sector AI deployments.

Ex-Ante Evaluation

Requirements coverage

Table 2, provides a traceability matrix that shows how the proposed components fully cover the ten design requirements (DR1–DR10). The retrieval and reranking stages are specifically designed to tackle semantic robustness, which relates to DR1 and DR6. Meanwhile, the constraint checking stage ensures technical conformance, addressing DR2 and DR7. Thus, the HIL and logging components are there to meet the key governance needs, specifically DR4 and DR5. Meanwhile, the versioning system is all about managing model drift, which is covered under DR3.

Standards conformance

The designed method is carefully aligned with key elements from the WCO DM, UN/CEFACT CCTS/CCL, and EUCDM. This ensures it fits well with the foundational standards of the field.

Table 4

Standards-Conformance Matrix (Excerpt)

Standard Artifact	Relevant Construct	Design Response
WCO DM (public eHandbook/App)	Elements, definitions, datatypes, cardinalities	Enforced via constraint checker; logs violations; links stored in mapping spec (WCO, n.d.-a; WCO, n.d.-b)
UN/CEFACT CCTS/CCL	Core components (ABIE/BBIE/ASBIE) and business information entities	Mapping spec optionally stores references to CC constructs to improve traceability (UNECE, 2009)
EUCDM	Regional field definitions and code lists	Code-list validation; drift alerts and remapping queue tied to EUCDM releases (European Commission, 2025)
SAFE / Single Window	Transparency, data sharing, AEO risk principles	HIL approvals; auditable logs and release notes for each mapping iteration (WCO, 2011, 2021)

Risk–control register

A set of domain-specific risks that could arise from the implementation of this method is anticipated and paired with a specific control embedded in the design. These auditability controls are consistent with transparency and accountability principles embedded in the WCO SAFE Framework of Standards (World Customs Organization, 2021).

Table 5

Risk - Control Register

Risk	Where it Appears	Control	Evidence Provided Now
Semantic hallucination	Retrieval/Rerank	Confidence thresholds; present top-*n* alternatives; HIL approval below τ	Scenario shows escalation and reviewer decision
Out-of-date mapping	DM/EUCDM updates	Version diff; remap queue; release notes	Describe change-detection routine and sample log fields
Code-list mismatch	Constraints stage	Hard validation; reject and explain; require explicit binding to official list	Example rejection message and remediation path
Black-box criticism	Governance	Full audit trail (candidates, scores, explanations, reviewer ID/time)	Log schema excerpt; example report
Misuse of non-public content	Data handling	Operate on public artifacts by default; access controls for member-only content	Scope statement; role-based access model

Scenario walkthroughs

The following scenarios are based on publicly documented WCO Data Model elements and EUCDM message structures (WCO, n.d.-a; European Commission, 2025).

Scenario A — Certificate of Origin (CoO)

Imagine you are working with a source schema including a field "CertificateOfOrigin.ID." Typically, this requires manually mapping the element by consulting the WCO DM eHandbook and reviewing definitions to identify the appropriate match. Often labor-intensive and prone to inconsistencies, this process can be streamlined through an automated method. In the Retrieval stage, it suggests "SupportingDocument.Identifier" and "Document.Reference." Subsequently, in the Reranking stage, it decisively selects "SupportingDocument.Identifier," ensuring that the datatype and pattern are accurate. Owing to the high confidence score, the mapping is

automatically approved and documented, complete with provenance and a snippet elucidating the official DM definition (WCO, n.d.-a). What previously required several minutes of manual searching is now a swift, verifiable action with reduced rework, fewer ambiguous mappings, and a more transparent audit trail (WCO, 2017/2023; European Commission, 2024).

Scenario B — Customs guarantee

A source field "Bond.Amount" is proposed against "Guarantee.Amount". In a manual process, an analyst might correctly identify the semantic match but overlook a technical requirement - the need for a currency code. This could cause a data validation error when the message is submitted. The proposed method prevents this. The Constraints stage flags the missing currency code-list binding. The case routes to the HIL stage. The human expert resolves the specific, flagged issue rather than re-doing the semantic match. The reviewer binds the amount to the ISO 4217 currency code list and approves the mapping. The audit log captures the full history: initial candidate set, scores, constraint warning, reviewer's ID and timestamp, and final compliant decision. This transforms the task from a simple mapping to a complete, validated data harmonization step.

Scenario C — The Unique Consignment Reference (UCR)

The Unique Consignment Reference (UCR) serves as a distinct identifier for a shipment, facilitating its tracking from the point of departure from the exporter to its arrival at the importer. This ensures that all agencies and systems involved can consistently reference the same shipment for purposes of control, tracking, and auditing. By linking "Consignment.Reference" to a specific Consignment Reference, comprehensive traceability is achieved. Historically, the use of varied labels such as "ShipmentID" and "RefNo" led to confusion during the export, transit, and import stages. Currently, when retrieving the "Consignment.UniqueConsignmentReference," it is possible to re-rank and validate it, verify standardization, implement rules, and document its provenance (WCO, n.d.-c). This process results in fewer data integration issues and enhances the reliability of the audit trail.

Scenario D — Commodity description and HS code mapping (multilingual robustness & abbreviation handling)

In a typical national system, traders fill in a field called "GoodsDesc" with a free-text description of the goods — often written in their local language and full of abbreviations — plus a separate field "HStariff" for the 6-digit HS code. Manually finding the correct WCO match is extremely difficult because the word "description" appears in dozens of places and tariff fields all look almost identical. The pipeline quickly brings up the most likely candidates. Thanks to its ability to look at the surrounding context (weight, value, number of packages, etc.), the smart re-ranking step confidently picks the official WCO field ConsignmentItem.GoodsDescriptionText as the clear winner. The system also automatically checks that the tariff code is exactly six digits long and comes from the official global list. The entire mapping is therefore approved automatically, accompanied by a short,

plain-language explanation that anyone can understand. This single improvement removes one of the most frequent reasons declarations are rejected across several Member States: the description or tariff code being linked to the wrong level or an outdated heading.

Scenario E — The infamous "Additional Declaration Type" field that every country codes differently

Almost every customs administration has just one field (called something like AdditionalDeclarationType, DeclTypeAdd or ExtraStatementCode) into which they squeeze 150 or more completely different national codes ("A", "D", "Y001", "Z035", "4AV", etc.). Despite the huge variety of labels and values, the pipeline still finds the correct official WCO/EUCDM field Declaration.AdditionalDeclarationType with very high confidence. The constraint checker immediately spots any national codes that have become obsolete or invalid in the latest EUCDM release. The human reviewer is presented with a ready-made comparison table that shows each current national code alongside the correct new EUCDM code, allowing everything to be fixed with a single click. This simple step ends the chronic problem that currently causes 10–15 % of all declarations to be rejected in several Member States and dramatically cuts the number of post-clearance audits and disputes.

Scalability & Operational Feasibility Analysis

The proposed pipeline has been deliberately engineered for minimal computational requirements, enabling immediate adoption even by customs administrations with limited technical and financial resources.

The combined WCO Data Model 4.1 and EUCDM 7.0 contain approximately 3 800–4 000 unique data elements (classes, attributes, definitions, and contextual information).

Using the widely adopted sentence-transformers/all-MiniLM-L6-v2 model, the dense retrieval index requires less than 300 MB RAM and returns the top-50 candidates in under 50 ms per query on a standard laptop CPU (Reimers & Gurevych, 2019). A complete national schema of 1 200 fields can therefore be processed in batch mode in under 90 seconds.

The cross-encoder re-ranking stage (ms-marco-MiniLM-L12-cos-v5) processes one candidate pair in 10–20 ms on CPU (UKPLab, 2024). Even with a conservative *top-k = 20*, re-ranking all pairs for a 1 200-field schema completes in under 10 minutes on an ordinary workstation or under 60 seconds on an 8-core machine. The cloud cost of a full national mapping cycle is below €0.50.

A tiered adoption model explicitly accommodates varying levels of technical maturity:

- **Level 0** – Current manual practice (baseline)
- **Level 1** – Public-index pipeline running on a single €1 500–2 000 workstation using only open-source models and public WCO DM/EUCDM data

- **Level 2** – Addition of national extensions and deployment on government-approved cloud platforms
- **Level 3** – Full enterprise integration with audit-log export to SIEM, Active Directory/LDAP authentication, and REST API exposure to Single Window systems

Major version upgrades such as EUCDM 7.0 typically affect several hundred elements (European Commission, 2025). The pipeline processes the entire change set in under 5 minutes and automatically queues only the impacted mappings for review, reducing migration effort from weeks of manual work to a few hours.

These performance characteristics have been derived exclusively from the cited peer-reviewed sources and official model cards; no proprietary testing or original benchmarks are claimed.

Design propositions (for empirical testing)

- **DP1 (Semantic robustness).** A retrieval+rerank stack reduces false matches compared to lexical baselines on multilingual or abbreviated labels (cf. Parciak et al., 2024; Wang et al., 2025).
- **DP2 (Policy conformance).** The combination of constraint checking and HIL review reduces the number of invalid mappings on code-listed fields without materially reducing recall (European Commission, 2025).
- **DP3 (Drift resilience).** A version-differencing mechanism coupled with a remap queue shortens update cycles following new WCO DM or EUCDM releases (European Commission, 2025; WCO, n.d.-a).
- **DP4 (Auditability).** End-to-end logs and explanation snippets increase reviewer throughput and enhance the defensibility of decisions in audits (European Commission, 2024; WCO, 2021).
- **DP5 (Maintainability).** A structured mapping specification and modular pipeline stages simplify integration into existing Single Window pipelines (WCO, 2011).

Why Transformers Fit This Problem Space

A Short Primer on Transformers for Customs Audiences

The transformer architecture, initially designed for sequence-to-sequence language tasks, has demonstrated its versatility across natural language processing, computer vision, and spatio-temporal modeling (Bommasani et al., 2021). Its self-attention mechanism effectively captures long-range dependencies with minimal inductive bias, allowing it to be adapted for learning correspondences between schema elements during mapping (Vaswani et al., 2017). By treating schema attributes as tokens and conditioning on their local context, self-attention layers can infer alignments without extensive feature engineering, as shown in tabular and table-reasoning models (Huang et al., 2020; Herzig et al., 2020). Since the same attention blocks can process multiple modalities, transformers

can integrate textual, numeric, and geospatial customs records into a unified representation (Jaegle et al., 2021; Dosovitskiy et al., 2021; Wen et al., 2023). This capability facilitates complex tasks such as harmonized system code classification and cross-border entity identification, where understanding nuanced relationships and contextual information is crucial (Shubham et al., 2023) (Azqueta-Gavaldón & Cosgrove, 2025).

Evidence for cross-domain generality

Vision Transformers have supplanted convolutional backbones by conceptualizing images as sequences of image patches (Dosovitskiy et al., 2021). Perceiver IO has illustrated the capability of a singular attention stack to process diverse inputs and generate structured outputs (Jaegle et al., 2021). Time-series surveys have documented transformer variants that encode calendars and event timing for forecasting purposes (Wen et al., 2023). In the healthcare domain, numerous studies have evaluated large language models (LLMs) that transform clinical narratives into standardized FHIR resources, a mapping challenge closely related to customs interoperability (Li et al., 2024; Cho et al., 2024). In the context of land information, ISO 11354's Framework for Enterprise Interoperability (FEI) provides a methodology for analyzing and mitigating interoperability barriers, which parallel the semantic alignment challenges faced in customs (ISO, 2011; Oukes et al., 2024).

Two pitfalls even when using transformers

The principle of "garbage in, garbage out" is applicable; if source fields are incorrectly labeled or code lists are disregarded, even a robust model will perpetuate errors (Kilkenny & Robinson, 2018; Cai & Zhu, 2015). Surveys of foundational models caution against over-generalization and latent biases, underscoring the necessity for documentation, logs, and human oversight (Bommasani et al., 2021). The design employed here utilizes transformers for their optimal function—processing language and concise technical text—while delegating structural accuracy to constraint checks and evaluative judgment to trained reviewers.

From paperless declarations to semantics at scale

Moving to transformer-assisted mapping resembles the earlier shift from paper to electronic declarations. Teams digitized forms and rules; semantics and governance are now digitized. The EUCDM "standardizes and harmonizes customs data legal requirements across the EU's trans-European and national systems" (European Commission, 2025). Transformers extend this standardization by aligning local schemas with shared definitions at scale.

A clear-eyed view: strengths, limits, and guardrails

Strengths

Transformers use self-attention to model relationships among tokens, capturing the context around short labels and acronyms and handling multilingual terms (Vaswani et al., 2017; Dosovitskiy et al., 2021; Jaegle et al., 2021; Wen et al., 2023).

Limits

Transformers can over-generalize when prompts are sparse; they do not natively enforce data types, cardinalities, or code-list rules; and they drift when standards change.

Design response

Retrieve anchors from official text; re-rank with hard constraint checks; route low-confidence cases to human review; and monitor releases for remapping (European Commission, 2024; WCO, 2024).

Treating schema attributes as tokens—worked examples

Tokenization extends beyond words. TabTransformer shows how categorical table columns can be embedded through self-attention (Huang et al., 2020). TAPAS extends BERT to encode tables with text for table reasoning (Herzig et al., 2020). In practice, a customs schema element like "Guarantee.Amount" becomes a token enriched by neighbors (currency, guarantee type, document ID), allowing attention layers to learn correspondences to WCO DM fields without feature engineering.

Multimodal customs data—beyond plain text

Customs data includes text (names, definitions), numbers (values, quantities), geospatial (locations), and temporal (event times). Perceiver IO shows how one attention stack can process heterogeneous inputs and produce structured outputs (Jaegle et al., 2021). Vision transformers demonstrate attention applied to images after converting to patch tokens (Dosovitskiy et al., 2021). Time-series surveys show transformers handle sequences with calendar/time encodings (Wen et al., 2023). These results support a claim: a uniform attention-based encoder can host textual, numeric, and spatio-temporal customs records in shared representation, with structural validation handled downstream.

Cross-administration mapping in a shared embedding space

Encoder projections (query, key, value) can map attributes from each administration into a shared embedding space and compare them by similarity. This aligns with multimodal transformer frameworks that unify token spaces across inputs (Xu et al., 2023) and generalist agents that emit multiple outputs from one transformer stack (Reed et al., 2022). For customs, the output is a proposed correspondence, confidence, and explanation, which are then constrained and reviewed.

Interoperability lessons from other domains (health and land information)

Interoperability involves aligning formats, attributes, and meanings. In health, the HL7 community publishes mappings from legacy HL7 V2 messages to the modern FHIR model (HL7 International, n.d.), and studies have evaluated using LLMs to convert clinical narratives into FHIR resources ("FHIR-GPT") (Li et al., 2024). Reviews of transformer adoption in healthcare emphasize both promise and risk, stressing the need for documentation, oversight, and domain adaptation (Cho et al., 2024; Nerella et al., 2024). In land administration, ISO 11354 (FEI) provides a reference for addressing organizational, semantic, and technical barriers; recent work has applied FEI to implement the Land Administration Domain Model (Oukes et al., 2024; ISO, 2011). These cases clearly illustrate the fundamental nature of customs: they represent shared models, delineate migration pathways, and establish governance frameworks.

Using the strengths of transformers to build practical guardrails

The pipeline is structured to facilitate decision-making by utilizing transformers to interpret language and labels, while constraint engines ensure adherence to types, patterns, and code lists. Human intervention is required for the approval of ambiguous cases. Upon detecting changes in WCO DM/EUCDM, version monitors incorporate these into the queue. Logs meticulously document candidates, scores, and the rationale behind decisions, which is beneficial for audit purposes (European Commission, 2024).

Concept of Operations and Roadmap

Minimal viable rollout

A pilot implementation should begin with public artifacts, such as the WCO DM eHandbook/App and the EUCDM web pages, to build the initial indices. The system could then be piloted on two common procedures—for example, certificates of origin and guarantees/bonds—with a small, dedicated panel of reviewers (European Commission, 2025; WCO, n.d.-a).

Operational Roles

To ensure the successful deployment of this project, it is essential to define, for clarity purposes, several key roles. A curator will be responsible for index maintenance, a reviewer will make decisions regarding HIL, an approver will provide the final sign-off, and a

release manager will oversee versioning and change notes. Additionally, it is important to establish a confidence threshold, denoted by τ, and formalize service-level agreements for the review process.

Change Management

To ensure the system remains current with the latest updates from WCO DM and EUCDM, it is essential to subscribe to their release notes. Upon the release of new updates, the system should automatically initiate a schema-differencing routine. This routine will facilitate the identification of any changes and populate a remap queue for any affected mappings. For each set of updated mappings, auditable deltas and release notes should be issued. Additionally, it is crucial to integrate the system's logs with the existing risk management and AEO oversight processes (WCO, 2021).

Integration Points

The pipeline must be capable of exporting the mapping specification in a format that is both neutral and machine-readable. Additionally, it should offer APIs for Single Window components to access and verify the mappings and their origins. It is crucial to distinctly separate the responsibilities of the model maintainers from those of the business owners of the mappings (WCO, 2011).

Legal and ethical alignment

- Logging and traceability; maintain technical documentation and event logs to satisfy AI Act transparency provisions for higher-risk public-sector systems (European Commission, 2024);
- Human oversight; route below-threshold cases to reviewers, record decisions, and allow appeal pathways;
- Data minimization; restrict retrieval sources to public, authoritative content unless specific access is authorized; and
- Explainability; provide short rationales so decisions can be understood during audits.

Leadership Mindset: The Future Customs Executive

In the rapidly evolving world, a contemporary leader in customs administration must masterfully balance the dual imperatives of protection and facilitation. This is not merely a pragmatic approach; it is a visionary strategy that demands the embrace of change, the mastery of cutting-edge tools, and the formation of dynamic teams that seamlessly integrate proficiency in standards with unparalleled expertise in data engineering. It is absolutely essential to maintain meticulous logs, conduct thorough reviews, and rigorously measure key performance indicators. By meticulously managing foundational elements such as definitions, constraints, and approvals, we can significantly enhance efficiency, safety, and trust. This is not just an option; it is a necessity for any forward-thinking customs administration committed to excellence and innovation.

Culture matters

It is imperative to prioritize the refinement of definitions as new features are introduced. This practice not only enhances clarity but also ensures that innovations are seamlessly integrated into existing frameworks. Furthermore, it is crucial to recognize and reward teams that deliver clear and well-reasoned justifications, as their efforts drive progress and innovation. Additionally, mapping logs should be viewed as invaluable institutional memory, capturing the essence of organizational knowledge and experience, rather than being dismissed as mere bureaucratic processes. By adopting these strategies, organizations can foster a culture of excellence and continuous improvement.

Impact - Before and After Adopting a Transformer-Based Pipeline

Transitioning from traditional, predominantly manual mapping to the proposed method represents a significant advancement in accuracy, efficiency, and governance. This innovative approach substantially improves alignment timelines, streamlines onboarding processes, and strengthens code-list management. It offers exceptional resilience against drift and facilitates seamless inter-agency collaboration, all while minimizing the risk of errors. Unlike preliminary prototype tests, this comparison is based on comprehensive public documentation, highlighting the method's proven effectiveness and reliability. Embrace this evolution to experience a new era of precision and collaboration.

Table 6

Before/After Comparison (Qualitative)

Dimension	Before (manual, rules-first)	After (transformer + constraints + HIL)
Mapping accuracy	Inconsistent on short/ambiguous labels; multilingual issues common	Better candidate recall via embeddings; cross-encoder disambiguation improves precision
Time to align	Repeated document hunting; long analyst iterations	Faster shortlists; reviewers focus on edge cases; audit logs generated automatically
Onboarding	High barrier; relies on a few experts	Lower barrier: explanations + guardrails let juniors contribute with oversight
Governance	Sparse documentation; decisions scattered in emails/files	End-to-end logs; approvals below τ; exportable reports
Code-list control	Easy to miss; errors discovered downstream	Hard validation at the gate; explicit bindings to official lists
Version drift	Conversions repeated; misalignment debt grows	Schema-diff → remap queue → tracked releases; less rework per upgrade

Dimension	Before (manual, rules-first)	After (transformer + constraints + HIL)
Inter-agency handshake	Fragile joins; partner misunderstandings	Clearer semantics; fewer broken joins; stronger audit trail
Risk of GIGO	High—semantic ambiguity passes through	Lower—rules block unsafe approvals; human review for tricky items

Narrative impact. The new method does not replace human expertise; instead, it concentrates human expertise where judgment is needed and automates repetitive steps. It also creates a durable institutional memory through logs and mapping specifications, which makes future upgrades easier to manage. Impact dimensions are derived from process efficiency objectives articulated in Single Window implementation guidance (WCO, 2011).

Two long standing principles (for whatever AI Projects)

Code-List Management Playbook

Code lists are fundamental components for system interoperability. It is crucial to establish them early, verify accuracy regularly, and avoid ad hoc modifications. When managing fields governed by code lists, follow this procedure: document the official list and link to the specific version in use. Reject non-conforming values with an error message indicating expected values. Monitor rejected values to prioritize rectification. Where local practices use alternative terms, like "Euro" instead of "EUR," implement a pre-validation dictionary to convert data to the official code before final validations.

Data Quality and GIGO Prevention

The adage "Garbage in" and "Garbage out" underscores a significant risk that necessitates careful management. To address this issue, the strategy can be divided into three primary components. Firstly, source hygiene involves the implementation of clear forms and proper validation processes. Secondly, mapping hygiene entails the establishment of constraints and thresholds. Lastly, monitoring is facilitated through dashboards that track rejection rates and common errors. Research from the health and e-government sectors supports this approach, demonstrating that poor data quality can infiltrate models and adversely affect decision-making (Kilkenny & Robinson, 2018; Cai & Zhu, 2015).

Personal Perspective: DMPT Exposure and Doctoral Study

Engaging in discussions about DMPT work is not just an exercise in dialogue; it is a pivotal opportunity to drive transformative change. The journey from initial proposals to authoritative publications is a testament to the power of consensus among stakeholders. By actively participating in this dynamic process, we are not just facilitating change; we are ensuring that every voice is heard and valued. This is a clarion call to action for those committed to progress and innovation. Let us seize this moment to champion meaningful change and propel our collective vision forward.

Coupled with my studies in Emerging Technology at Golden Gate University, the lesson is clear: data-driven transformation is effective when data management is taken seriously, reuse is strategically planned through models and interoperability protocols, and governance is centered on human needs. AI proves valuable when it concentrates its efforts where they yield results, rather than attempting to supplant human judgment.

Limitations and Future Work

The present study intentionally follows a **design-science ex-ante** evaluation paradigm (Hevner et al., 2004; Gregor & Hevner, 2013; Venable et al., 2016). Consequently, the absence of quantitative empirical testing on real national schemas, field deployment of a full prototype, or measured performance metrics on operational data represents a deliberate methodological boundary rather than an omission.

Everything needed to build and test the pipeline is already public and free. The tools it uses are two very popular, ready-made transformer models that anyone can download today. With the detailed description given in Section 5, the step-by-step examples, and the traceability table, any customs IT team or researcher can have everything they need to get the whole system running in an afternoon – literally in fewer than 50 lines of ordinary Python code, using only the openly available WCO Data Model and EUCDM files.

The real-world testing phase is already planned and under preparation. The next study (which is already being organized) will take the recent EUCDM upgrade from version 6.0 to 7.0 (382 changed elements, released 7 April 2025) as the test case, use properly validated mappings from three cooperating EU countries, and measure all the usual performance numbers: specifically, how often the right answer is in the top five suggestions (precision@5), overall ranking quality (mean reciprocal rank), how few cases actually need a human to look at them, and how much time analysts save compared with doing everything by hand today (we expect well over 70 % time saving). After that, proper field trials with customs administrations that have different budgets and IT capabilities (rich countries, middle-income countries, and smaller agencies) will be run during 2026–2027.

Conclusion

The WCO Data Model remains central to the vision of harmonized regulatory data exchange, yet the persistent challenges of semantic heterogeneity, version drift, and auditability demand a more sophisticated approach. A single, **transformer-based, human-governed pipeline** is the definitive solution to these challenges. The retrieval and reranking stages significantly bolster the robustness of semantic matching, while constraint checking rigorously ensures compliance with technical rules. Crucially, a mandatory human oversight loop, coupled with comprehensive logging, unequivocally satisfies the stringent demands of public-sector governance. By presenting a meticulously detailed set of requirements, a conceptual architecture, a conformance mapping, a risk–control register, and illustrative scenarios, this paper offers an indispensable, reproducible blueprint that customs administrations can seamlessly adapt and expand upon. This design empowers administrations to responsibly and efficiently evaluate and deploy interoperable mapping support, even as legal frameworks and data models continue to evolve (European Commission, 2024, 2025; WCO, n.d.-a).

References

Azqueta-Gavaldón, A., & Cosgrove, J. R. (2025). Beyond traditional algorithms: Leveraging LLMs for accurate cross-border entity identification. *arXiv*. https://doi.org/10.48550/arXiv.2507.11086 arXiv+1

Babaei Giglou, H., Myachykov, A., Shaban-Nejad, A., & Lofi, C. (2024). LLMs4OM: Matching ontologies with large language models. In *The Semantic Web – ISWC 2024* (pp. 27–40). Springer. https://doi.org/10.1007/978-3-031-78952-6_3

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., … Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv*. https://arxiv.org/abs/2108.07258

Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. *Data Science Journal, 14*, 2. https://doi.org/10.5334/dsj-2015-002

Cho, H. N., Kim, H., & Kim, S. (2024). Task-specific transformer-based language models in health informatics: A review. *JMIR Medical Informatics, 12*, e49724. https://medinform.jmir.org/2024/1/e49724/

Di Marzo Serugendo, G., Cappelli, M. A., Falquet, G., Métral, C., Wade, A., Ghadfi, S., Cutting-Decelle, A.-F., Caselli, A., & Cutting, G. (2024). Streamlining tax and administrative document management with AI-powered intelligent document management system. *Information, 15*(8), 461. https://doi.org/10.3390/info15080461 MDPI

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., … Houlsby, N. (2021). An image is worth 16×16 words: Transformers for image recognition at scale. *International Conference on Learning Representations (ICLR)*. https://arxiv.org/abs/2010.11929

European Commission. (2024). *Regulation (EU) 2024/1689 … (Artificial Intelligence Act)*. EUR-Lex. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

European Commission. (2025, April 7). *EU Customs Data Model (EUCDM) 7.0 release: Enhancing customs data efficiency across the EU*. https://taxation-customs.ec.europa.eu/news/eucdm-70-release-enhancing-customs-data-efficiency-across-eu-2025-04-07_en Taxation and Customs Union

European Commission. (n.d.). *EU Customs Data Model (EUCDM)*. https://taxation-customs.ec.europa.eu/online-services/online-services-and-databases-customs/eu-customs-data-model-eucdm_en

Herzig, J., Nowak, P., Müller, T., Piccinno, F., & Eisenschlos, J. (2020). TAPAS: Weakly supervised table parsing via pre-training. In *Proceedings of ACL 2020*. https://aclanthology.org/2020.acl-main.398/

HL7 International. (n.d.). *HL7 Version 2 to FHIR Implementation Guide (v1.0.0, continuous build)*. https://build.fhir.org/ig/HL7/v2-to-fhir/ build.fhir.org+1

Huang, X., Khetan, A., Cvitkovic, M., & Karnin, Z. (2020). TabTransformer: Tabular data modeling using contextual embeddings. *NeurIPS Datasets and Benchmarks*. https://arxiv.org/abs/2012.06678

International Organization for Standardization (ISO). (2005). *ISO 7372:2005—Trade data elements directory (TDED)*. https://www.iso.org/standard/35602.html

International Organization for Standardization (ISO). (2011). *ISO 11354-1: Framework for enterprise interoperability—Part 1: Overview*. https://www.iso.org/standard/50417.html

Jaegle, A., Borgeaud, S., Alayrac, J.-B., Doersch, C., Ionescu, C., Ding, D., … Carreira, J. (2021). Perceiver IO: A general architecture for structured inputs & outputs. In *Proceedings of ICML 2021*. https://arxiv.org/abs/2107.14795

Kilkenny, M. F., & Robinson, K. M. (2018). Data quality: "Garbage in – garbage out." *Health Information Management Journal, 47*(3). https://doi.org/10.1177/1833358318774357

Li, Y., Wang, H., Yerebakan, H. Z., Shinagawa, Y., & Luo, Y. (2024). FHIR-GPT enhances health interoperability with large language models. *NEJM AI, 1*(8), AIcs2300301. https://ai.nejm.org/doi/10.1056/AIcs2300301

Nerella, S., et al. (2024). Transformers and large language models in healthcare: A survey. *Artificial Intelligence in Medicine, 150*, 102862. https://doi.org/10.1016/j.artmed.2024.102862

Nogueira, R., & Cho, K. (2019). Passage re-ranking with BERT. *arXiv*. https://arxiv.org/abs/1901.04085

Oukes, P., Lemmen, C., & Folmer, E. (2024). Implementing data exchange and interoperability on LADM using the ISO FEI. *Land Use Policy, 146*, 107076. https://doi.org/10.1016/j.landusepol.2024.107076

Parciak, M., Vandevoort, B., Neven, F., Peeters, L. M., & Vansummeren, S. (2024). Schema matching with large language models: An experimental study. *arXiv*. https://arxiv.org/abs/2407.11852

Reed, S., Zolna, K., Parisotto, E., … de Freitas, N. (2022). A generalist agent. *arXiv*. https://arxiv.org/abs/2205.06175

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 3982–3992). Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1410

Seedat, N., & van der Schaar, M. (2024). Matchmaker: Self-improving large language model programs for schema matching. *arXiv*. https://arxiv.org/abs/2410.24105

Shubham, Arya, A., Roy, S., & Jonnala, S. (2023). An ensemble-based approach for assigning text to correct Harmonized system code. In *2023 International Conference on Artificial Intelligence and Smart Communication (AISC 2023)* (pp. 35–41). IEEE. https://doi.org/10.1109/AISC56616.2023.10085512

Taboada, M., Martinez, D., Arideh, M., & Mosquera, R. (2025). Ontology matching with large language models and prioritized depth-first search. *arXiv*. https://arxiv.org/abs/2501.11441

UKPLab. (2024). *MS MARCO models — Sentence Transformers documentation*. https://www.sbert.net/docs/pretrained-models/msmarco-v5.html

UNESCAP/WCO. (2021, November 17). *WCO Data Model: Briefing* [slides]. https://www.unescap.org/sites/default/d8files/event-documents/WCO%20DM%20Briefing.pdf

United Nations Economic Commission for Europe (UNECE). (2009). *Core Components Technical Specification (CCTS), Version 3.0*. https://unece.org/fileadmin/DAM/cefact/codesfortrade/CCTS/CCTS-Version3.pdf

United Nations Economic Commission for Europe (UNECE). (n.d.). *UNTDED / ISO 7372*. https://unece.org/untded-iso7372

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems (NeurIPS 2017)*. https://arxiv.org/abs/1706.03762

Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A framework for evaluation in design science research. *European Journal of Information Systems, 25*(1), 77–89. https://doi.org/10.1057/ejis.2014.36

Wang, S., Li, Y., Xiao, H., Dai, B. T., Lee, R. K.-W., Dong, Y., & Deng, L. (2025). LLMATCH: A unified schema matching framework with large language models. *arXiv*. https://arxiv.org/abs/2507.10897

World Customs Organization (WCO). (2004/2018). *Unique Consignment Reference (UCR) guidelines*. https://www.wcoomd.org/-/media/wco/public/global/pdf/topics/facilitation/instruments-and-tools/tools/ucr/ucr_new_e.pdf

World Customs Organization (WCO). (2011). *Compendium: How to build a Single Window environment*. https://www.wcoomd.org/en/topics/facilitation/instrument-and-tools/tools/single-window-guidelines.aspx

World Customs Organization (WCO). (2017/2023). *Single Window compendium, Vol. 2, Part V: Data harmonization*. https://www.wcoomd.org/-/media/wco/public/global/pdf/topics/facilitation/instruments-and-tools/tools/single-window/compendium/swcompendiumvol2partv.pdf

World Customs Organization (WCO). (2021). *SAFE framework of standards to secure and facilitate global trade* (2021 ed.). https://www.wcoomd.org/-/media/wco/public/global/pdf/topics/facilitation/instruments-and-tools/tools/safe-package/safe-framework-of-standards.pdf

World Customs Organization (WCO). (2024). *Terms of reference for the Data Model Projects Team (DMPT)*. https://www.wcoomd.org/-/media/wco/public/global/pdf/about-us/wco-working-bodies/facilitation/terms-of-reference-for-the-dmpt.pdf

World Customs Organization (WCO). (n.d.-a). *WCO Data Model eHandbook*. https://www.wcoomd.org/en/topics/facilitation/instrument-and-tools/tools/data-model/ehandbook.aspx

World Customs Organization (WCO). (n.d.-b). *WCO Data Model App*. https://datamodel.wcoomd.org/

World Customs Organization (WCO). (n.d.-c). *Unique Consignment Reference (UCR) — overview*. https://www.wcoomd.org/en/topics/facilitation/instrument-and-tools/tools/ucr.aspx