✉ info@databoiler.com https://www.databoiler.com

Evolution from ISIN to UPI+ and PKI / DPKI

The use of derivatives identifiers is heavily integrated across global financial hubs, with regulations dictating precisely which identifiers must be used depending on whether the trade happens **Over-the-Counter (OTC)** or on a **Trading Venue (Exchange-Traded)**. In jurisdictions like the European Union (ESMA), regulators use the OTC ISIN alongside the UPI. ISIN is incredibly effective for venue transparency and investor protection because it identifies the exact instrument being traded. However, it is overly granular for many OTC derivatives. Minor lifecycle events or custom dates generate entirely new ISINs, causing **data explosion** and making global macroeconomic risk aggregation highly difficult.

UPI is ideal for cross-border data aggregation. It balances anonymity with product traits, creating a global standard for tracking market-wide concentration risks across different countries. The downside of UPI is the lack of granular terms (such as expiry dates or strike prices) required by regulators for detecting specific insider trading or market manipulation. Investment firms operating in both the US and Europe often use data layers function to solve structural problems – the UPI tracks systemic liquidity, the ISIN handles market abuse forensic, then leverage the Legal Entity Identifiers (LEIs) for counterparty risk tracking to isolate credit contagion, and apply filter via Classification of Financial Instruments (CFI) to separate for example all *"FX Forwards"* from option or interest rate portfolios.

When the DSB issues a UPI for an Interest Rate Swap, its embedded template fields focus solely on the structural *"product"* characteristics. Because it ignores exact dates, a spot-starting 5-year SOFR swap and a 1-year forward-starting 5-year SOFR swap share the exact same UPI code. To distinguish a spot swap from a forward swap, regulatory bodies require reporting entities to supplement the UPI code with auxiliary, transactional data fields. For derivatives transaction-level reporting, if the Effective Date is mathematically greater than the Execution Timestamp (plus standard settlement conventions), the trade repository flags it as a forward-starting contract.

Under European transparency rules (MiFIR/EMIR Refit), the ESMA operates the Financial Instruments Reference Data System (FIRDS). It links the UPI directly and when one querying FIRDS using the UPI, the Core UPI Library returns a wider table containing the distinct regulatory attributes, such as *"Term of Contract"* and *"Forward Starting Term."* Another way to achieve the function is by querying an OTC ISIN in the DSB reference data library to retrieve the *"Contract Term (Value/ Unit)"* as a standard data attribute (e.g., 5 Years). Despite those benefits, regulators chose not to adopt the approach because doing so causes the single UPI splits into thousands of fragments, making the UPI lose its primary function as a macro-level *"product aggregator"* – i.e., an Identifier Explosion duplicating ISIN.

ESMA and the DSB have evaluated the possibility to make contractual terms and forward-starting statuses directly available within the UPI Reference Data Library in the past. If these attributes were made available natively inside the UPI library, the benefits include: eliminates custom date logic, simplifies cross-broader messaging, and plugs data gaps in post-trade systems. If investment firms were to build dynamic, real-time data pipelines to stream these elements on the fly, the process involves:

- hit separate reference databases simultaneously upon a trade is executed,
- execute an algorithm to stitch the UPI attributes and localized ISIN elements together in a single message, and
- map them perfectly under strict time constraints (often T+1 or within minutes of execution);
- because this happens dynamically in milliseconds for every single transaction, operational risk is driven by system instability and network latency issues.

It is unjust to lay all the blame on investment firms for reporting delays or corrupted data streams lead to non-compliance penalties, regulatory audits, and severe financial fines. Thankfully, a compromise emerged where DSB championed the UPI-ISIN Wrapper. DSB acts as a translator where investment firms can input the regional code (CUSIP, SEDOL, FIGI) and the

system automatically find the correct UPI to ensure a 100% accurate, single-step lookup while maintaining licensing compliance. Yet, the data friction between UPI+ and ISIN is still ongoing. Exchange-Traded Derivatives (ETD) is predominantly 100% ISIN-only. Approximate 60% of OTC Swaps and FX is UPI-only for G20 markets, while the EU and UK represent the remaining 40% roughly in adopting a hybrid UPI+ISIN framework. In turn, the ESMA and the UK FCA allow firms to populate a single *"Instrument Identification Code"* field preferentially with a UPI where available, dropping back to an ISIN only when a contract is bound to an active trading venue.

Short-Term: Immediate Parallel Data Enrichment Reference Library

Semantic Inferencing, Handling Schema Drift and Variety

As trades mature, automated scripts frequently struggle to handle the rounding math of, for example *roll conventions*, which fragments tracking data. We encourage regulators, service bureau of reference data, clearinghouses, consolidated tape provider (CTP), and other market participants to explore the use of modern semantic databases for **short-term quick wins**.

For example, when inferring the contractual term (tenor) and forward-starting status from *Effective Date*, *Expiry Date*, and *Roll Convention*, Progress MarkLogic can handle the semantic data integration, while Semaphore manages the metadata intelligence and entity enrichment that grounds in advanced Retrieval-Augmented Generation (RAG) for high precision retrieval and contextual prompting. This unique toolset is best for handling Schema Drift and Variety. Depending on the users' requirements for institutional-grade pricing precision that accounts for every holiday shift across complex multi-city roll conventions or certain *"rounding"* is acceptable, this semantic tech-stacks can achieve significant efficient gain and cost savings over traditional structural database.

The Architectural Core: Dual-Track Execution (Direct Streaming vs. Semantic Inference)

To maximize resource deployment across the financial ecosystem, this business case rejects a *"one size fits all"* reporting framework. Instead, it implements a Dual-Track Data Architecture that handles ETDs and OTC contracts according to their native data characteristics.

Track 1: Direct Streaming for Exchange-Traded Derivatives (ETDs)

For listed futures and options, contractual terms, ex-dates, and strike prices are entirely standardized and set by the exchange at inception. ETD platforms natively bundle these terms inside the standard **Exchange ISIN**. Because this data is predefined by the venue, it can be streamed directly into transaction networks, consolidated tapes, and regulatory systems with **zero development cost** or upstream modification required by market participants. The streaming **Consolidated Tape** runs at maximum bandwidth efficiency. It avoids payload bloating because subscribers can cross-reference the streamed ETD ISIN against a local, static security master file rather than processing dynamic strings in real-time.

Track 2: Downstream Semantic Inference for Over-the-Counter (OTC) Derivatives

Unlike ETDs, OTC swaps are bespoke, bilaterally negotiated contracts. Forcing trading desks to explicitly append and stream a calculated Contractual Term field across multi-vendor networks (e.g., Murex, Calypso, Bloomberg) creates immense technical debt and crippling regression risks. The OTC stream is kept completely clean, passing only the standard UPI string and raw date fields (Effective Date, Expiry Date, Roll Convention).

Upon ingestion, a **Template Driven Extraction (TDE)** template in an **Automated Data Layer Processing (MarkLogic)** automatically executes deterministic date math to calculate a standard duration string (e.g., 5Y). **Metadata Intelligence Enriched (Semaphore): Progress Semaphore** reads the inferred term and applies financial rules to categorize the swap under standardized regulatory taxonomy concepts for **Metadata Intelligence Enrichment** (e.g., "Medium-Term Interest Rate Swap" or "Forward-Starting Swap" if Effective Date > Execution Timestamp). Users requiring institutional-grade pricing precision can always **click a record on-demand** to dynamically query the precise, non-approximate contract terms directly from the core reference data library.

Stand up the Progress MarkLogic / Semaphore caching library running the date-inference data pipeline for OTC contracts, while establishing the direct-ingest pipeline for streamed ETD ISIN data. Market participants query this library on-demand via a secure REST API. This instantly resolves the operational visibility gap across both UPI+ and ISIN regimes without requiring months of cross-platform code changes on active transaction streams.

To eases the political standoff that involves licensing fees, access limits, and vendor locking from specific regulatory data hubs) in the long-term, let's step aside from a data formatting argument for a moment, to consider an **access layer solution** using centralized or decentralized Public Key Infrastructure (PKI / DPKI).

Long-Term: A 3-Phase Cryptographic Identity Evolution (PKI to DPKI)

Decoupling the "Product" from the "Lifecycle" via Twin Certificates

Envisioning a system implementation that supports **two** distinct certificate structures instead of relying on DSB centralized registry to issue competing alphanumeric string templates:

- **Public-key certificates:** Certify the absolute core identity of an entity or product.
- **Attribute certificates:** Certify variable privileges, states, or contract settings linked to that identity.

The stable base product (the UPI definition) is cryptographically signed as the immutable, permanent public-key master record. The granular, region-specific terms that cause identifier explosion – like maturity adjustments, execution timestamps, or European venue compliance data – are appended as digitally signed Attribute Certificates.

Resolving the "Data Streaming versus Library Lookup" Conflict

In a DPKI system, when a trade is executed, the node pushes the trade certificate directly to a shared, replicated distributed ledger. Because DPKI makes the entire certificate history and metadata globally available locally at every node, a US firm reads the clean UPI layer instantly, while an EU firm's node extracts the cryptographically verified attribute blocks (the ISIN-equivalent data) in parallel. No streaming overhead is required; the data is bound natively to the cryptographic object.

Eliminating Centralized Licensing and Aggregation Friction

DPKI establishes trust by global mathematical consensus across financial networks, completely removing single points of failure or data monopoly. Regulators from the ESMA and the US CFTC/ SEC can audit the same immutable ledger independently. This allows the essential mathematically roll up risk at the root-hash level globally (UPI function), while regional regulators query down the signature chain to extract specific transaction-date forensics (ISIN function), all without paying for disparate, disjointed indexing services.

Cryptographic Proof of Complex Rules

Ideally, we would like to see a long-term migration of reference data management to DPKI, where a smart contract will act as an authoritative, immutable validator node. Instead of upstream systems passing varying string interpretations of a contract's term, the DPKI architecture runs a zero-knowledge or deterministic rule ledger. It signs off on the precise contractual tenor at inception, eliminating interpretation deltas between counterparties and global trade repositories. That being said, today's distributed ledger technology (DLT) consensus algorithms (the backbone of DPKI) introduce latency that has yet to match the speed of the DSB's existing centralized, automated cloud infrastructure. Forcing thousands of global bank nodes to reach consensus on a blockchain for every single swap creation would stall market plumbing.

A Practical Three-Phased Hybrid Approach

To transition smoothly without disrupting market operations, the DSB can implement a three-phased hybrid approach.

Phase 1: Implement Centralized PKI (The Identity Foundation)

The DSB remains a centralized utility but upgrades its generation output. Instead of returning a raw alphanumeric text string (e.g., UPI: A1B2C3), it operates as a Certificate Authority (CA). It issues a standard Public-Key Certificate representing the immutable UPI product. Firms pull this certificate using standard REST APIs. No decentralized ledger is required yet, but the identifier is now a cryptographically signed object, completely eliminating data spoofing or translation errors downstream. This Centralized PKI setup embeds data security natively into the trade object while preserving backward compatibility.

Phase 2: Deploy Attribute Certificates for Jurisdiction Mapping

The DSB utilizes Privilege Management Infrastructure (PMI) to issue separate, dynamic Attribute Certificates that plug into the core UPI master certificate. The European version attaches a separate attribute block mapping specific contract tenors and venue tags (ISIN style), while the US version of the certificate contains basic product attributes (UPI+ style). This solves the political standstill without causing *"identifier explosion."* The underlying database object is exactly the same, but different

regulatory lenses strip away or view specific cryptographic attributes based on local laws. This Solve the US versus EU debate by attaching separate, digitally signed Attribute Certificates to the master UPI. A US firm reads the clean UPI layer; an EU firm extracts the venue and expiration data (ISIN parity) from the exact same object.

Phase 3: Introduce DPKI for Central Bank and CCP Node Syncing

For the top-tier institutions that require heavy cross-border reconciliation – specifically Central Banks, Trade Repositories (TRs and SDRs), and Clearinghouses (CCPs) – the DSB establishes a federated DPKI network using a Zero-Trust, high-throughput hybrid architecture (such as Hyperledger Fabric with Zero-Knowledge Proofs). The DSB publishes **master root certificates** onto this closed network (i.e., the highest-level digital certificates that if a node possesses this root certificate, it can cryptographically verify that any other node or transaction on the network is legitimate). In turn, Central Banks can audit systemic risk instantly via their local nodes without checking with a centralized intermediary or running slow manual lookup queries. It maintains high performance for investment firms, while providing decentralized, real-time auditability to global macro-prudential regulators. By establishing a permissioned consortium blockchain node network across Central Banks, Trade Repositories, and Clearinghouses, regulators gain zero-latency, cross-border auditability without relying on an expensive, single-point-of-failure database lookup.

Strategic Justification and Ecosystem Resource Optimization

For Market Participants: Eliminates upstream refactoring costs across both trading classes. ETDs use existing exchange infrastructure, while OTC contracts leverage the downstream data layer to bypass payload alterations.

For Global Regulators: Central banks tracking macro systemic risk can query the clean, aggregated UPI data stream to monitor global exposures. Concurrently, market surveillance teams query on-demand inferred terms from the library to audit precise execution-to-expiry windows for market abuse detection.

Data Provider Relief: Prevents identifier explosion within the DSB. The UPI registry remains clean and stable, while the computing cluster handles highly volatile OTC contract terms in an elastic, on-demand query layer.

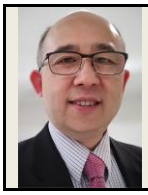
Other Remarks – the latest US regulatory development

The recent US CFTC's proposal on Event Contracts alternate Reporting Framework could be disruptive to the derivatives ecosystem – **prediction markets** will not generate or require a UPI or an OTC ISIN for these contracts. They will bypass standard Swap Data Repositories (SDRs) entirely, reporting directly to the Commission. NOTE: ESMA classify many event contracts as binary options or derivatives under MiFID II. Because retail binary options are broadly banned in the EU, platforms face an uphill battle trying to register as financial instruments. Also, nine European gambling regulatory authorities recently formed a joint international initiative to actively monitor, block, and penalize unlicensed prediction markets.

Also, there is a sweeping joint SEC-CFTC push to **harmonize**, modernize, and eliminate duplicative reporting between **security-based swaps** (SEC) and **standard swaps** (CFTC). The proposed unify data definitions would require to match the global Critical Data Elements (CDE) framework. While it sounds counterintuitive – since a joint rulemaking means both agencies will force payload rewrites anyway – the reality of a consolidated US reporting structure makes a downstream semantic data layer even more valuable. Pursuing semantic inference remains the superior strategic choice over streaming contractual terms.

In addition, the CFTC finalized an order to completely sunset **Large Trader Reporting requirements** under Part 20 for physical commodity swaps. This development would remove legacy, duplicative reporting structures that cost market participants millions annually. It validates the data consolidation thesis of this business case.

***



Data Boiler is a Pioneer in FinTech with patented inventions (US, Canada, Singapore, Japan, Australia, and 20 European countries) in signal processing, trade analytics, machine learning, time-lock cryptography, etc. We frequently comment on regulatory policy both domestically and abroad with over 12 years in business. A type C Member of the European Commission's Data Expert Group + former committee of BITS (Bank Policy Institute).