January 2026

CONVERGENCE: THE GREAT TOKENIZATION

HOW TWO DEFINITIONS OF TOKENIZATION ARE MERGING TO CREATE DATA MARKET INFRASTRUCTURE

AND HOW CARBON ARC BUILT A COMPOSABLE WORLD MODEL

Convergence: The Great Tokenization

How Two Definitions of Tokenization Are Merging to Create Data Market Infrastructure

And How Carbon Arc Built a Composable World Model[1]

Executive Summary

Within the next decade, a fundamental shift will occur in how humans and economic systems operate. As autonomous AI agents proliferate across enterprises, the digital economy is transitioning from 8 billion human actors to a hybrid model incorporating tens of billions of non-human identities—AI agents, service accounts, and automated systems capable of independent decision-making and related action—with the AI agents market alone projected to reach over $50 billion by 2030.[2][3][4]

This projection may sound provocative. For some, it may feel unsettling. But it reflects the convergence of independent forces that have been developing for decades, each following its own trajectory, now aligning in ways that make this agentic future not speculative but probabilistically inevitable.

At the center of this transformation lies a term that carries two distinct meanings: *tokenization*. Understanding the distinction—and the relationship—between tokenization as it relates to information and tokenization as it relates to value is essential to grasping the infrastructure requirements of an AI-driven economy.

The tokenization of information and the tokenization of digital assets have each developed along separate timelines, for separate purposes, as part of two revolutions. They are now converging. That convergence creates the foundation for something neither could achieve alone: a data market infrastructure.

The emergence of data market infrastructure—clearing, standardized instruments, real-time settlement, embedded compliance, separation of custody from access—mirrors the infrastructure that transformed financial markets from bilateral negotiation to scalable, institutional-grade systems. It is now being built for data.

Carbon Arc sits at the center of this buildout—performing the clearing function, producing structured instruments, enabling atomic settlement, embedding compliance, and separating custody from access. Through a composable world model exposed via an AI-native context layer, it transforms real-world activity into an entity-centric, continuously updated representation of state, behavior, and relationships that AI systems can reason over directly. As data increasingly functions as an asset class, Carbon Arc's operational infrastructure will hold structural advantages analogous to those held by early clearing and settlement infrastructure in financial markets.

---

1 *This white paper examines the convergence of AI language processing infrastructure and blockchain-based asset tokenization, and their implications for the emergence of data as an institutional asset class.*

2 World Economic Forum. (2025, September). "Unsecured AI agents expose businesses to new cyberthreats." *WEF Stories*. https://www.weforum.org/stories/2025/09/unsecured-ai-agents-cyberthreat/ (citing projection that "the volume of non-human and agentic identities is now expected to exceed 45 billion by the end of this year").

3 Okta Inc. (2025, April). *Securing non-human identities with Okta* [Datasheet]. https://www.okta.com/sites/default/files/2025-06/Securing-Non-Human-Identities-with-Okta.pdf (noting non-human identities "now outnumber humans 50 to 1").

4 Amazon Web Services. (2025, June 12). "The rise of autonomous agents: What enterprise leaders need to know about the next wave of AI." *AWS Insights*. https://aws.amazon.com/blogs/aws-insights/the-rise-of-autonomous-agents-what-enterprise-leaders-need-to-know-about-the-next-wave-of-ai/ (projecting "The AI agents market itself is expected to grow to $52.6 billion by 2030").

Part I: A Chronology of Two Tokenizations

The First Revolution: Tokenization of Information (1950s–Present)

The tokenization of language for machine processing traces back to the earliest days of computational linguistics.

1950s–1960s: Foundations. Early natural language processing systems at Georgetown began breaking text into discrete units for machine analysis.[5] These primitive tokenizers operated on simple rules—splitting text at whitespace and punctuation—but established the foundational concept: machines cannot process language directly; they must first decompose it into computable fragments.

1970s–1980s: Statistical Methods Emerge. Researchers at IBM and Bell Labs pioneered statistical approaches to language modeling. The insight that language could be modeled probabilistically—predicting the likelihood of word sequences based on observed frequencies—transformed tokenization from a preprocessing step into a core architectural decision. Different tokenization schemes produced dramatically different model performance.

1990s–2000s: Subword Tokenization. The limitations of word-level tokenization became apparent as systems encountered out-of-vocabulary terms. Byte Pair Encoding (BPE), originally developed for data compression, was adapted for natural language processing, enabling models to handle novel words by decomposing them into familiar subword units. A specialized term like "tokenization" could be processed as "token" + "ization"—fragments the system had encountered frequently in training data.

2017–Present: The Transformer Era. The introduction of transformer architectures fundamentally changed what tokenization enables. Modern large language models convert tokens into numerical representations called embeddings, which encode semantic properties. Through attention mechanisms, these models dynamically compute relationships between tokens—examining how each fragment relates to others across the entire context window. The result: systems that can generate remarkably coherent text by predicting one token at a time.

The critical insight remains unchanged across seven decades: AI systems do not comprehend language—they compute numbers. The tokenizer is the bridge between human expression and machine computation.

The Second Revolution: Tokenization of Value (2008–Present)

The tokenization of assets on digital infrastructure followed a separate developmental arc, driven by different communities pursuing different objectives.

2008–2013: Cryptographic Foundations. The introduction of Bitcoin demonstrated that value could be represented and transferred through cryptographic tokens without trusted intermediaries. While initially dismissed as a curiosity, the underlying innovation—immutable, verifiable ownership records maintained through distributed consensus—represented a fundamental advance in how ownership could be structured.

2014–2017: Programmable Assets. Ethereum and subsequent platforms introduced smart contracts—self-executing code that automates transfers, compliance checks, and settlement rules. This expanded tokenization beyond simple value transfer to programmable financial instruments. Assets could carry their own rules: who could hold them, under what conditions they could transfer, what rights attached to ownership.

Institutional Infrastructure. Traditional financial institutions began building production-grade tokenization infrastructure. The Canton Network, supported by Goldman Sachs, HSBC, BNP Paribas, Deutsche Börse,

[5] Hutchins, W. John. (2004). "The Georgetown-IBM experiment demonstrated in January 1954." *Proceedings of the 6th Conference of the Association for Machine Translation in the Americas*, 102–114. https://aclanthology.org/2004.amta-papers.12/ (describing the first public demonstration of machine translation, a collaboration between IBM and Georgetown University that translated Russian sentences into English using an IBM 701 mainframe).

Depository Trust & Clearing Corporation ("DTCC"), Citadel, and Bank of America, emerged to manage institutional-grade tokenized assets. Strategic investors including Liberty City Ventures began deploying capital specifically into the tokenization infrastructure layer, recognizing the structural opportunity in data and asset digitization. Broadridge's Distributed Ledger Repo platform began processing repurchase agreements, scaling to over $280 billion in daily volume by 2025.[6] JPMorgan's Onyx platform launched tokenized collateral settlements.[7] BlackRock introduced BUIDL, a tokenized money market fund, growing to over $2 billion in assets.[8]

2022–Present: Stablecoin Maturation. Regulated stablecoins achieved institutional credibility. Paxos issues USDP and PayPal's PYUSD through established regulatory frameworks. Circle's USDC operates across multiple blockchains with regular attestations. The Global Dollar Network, developed in partnership with the Monetary Authority of Singapore and over 70 global institutions, enables seamless USD-pegged transfers without traditional correspondent banking relationships. Franklin Templeton launched the first tokenized money market fund on a public blockchain, while WisdomTree introduced tokenized gold instruments.

The Current State. The Canton Network currently manages over $6 trillion in tokenized assets spanning repurchase agreements, fund shares, and other instruments, processing approximately $280 billion in daily repurchase agreements through atomic swaps.[9] DTCC operates as both network validator and clearing agent, aligning modern settlement infrastructure with existing regulatory frameworks.[10] It provides regulators with familiar oversight mechanisms, complete audit trails for SEC and CFTC compliance, and legal certainty rooted in established clearing rules. Tokenized treasuries now exceed $7 billion in total value locked across platforms including BlackRock BUIDL, Franklin Templeton, Ondo Finance, Maple, and Centrifuge.[11]

The Convergence Point

These two revolutions—developing independently for decades—are now converging around a common requirement: infrastructure that can price, clear, and settle data transactions at machine speed.

AI systems tokenize information to process it computationally. Financial systems tokenize assets to transfer them instantly. The data economy requires both: intelligence must be structured into computable units *and* those units must clear and settle economically at the moment of consumption.

---

[6] Broadridge Financial Solutions. (2025, September 10). "$280 Billion in Average Daily Processed Trade Volumes on Broadridge Distributed Ledger Repo Platform" [Press Release]. https://www.broadridge.com/press-release/2025/billions-in-average-daily-processed-trade-volumes-on-broadridge-dlt-repo-platform.

[7] J.P. Morgan. (2023, October 11). "JPMorgan's Tokenized Collateral Network (TCN) facilitates collateral settlement for a live-client OTC derivative transaction" [Press Release]; see also CoinDesk. (2023, October 11). "JPMorgan Debuts Tokenized BlackRock Shares as Collateral with Barclays." https://www.coindesk.com/business/2023/10/11/jpmorgan-debuts-tokenized-blackrock-shares-as-collateral-with-barclays (reporting first live blockchain-based collateral settlement on JPMorgan's Onyx platform, with BlackRock tokenizing money market fund shares transferred to Barclays).

[8] CoinDesk. (2025, December 30). "BlackRock's BUIDL hits $100M in dividends and passes $2 billion in assets." https://www.coindesk.com/markets/2025/12/30/blackrock-s-buidl-hits-usd100m-in-dividends-and-passes-usd2b-in-assets.

[9] Canton Network and Chainlink. (2025, September 24). "Canton Network and Chainlink Enter Into Strategic Partnership To Accelerate Institutional Blockchain Adoption" [Press Release]. https://www.canton.network/canton-network-press-releases/canton-network-and-chainlink-enter-into-strategic-partnership-to-accelerate-institutional-blockchain-adoption- (noting "The Canton Network supports over $6T in on-chain real-world assets, processing $280bn in repos daily").

[10] DTCC. (2025, December 17). "DTCC and Digital Asset Partner to Tokenize DTC-Custodied U.S. Treasury Securities on the Canton Network" [Press Release]. https://www.dtcc.com/news/2025/december/17/dtcc-and-digital-asset-partner-to-tokenize-dtc-custodied-us-treasury-securities.

[11] Yellow.com Research. (2025). "Tokenized U.S. Treasuries Hit $7.3B in 2025: Complete Guide to Digital Treasury Bonds." https://yellow.com/en-US/research/tokenized-us-treasuries-hit-dollar73b-in-2025-complete-guide-to-digital-treasury-bonds; see also INX. (2025, October 30). "Tokenized Treasuries: The Safest Way to Earn Yield On-Chain in 2025" (noting market "exploded from under $100 million two years ago to over $8 billion in October 2025"). https://www.inx.co/tokenized-treasuries-the-safest-way-to-earn-yield-on-chain-in-2025/.

This convergence creates the foundation for data as a true asset class—not data-as-product sold through enterprise sales cycles, but data-as-instrument that can be priced, traded, and settled with the fluidity of financial securities. Carbon Arc has built the payment processing infrastructure that makes this possible.

Part II: The Economics of Current AI Infrastructure and the Compute Constraint

The Resource Reality

Every AI query—whether requesting a recipe, summarizing a document, or debugging code—requires the model to compute attention across all tokens in the context window and generate responses token-by-token, with each generation step involving billions of mathematical operations across the model's parameters. This computational demand explains why data center infrastructure has become a critical bottleneck, with developers exploring unconventional solutions including space-based computing to address terrestrial limitations in power, cooling, and regulatory constraints.[12]

The numbers are stark. Training frontier models costs hundreds of millions to billions of dollars.[13] Inference costs compound the problem: the system cannot selectively engage relevant portions of its learned patterns—it must process every request through its entire architecture, a computational inefficiency equivalent to "boiling the ocean."

Models have finite context windows—limits on how many tokens they can process simultaneously. This constraint directly impacts both functionality and cost structure. Token count determines message length, response capacity, and computational expense.

The Structural Disadvantage of Optimization-Dependent Business Models

A generation of AI platforms has emerged promising to revolutionize financial and enterprise workflows. Hebbia provides document analysis at scale. AlphaSense delivers market intelligence through natural language search. Perplexity offers AI-powered search synthesis. These platforms demonstrate real capabilities within their architectural paradigm.

However, the limitation is structural, not technical. These platforms apply AI capabilities on top of existing data architectures—optimizing search across documents that remain fundamentally unstructured. They accelerate workflows that should not exist in an AI-native world.

The critical insight: companies whose business models depend on the work needing to be optimized face a structural disadvantage. Their value proposition is reducing friction in processes that better infrastructure would eliminate entirely.

Consider the architecture: when a document analysis platform processes a contract, it must ingest the full text, tokenize it, run it through attention mechanisms, and generate summaries or answers. The platform's value comes from making this expensive computation more efficient. But the underlying cost structure—compute-

[12] Stanford Institute for Human-Centered Artificial Intelligence. (2024). *AI Index Report 2024*. https://hai.stanford.edu/news/inside-new-ai-index-expensive-new-models-targeted-investments-and-more (estimating OpenAI's GPT-4 training cost at $78 million and Google's Gemini Ultra at $191 million in compute alone); see also Epoch AI. (2025, January 13). "How much does it cost to train frontier AI models?" https://epoch.ai/blog/how-much-does-it-cost-to-train-frontier-ai-models (projecting largest training runs will exceed $1 billion by 2027).

[13] Starcloud (formerly Lumen Orbit). (2024, September). *White Paper: Why we should train AI in space*. https://lumenorbit.github.io/wp.pdf; see also GeekWire. (2024, December 12). "Lumen Orbit, a Seattle-area startup that wants to put data centers in space, raises $11M." https://www.geekwire.com/2024/lumen-orbit-a-seattle-area-startup-that-wants-to-put-data-centers-in-space-raises-11m/ (reporting on Y Combinator-backed ventures developing orbital data centers to address AI energy demands); IBM. (2024, December). "Are data centers in space the future of cloud storage?" https://www.ibm.com/think/news/data-centers-space (describing multiple companies pursuing space-based computing infrastructure).

intensive processing of unstructured text—remains unchanged. Margins compress as the underlying models improve and compute costs decline.

Three Structural Failures of Current AI Platforms:

Compliance as overlay, not infrastructure. Every novel use case requires review. Every data source requires separate licensing negotiation. Every output requires human attestation. The AI accelerates analysis; the compliance bottleneck remains. For machine-speed operations serving autonomous agents, this is architecturally fatal.

Economic misalignment. Flat subscription pricing—the dominant model—disconnects value delivered from compensation received. When AI agents query thousands of times per hour, subscription economics become incoherent. More fundamentally, subscriptions cannot support the micropayment flows that data-as-asset-class requires.

Human-centric architecture. These platforms were built for analysts who read, evaluate, and wait seconds for responses. Autonomous agents require structured data at machine speed: verified, entity-centric, compliance-cleared, and economically settled at consumption. Current platforms cannot pivot—their architectures would require complete reconstruction.

The Data Access Problem Compounds Compute Constraints

Standard LLMs rely entirely on patterns encoded during training from a static corpus captured at a point in time, drawn predominantly from public sources. Without retrieval-augmented generation or similar architectures that can query external data sources, the precision that would come from structured, verified, real-time data remains largely inaccessible.

The inefficiency compounds: the vast majority of valuable data—proprietary data assets held by enterprises, institutions, and individuals—remains locked away. The International Data Corporation estimates that 90% of unstructured data remains unanalyzed[14][15]. This 90% remains siloed not because owners are unaware of its value, but because no infrastructure exists to extract that value safely while preserving ownership rights and ensuring compliance.

The industry's pivot toward synthetic data generation reflects this reality—when real-world transactional data proves too difficult to access and structure, simulating plausible alternatives becomes the path of least resistance. But simulation, however sophisticated, cannot substitute for observation.

The combination of compute constraints and data access limitations creates a structural ceiling on current AI capabilities. Breaking through requires not better optimization of existing architectures, but fundamentally different infrastructure—infrastructure that structures data for efficient retrieval rather than exhaustive computation, and that embeds compliance and economic settlement at the transaction layer.

Part III: Real-Time Transaction Data and the Organizational World Model

Beyond Static Corpora: The Case for Dynamic Intelligence

Current AI systems operate on a fundamental limitation: they learn from static snapshots. Training corpora are frozen at points in time. The models cannot perceive the living, dynamic world they are meant to serve.

---

[14] United States Data Science Institute. (2025, May 3). *Huge potential of dark data – Revealing untapped opportunities*. USDSI Data Science Insights. https://www.usdsi.org/data-science-insights/huge-potential-of-dark-data-revealing-untapped-opportunities.

[15] Box, Inc. & IDC. (2023). "Untapped Value: What Every Executive Needs to Know About Unstructured Data" [White Paper]. https://www.box.com/resources/unstructured-data-paper (noting "90% of your business data" is unstructured according to IDC); see also Hitachi Vantara, "Bringing Dark Data to Light," https://www.hitachivantara.com/en-us/blog/bringing-dark-data-to-light (citing IDC finding that "about 90% of unstructured data is never analyzed").

When a company changes leadership, when a fund adjusts its portfolio, when a regulatory filing reveals new ownership structure—the model remains ignorant until retrained.

This creates a structural gap between AI capability and organizational reality. The most valuable intelligence about organizations—their current state, recent behavior, evolving relationships—exists in real-time transaction data that no training corpus captures.

The Synthetic Data Problem

The AI industry has increasingly turned to synthetic data to address training data limitations.[16] Models generate training examples for other models. The approach has merit for certain applications—augmenting sparse datasets, creating privacy-preserving alternatives to sensitive data, generating edge cases for testing.

But synthetic data carries an inherent limitation: it can only recombine patterns present in its source data. Synthetic data cannot introduce new facts, nor can it capture organizational reality that the underlying models never observed. A model trained on synthetic financial data cannot learn that a particular company restructured last week, that a fund manager changed strategy after a personnel departure, that a regulatory filing revealed previously undisclosed ownership.

The appeal of synthetic and simulated data is understandable—it sidesteps the formidable challenges of acquiring, structuring, and maintaining rights to real-world transactional data at scale. These challenges are not merely technical; they require building relationships with data providers, establishing compliance frameworks, and creating infrastructure that preserves ownership while enabling access. When the alternative is grappling with these institutional complexities, generating data computationally becomes an attractive path of least resistance.

For organizational intelligence—understanding entities as they actually exist and behave—synthetic data is structurally insufficient. What matters is not statistically plausible patterns but observed reality: actual transactions, actual filings, actual behavioral signals from actual organizations.

Carbon Arc's Approach: Real-Time Transaction Data for Organizational World Modeling

Carbon Arc is constructing what current AI architectures cannot: a continuously updated, entity-centric representation of organizational reality—a world model built from actual transactional data, behavioral signals, and relationship structures rather than inferred from text or generated synthetically.

This world model is not a database. It is a living representation of state, behavior, and relationships that evolves as the organizations it represents evolve. The model ingests real-time signals: SEC filings as they publish, corporate announcements as they release, transaction data as it clears, relationship changes as they occur.

The architectural distinction is fundamental:

Traditional data platforms provide search across documents—natural language queries against unstructured text, returning ranked results that humans must interpret. The platform optimizes retrieval; the human provides understanding.

Carbon Arc provides a queryable model of organizational reality itself—structured through unified ontology, governed by embedded compliance, and economically accessible through consumption-based pricing. AI systems can reason over grounded truth rather than probabilistic inference from unstructured text.

---

[16] Gartner. (2023, August 1). "Gartner Identifies Top Trends Shaping the Future of Data Science and Machine Learning." https://www.gartner.com/en/newsroom/press-releases/2023-08-01-gartner-identifies-top-trends-shaping-future-of-data-science-and-machine-learning (predicting 60% of AI training data will be synthetic by 2024, up from 1% in 2021); Shumailov, I. et al. (2023). "The curse of recursion: Training on generated data makes models forget." *arXiv preprint arXiv:2305.17493* (documenting "model collapse" risks when AI systems are trained iteratively on AI-generated data).

Entity Resolution as Foundation. The world model begins with entity resolution: mapping disparate data sources to verified organizational identities. A company appears differently in SEC filings, news articles, transaction records, and corporate registries. Carbon Arc's entity resolution infrastructure unifies these representations into coherent organizational identities that persist across data sources and time.

Behavioral Signals, Not Just Static Attributes. Traditional organizational data captures attributes: headquarters location, officer names, filing history. Carbon Arc's world model captures behavior: how organizations act, how their actions change over time, how they respond to market conditions and regulatory requirements. Behavioral patterns reveal organizational reality that attribute databases cannot capture.

Relationship Mapping. Organizations exist within webs of relationships: ownership structures, counterparty networks, supply chains, competitive dynamics. Carbon Arc's world model represents these relationships as first-class objects, continuously updated as relationships form, evolve, and dissolve.

The Query Paradigm Shift

When an AI system needs to understand beneficial ownership, evaluate counterparty risk, or map organizational relationships through traditional platforms, it must search documents—retrieve relevant filings, parse unstructured text, infer relationships from linguistic descriptions. The process is computationally expensive, error-prone, and fundamentally limited by what documents happen to describe.

Through Carbon Arc's world model, the same query operates on structured organizational reality. Beneficial ownership is not inferred from textual descriptions but retrieved from resolved ownership hierarchies. Counterparty risk is not estimated from news sentiment but calculated from observed transaction patterns. Organizational relationships are not extracted from prose but queried from explicit relationship structures.

This is the intelligence layer that autonomous agents require: not optimized search over documents, but direct access to organizational reality as it exists now.

Part IV: Carbon Arc as Data Payment Processor—Where the Two Tokenizations Meet

The Infrastructure Inflection Point

The convergence of AI-driven demand and modern payment rails creates the preconditions for data to function as a true asset class. But preconditions are not sufficient. What is required is the full stack of market infrastructure that makes financial markets function: clearing, standardized instruments, real-time settlement, embedded compliance, and separation of custody from access.

Carbon Arc is building that stack. The parallels to financial market infrastructure are not metaphorical—they are structural. Carbon Arc functions as a data payment processor: clearing data transactions, settling consumption economically, and enabling the micropayment flows that data-as-asset-class requires.

Data Clearing: The Foundation

In financial markets, clearing serves a specific function: it standardizes instruments, resolves counterparty identity, nets obligations, and guarantees settlement. Raw trade instructions cannot flow directly between parties at scale—they must pass through infrastructure that transforms bilateral chaos into multilateral order.

Data faces an identical problem. Raw data assets are heterogeneous, inconsistently structured, and lack standardized identity resolution. Carbon Arc performs the clearing function for data:

Entity Resolution. Disparate records map to verified identities. The same organization appearing across SEC filings, news articles, and transaction records resolves to a single canonical entity.

Standardization. Proprietary formats transform into interoperable structures. Intelligence from diverse sources becomes queryable through unified ontology.

Rights Enforcement. Usage permissions validate before access occurs. Compliance embeds at the transaction layer, not as manual review overlay.

This is the equivalent of central counterparty clearing for data—analogous to the role the DTCC plays for U.S. listed equities. In securities markets, the DTCC sits between buyers and sellers, guaranteeing that trades settle even if one party defaults, standardizing how ownership transfers, and maintaining the authoritative record of who owns what. Before such infrastructure existed, securities trading was a bilateral affair—slow, risky, and limited in scale. The DTCC's clearing function is what enabled equity markets to operate at institutional scale with trillions of dollars in daily volume. Carbon Arc is building the equivalent infrastructure layer for data—the foundation that makes institutional-scale data transactions possible.

Structured Intelligence as Tradeable Instruments

Consider the evolution from physical asset ownership to securities. A security is not the underlying asset—it is a standardized, divisible, transferable representation of rights to that asset. Securities are more accessible than the assets they represent precisely because they are structured, priced, and tradeable within established frameworks.

Carbon Arc's intelligence layer performs an analogous transformation. Raw data is the underlying asset—valuable but illiquid, difficult to price, and impractical to transact at scale. The structured framework objects that Carbon Arc produces function as data instruments: machine-readable, modular representations that can be independently priced and consumed.

Carbon Arc's pipeline—ingestion and structuring, then modular output generation—mirrors the process that transformed illiquid assets into tradeable instruments. Each intelligence unit becomes discrete: clearable, settleable, priceable independently.

Consumption-Based Pricing as Atomic Settlement

Financial markets spent decades evolving toward real-time settlement. The gold standard today—atomic settlement in electronic FX markets, instant repo clearing on institutional networks—requires that instruments be structured for real-time clearing at the moment of transaction. You cannot settle what you cannot price; you cannot price what you cannot measure; you cannot measure what is not structured.

This is why structuring data into discrete, measurable units is prerequisite to real-time data commerce. Carbon Arc's consumption-based pricing model mirrors atomic settlement: consumption triggers clearing triggers payment in a single integrated flow.

The payment processor function:

- AI agent queries Carbon Arc's intelligence layer
- Query resolves to specific data consumption
- Consumption measures against defined pricing
- Payment settles atomically at query completion
- Data owner receives attribution and compensation

Without the underlying data structure that makes each query a discrete, measurable, priceable event, real-time settlement at the moment of consumption is impossible. Carbon Arc has created the infrastructure within which AI agents can transact as economic counterparties—agents that can pay their own bills. The process of creating this infrastructure has put more than fourteen patents in flight for the company, and they are still building.

Compliance Embedded at the Infrastructure Layer

Here is a structural truth about functioning markets: rules are enforced by the system, not by individual review of each transaction.

When an equity trade executes, compliance with securities laws is not negotiated bilaterally. The infrastructure (such as an ISDA Master Agreement) enforces those rules automatically. This is what makes markets scalable. Individual legal review of each transaction is not merely inefficient; it is architecturally incompatible with market function.

Data transactions today remain trapped in the pre-infrastructure paradigm: bespoke, negotiated licensing arrangements; manual compliance review and lengthy due diligence processes; bilateral contracts for each relationship. This cannot scale to an agentic economy processing billions of queries daily.

Carbon Arc embeds compliance at the infrastructure layer. Every data asset entering the stack undergoes provenance verification and rights validation at ingestion. Every subsequent transaction inherits and enforces those compliance parameters automatically. The rules are encoded into the transactional layer itself.

This is not merely protective—it is what makes institutional-scale data commerce architecturally possible.

And here is the compounding insight: each transaction reinforces the compliance architecture. The accumulated body of cleared transactions, validated usage patterns, and enforced rights creates operational precedent. This precedent may ultimately inform the regulatory frameworks that govern data as an asset class, positioning Carbon Arc not merely as responding to regulation but establishing operational norms from which regulation will emerge.

Atomic Ownership Rights and the Custody Transformation

The case for atomic ownership in financial assets is compelling but faces adoption barriers—incumbent resistance rooted in revenue protection. Data ownership faces no such entrenched intermediary structure. The market for data-as-an-asset-class is nascent, without legacy infrastructure to displace or incumbent tollbooths to circumvent. This creates an opportunity to establish atomic ownership norms from inception rather than retrofitting them onto existing structures.

Carbon Arc's architecture implements atomic ownership for data assets. Every data element ingested into the platform carries provenance metadata identifying its source. Every consumption event triggers attribution to that source. Every payment flows to the data owner based on actual usage, not estimated allocation of subscription revenue.

This transforms the economics of data contribution fundamentally. Under current models, data asset owners face a binary choice: retain proprietary control and forgo monetization, or surrender data to platforms that aggregate, repackage, and resell it—capturing most of the value while data originators receive flat licensing fees disconnected from actual usage. The 90% of enterprise data that remains siloed is siloed for rational reasons: the value extracted from sharing does not justify the control surrendered.

Atomic ownership rights invert this calculus. Data owners can contribute to the AI economy while retaining granular control over usage and receiving direct compensation proportional to value delivered. The infrastructure makes contribution economically rational for holders who currently—rationally—refuse to participate.

Separation of Custody, Access, and Settlement

Mature financial markets separate these functions by design. Custodians hold assets. Trading venues provide access. Clearing infrastructure settles transactions. This separation enables each layer to specialize, creates appropriate checks on each function, and allows asset owners to retain ownership while enabling liquid markets.

Data infrastructure today conflates these functions. Platforms that provide access typically also control custody. Payment mechanisms are entangled with access controls. This architectural conflation is why data owners face a binary choice: retain control and forgo value, or surrender control to extract value.

Carbon Arc's architecture separates these functions:

Data owners retain custody. Proprietary data assets remain under owner control. Carbon Arc does not warehouse data—it provides structured, permissioned access.

The platform provides access. AI systems query the intelligence layer within defined guardrails. Access is metered, measured, and governed.

Settlement infrastructure executes. Micropayments flow upon consumption in real time. Attribution traces to data sources. Economic value distributes to data owners.

This separation enables the 90% of siloed data to participate in the AI economy without surrendering the control that caused owners to silo it in the first place.

The Liquidity Transformation

Beyond individual data owners, atomic ownership enables something markets have never achieved: true liquidity in information assets.

Liquid markets require standardization—homogeneous units that can be priced, traded, and settled without extensive bilateral negotiation. This is why commodity markets require grading standards, why securities markets require CUSIP identifiers, why derivatives markets require standardized contracts.

Current data transactions are bespoke by necessity. Every enterprise dataset is structured differently, governed by unique licensing terms, with idiosyncratic compliance requirements. Negotiating access requires legal review, technical integration, and often months of diligence. This is the antithesis of liquidity.

Atomic ownership—combined with standardized provenance, embedded compliance, and micropayment settlement—creates the preconditions for liquid data markets. When every data asset carries machine-readable rights, every consumption event settles atomically, and every participant operates within consistent infrastructure, data can trade with the fluidity of financial instruments rather than the friction of enterprise sales cycles.

This is the market that Carbon Arc is building infrastructure to serve. Not data-as-product (the current model), but data-as-asset-class (the model that AI economics demand).

Part V: The Incumbent Resistance and the Structural Opening

Why Incumbents Cannot Build This Infrastructure

A reasonable observer might ask: if modern data payment infrastructure offers such clear advantages—atomic settlement, embedded compliance, consumption-based economics—why hasn't it emerged from incumbent players? The technology components exist. The theoretical benefits are understood.

The answer is not technological inadequacy. It is structural incentive misalignment. Efficient data infrastructure is antithetical to the business models of the institutions that control current data market plumbing.

The Intermediation Economy:

Opaque licensing enables price discrimination. Data vendors profit from bilateral negotiations where pricing depends on buyer sophistication and negotiating leverage. Transparent, consumption-based pricing would compress margins and eliminate information asymmetry that benefits sellers.

Bundled products create artificial switching costs. Platforms that combine data access with analytics, visualization, and workflow tools lock customers into ecosystems. Modular, interoperable data instruments would enable buyers to mix and match—threatening bundling economics.

Subscription models disconnect value from compensation. Flat fees benefit heavy users and subsidize light users, creating cross-subsidization that incumbents optimize around. Usage-based pricing would realign economics in ways that threaten established customer relationships.

Friction justifies intermediation. Every complexity in current data transactions—licensing negotiation, compliance review, technical integration—represents a service that intermediaries provide and charge for. Infrastructure that eliminates friction eliminates the intermediary's role.

The pattern is consistent: the inefficiencies of current data infrastructure are not bugs to be fixed—they are revenue sources to be protected.

The Data Buyer Cold Start Problem:

Data buyers also benefit from friction, creating a two-sided resistance to change. Large institutions with massive teams and data advantages move slowly—their competitive advantage partially lives in expensive complexity that smaller competitors cannot replicate. This creates a cold start problem: infrastructure that reduces complexity threatens the moat that sophistication provides.

The Platform Shift Opportunity

Platform shifts create opportunities precisely when incumbents are structurally inhibited from pursuing them.

The transition from mainframe to client-server computing created opportunities because IBM's business model depended on mainframe economics.[17] The transition from on-premise to cloud created opportunities because enterprise software vendors' business models depended on perpetual licensing. The transition from human-centric to AI-native data infrastructure creates opportunities because incumbent data vendors' business models depend on opacity, bundling, and intermediation.

Carbon Arc, built from inception for AI-native data access, is building the infrastructure that incumbents cannot build—not because they lack technical capability, but because that infrastructure would undermine the economics they depend upon. This structural opening has attracted investors like Liberty City Ventures, whose thesis centers on the convergence of tokenization infrastructure and AI-native data systems—precisely the intersection where incumbent business models create the greatest resistance to innovation. This is not a feature advantage or a timing advantage. It is a structural advantage rooted in the misaligned incentives of legacy players.

<u>Part VI: The Practical Impact</u>

Efficiency Gains

The two definitions of tokenization, developed independently over decades, are now converging to enable something neither could achieve alone: data functioning as a true asset class—with the clearing infrastructure, standardized instruments, real-time settlement, embedded compliance, and custody separation that designation requires.

---

[17] Christensen, Clayton M. (1997). *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Harvard Business School Press (analyzing how disruptive innovations displaced market leaders, with extensive documentation of the disk drive industry's successive disruptions from 14-inch mainframe drives to desktop and laptop formats); see also Christensen Institute. "Disruptive Innovation Theory." https://www.christenseninstitute.org/theory/disruptive-innovation/.

The efficiency gains are substantial. When AI systems can retrieve precise, verified information rather than searching exhaustively across training data, latency drops dramatically. When inputs are structured and authoritative rather than inferred from unstructured text, error rates decline by orders of magnitude.

Machine-Speed Intelligence Access

The agentic economy will operate at machine speed. When tens of billions of AI agents are making autonomous decisions—executing trades, managing supply chains, optimizing operations—they cannot wait for batch processes, overnight updates, or manual data refreshes. They require infrastructure that responds in milliseconds, not minutes.

Carbon Arc's MCP architecture was designed for this reality. By providing standardized, low-latency access to structured intelligence, the platform enables AI agents to query exactly what they need, when they need it, with response times measured in milliseconds. The MCP server acts as the interface layer between AI systems and Carbon Arc's cleared data infrastructure—translating agent queries into precise data operations and returning structured results that agents can act on immediately.

This is not theoretical. AI systems already use Carbon Arc's MCP server to access real-time entity data, market intelligence, and analytical outputs during live operations.[18] The infrastructure exists, functions at production scale, and demonstrates what data commerce looks like when optimized for machine participants.

Conclusion: The Infrastructure Imperative

The convergence of AI-driven demand for structured, verifiable data and modern settlement infrastructure creates the conditions for a new asset class to emerge. But asset classes do not emerge from technology alone—they require market infrastructure that incumbent players are structurally inhibited from building.

The institutions that control financial market plumbing—the custodians, clearinghouses, data vendors, and intermediaries whose business models depend on opacity, fragmentation, and settlement friction—face structural barriers to building the infrastructure the AI economy requires. Every efficiency gain threatens a revenue stream. Every transparency improvement undermines a competitive moat. Every reduction in intermediation eliminates a tollbooth.

This creates the structural opening that platform shifts always create: incumbents optimize within the existing paradigm while new entrants build for the paradigm to come.

The history of financial markets demonstrates this pattern clearly. Securities required clearinghouses. Derivatives required standardized contracts. Electronic trading required matching engines and settlement systems. Each evolution from bilateral negotiation to institutional-scale markets required purpose-built infrastructure.

Data stands at this inflection point. The 90% of global knowledge currently siloed—held by enterprises, institutions, and individuals unwilling to surrender control for uncertain returns—represents the addressable opportunity. Unlocking it requires infrastructure that preserves ownership while enabling value extraction: clearing for data, standardization into tradeable instruments, real-time settlement at the moment of consumption, compliance embedded at the infrastructure layer rather than negotiated bilaterally, and AI-native access patterns built for machine speed.

But infrastructure alone is insufficient. What distinguishes viable infrastructure from aspirational architecture is whether it enables systems to operate on reality rather than representations of reality. This is the

[18] Carbon Arc Corporation. (2025, November 3). "Carbon Arc Launches MCP Server to General Availability, Powering the Next Generation of AI-Driven Insight" [Press Release]. https://fisd.net/carbon-arc-launches-mcp-server-to-general-availability-powering-the-next-generation-of-ai-driven-insight (announcing general availability of Carbon Arc's Model Context Protocol Server, enabling LLMs and AI agents to connect directly to Carbon Arc's data stack for real-time insights through conversational queries).

fundamental limitation of current AI systems: they operate on static snapshots, training corpora frozen at points in time, unable to perceive the living, dynamic world they are meant to serve.

Carbon Arc is constructing what current AI architectures cannot: a continuously updated, entity-centric representation of organizational reality—a world model built from actual transactional data, behavioral signals, and relationship structures rather than inferred from text. This world model is not a database. It is a living representation of state, behavior, and relationships that evolves as the organizations it represents evolve.

Carbon Arc is building those rails for the AI economy—not AI as a feature applied to legacy data architecture, but the foundational layer that transforms data into a functioning asset class.

The company functions as clearing infrastructure for heterogeneous data, produces standardized instruments through its intelligence layer, enables real-time settlement via consumption-based pricing, embeds compliance at the transaction level, and implements atomic ownership rights that preserve data owner control while enabling liquid markets.

What distinguishes this approach is architectural conviction rather than incremental optimization. Traditional data platforms were built for human analysts operating at human speed within human compliance frameworks. The MCP implementation demonstrates that Carbon Arc is building for a different reality: autonomous agents transacting at machine speed, requiring verified data rather than probabilistic retrieval, and settling consumption atomically at the moment of value exchange.

When AI agents need to understand beneficial ownership, evaluate counterparty risk, or map organizational relationships, they will not search documents—they will query the world model. When tens of billions of autonomous agents make decisions in real-time markets, they will require infrastructure that reflects reality as it exists now, not as it was described months ago in training data.

The platforms that establish this operational infrastructure will hold structural advantages analogous to those held by early clearinghouses, early exchanges, and early data vendors who built the rails that financial markets came to depend upon. But the advantage is not merely in providing infrastructure—it is in constructing the representation of reality that AI systems will query billions of times daily.

Carbon Arc—not the incumbents structurally inhibited from disrupting their own business models, not the transitional AI platforms optimizing human workflows, not the institutions that profit from opacity—is positioned to define what comes next.

That positioning rests on a foundation already built and operational: production-grade infrastructure processing real transactions, generating real intelligence, and enabling real economic exchange between data owners and AI systems. The world model exists. The clearing infrastructure functions. The compliance framework operates. The settlement mechanisms execute. The agentic economy is not theoretical. It is emerging now. And the infrastructure it requires is being built by those unconstrained by legacy business models and uncompromised by incumbent incentives.

This is Carbon Arc's structural advantage—and the market opportunity it represents.

.