



THE CENTRALIZED DATA CATALOG

Solutions for a Digitized Industry

Michele I. Kelsey
Richard F. Clements

Forward by Brian Buzzelli

October 23, 2023



Table of Contents

Forward by Brian Buzzelli	2
A Single, Centralized Data Catalog to Transform the Market.....	4
Capabilities for Both Vendors and Consumers	5
Neutrality.....	5
Standardized, Digitized, Common Language	6
Data Identifiers and Common Taxonomy.....	6
On the Question of Transparency	7
Addressing Financial Services Industry Pain Points	7
Utilization of Multiple Data Sources.....	7
Data Usage Rights.....	8
Regulation, Compliance, and Risk Management	8
Looking Ahead to a Digitized Future	8
The VendEx Data Ecosystem: Bringing Vendors and Consumers Together.....	10
Vendor-Friendly Approach	10
VSource: The Dynamic Digital Data Catalog	10
VLink: Market Data Customer Connection	10
VendEx Data Registry™: The Golden Source of Standardized Data.....	11
The VendEx Identifier (VID): Industry Standard Data Identifier	11



Forward by Brian Buzzelli

Author of *Data Quality Engineering in Financial Services*, O'Reilly (2022)

The financial services industry is undergoing a significant transformation driven by the explosion of financial data and the adoption of artificial intelligence (AI). To meet the evolving demands, there's a growing call for a common data taxonomy and granular data product identification in the form of a data catalog that can serve both vendors and consumers of data. The aim of the data catalog is to provide a standardized, normalized, and neutral platform where vendors can control their data specifications and consumers can search, access, and use data product identification and information effectively. A common data catalog and data identification would address many industry challenges, including interoperability, data lineage, and data usage rights. The use of common identifiers and a granular taxonomy is vital to describe and tag datasets in a standardized manner, allowing for more efficient data management. Additionally, the data catalog can play a crucial role in addressing regulatory compliance, risk management, and contractual usage rights, ultimately benefiting the industry as a whole.



VendEx Solutions is a notable player in this space, providing a dynamic digital data catalog called VSource. This catalog enables data consumers to research and compare products from over 4,000 vendors and facilitates communication between vendors and consumers. The VendEx Data Registry™ standardizes and categorizes available vendor market data products, allowing consumers with permissioned access to vendor product information. The proprietary VendEx Identifier (VID) serves as a standard data identifier, offering a host of benefits to data consumers, including faster product searches, accurate vendor product billing, consistent linkage between contracts and invoices, improved reporting, and reduced risk. This approach aims to bring both vendors and consumers together through a neutral, centralized platform, providing a scalable and cost-effective solution to industry challenges in an increasingly digitized financial services landscape.



Brian Buzzelli is a proven leader in transforming data architecture within the financial services industry, having extensive expertise in information management, data quality and governance, business and technical development, and professional services. His innovative product and strategic business development skills, underpinned by a deep understanding of the industry, technical expertise, and effective data management leadership, are instrumental in achieving business strategy and revenue goals for delivering data, products, and services to the investment management sector. Brian's strong commitment to delivering high-quality, data-driven solutions has optimized performance, enhanced efficiency, and elevated data quality for global financial services and asset management firms. His recently published *Data Quality Engineering in Financial Services*, O'Reilly (2022) focuses on the application of rigorous quality assurance techniques used in manufacturing applied as quantitative data quality validations to data used in the financial industry. He is enhancing his data quality and business strategy expertise with the application of AI through an executive program at MIT CSAIL (Computer Science and Artificial Intelligence Lab).



A Single, Centralized Data Catalog to Transform the Market

Financial services has always been at the forefront of the data industry, from the invention of the ticker tape in 1867 to its demise 100 years later when television and computers were leveraged to convey financial information.

Financial data and its applications have increased exponentially in the last decade. Some datasets have developed as a result of new instrument types or new market disciplines, for example, cryptocurrency and ESG. Others have emerged based on the non-traditional approaches financial professionals are taking to accomplish their goals, such as using non-financial data (alternative data) to analyze risk and viability. In the 1990s, the phrase "Big Data" became popular. In the Roaring 20's (the 2020s, that is), the new buzz phrase is AI. As AI matures, there will be a need for more and more data, as well as newer ways to process and utilize this data.

A recent movement in the financial industry has been the call for data catalogs, with some financial institutions receiving feeds from vendors that have a semblance of a catalog. A data catalog is a necessary component of an industry-wide solution, but to function as such, it must serve the needs of both vendors and data consumers. It is also clear that for this to happen, vendors must have a level playing field to be treated fairly and equally.

An effective solution requires a centralized data catalog built by a neutral third party. The solution represents both vendors and consumers of data fairly and equally, where vendors have control over the content and distribution of their data specifications.

Separate data catalogs built by individual consumers, receiving information from vendors delivered via disparate technologies with varying definitions and unique ways of identifying and describing the details of the data, will make the exercise costly and untenable. Many firms that have started down this path are beginning to realize the enormous amount of effort needed to standardize these data catalogs, as well as the cost of maintaining them. The technology, the definitions, and connectivity are not standardized, limiting the potential uses and benefits or requiring inordinate amounts of work from each institution to normalize the information.

Over the years, the financial industry has come together to address issues such as pricing and valuation services, centralized loan-level data services, and centralized electronic trading systems for OTC instruments. Similarly, data catalogs call for collaboration from both vendors and consumers of data to represent and address the current and future industry challenges.



Capabilities for Both Vendors and Consumers

For a true solution to exist, there must be connectivity and compatibility with other consumers, with vendors providing information that can be normalized for the consumer. Among the primary issues are interoperability of technologies and a common language and vocabulary. How will the industry communicate, particularly when technologies among providers differ? And what happens when (post-integration) one of them makes a change? At the same time, a critical concern among vendors is and will continue to be protecting proprietary data field information and ensuring it is viewed only by the desired parties.

A data catalog is the foundation that can truly digitize the industry and bring about transformational change in financial market data. That said, the solution must be properly constructed to be effective. A data catalog solution that will support the needs of the industry must be normalized, centrally maintained, represent all vendors fairly, and offer a utility that enables vendors to control the content and distribution of information about their data.

No vendor is an island; vendors building their own catalogs will not help their consumers. Similarly, consumers who have tried to build a data catalog have come to recognize the high cost of normalizing, not to mention maintaining the data.

Data catalogs built by individual entities with unintegrated, nonnormalized data and technology will not provide the building blocks to support a true data solution between vendors and consumers. These building blocks include interoperable technology, connectivity, standardization, a common language, and a taxonomy that allows for a granular way of describing, identifying, tagging, and actioning datasets.

Neutrality

Neutrality is essential to developing and maintaining a market data ecosystem to serve both vendors and data consumers. A solution provider must be a neutral third party that builds and maintains the platform and acts as an independent resource and facilitator without benchmarking or rating any vendor, solution, or dataset.

The cost of storage, aggregation, and normalization of data is considerable, particularly when each consumer and provider of data is managing their own catalog and attempting to map it to their vendors, and vice versa. Vendors must have the tools to address these key industry issues, including control over their data specifications: vendors must have the ability to permission and specify exactly who may access their information and what information is revealed.



The industry recognizes the need for a centralized data catalog for products and services. However, just as importantly, there must be a vendor registry for discrete data fields where vendors can control the information and capabilities that uniquely set them apart and control who can view this information. Vendors have made significant investments in their data solutions. Operating within a centralized solution that fairly and equally represents vendors will be key to addressing industry needs.

Standardized, Digitized, Common Language

A data utility for the industry must be standardized and normalized, using a common language, vocabulary, and set of conventions (for example, a Data Style Book). At the same time, it must be able to capture a high level of granularity to describe the data.

The objective is to provide a "Golden Source," or single version of the truth – a single, well-defined, precise description of the data, which flows throughout the vendor and consumer market data ecosystem. A data point or descriptor need only be changed once, and that change cascades throughout the system.

Another element that is critical to identify is data lineage. Where did the data originate? Is it being redistributed by another party? In what way, if any, did the provider modify or add value to the data? Again, this must be captured to ensure that if a data characteristic changes, it is changed throughout the system.

To achieve this successfully, the vendors must contribute, be represented equally and fairly, and access, update, and control their information.

A data identifier is essentially a bar code for data, assigning alphanumeric codes to key attributes that describe the data at a granular level.

Data Identifiers and Common Taxonomy

The ability to determine precisely what a dataset contains is a vital part of managing a data portfolio. A data identifier is essentially a bar code for data, assigning alphanumeric codes to key attributes that describe the data. It is not enough to say, "We provide data on loans." What kind of loans? Syndicated? Middle market? Auto loans? Does the data include news, pricing, valuations, terms and conditions, symbology, analytics? Which organization originated the data? Is the vendor providing the data, or are they redistributing data from another vendor?

There must also be a common language upfront and the ability to generate a data identifier at the most detailed level. Key elements of data identifiers, which are built from the data taxonomy,



need to include type of solution, classification (e.g., asset class, ESG), geographic coverage, data update frequency, languages available, data provider (which may differ from the originator), and many more. Based on the type of data, there may be additional descriptors. For example, in the case of foreign exchange, there would be a need to identify fields, such as base and quote currency and tenor.

On the Question of Transparency

Further digitization of market data is necessary and inevitable. The notion that digitization of any market creates increased transparency, which will lead to pressure on margins, has been around for at least two decades.

However, looking at the financial instruments that have become digitized over this period, it is a dated concern.

What has actually come to pass is that digitization has benefitted these markets materially in terms of liquidity, identification of counterparties, management of risk, compliance, and regulatory requirements, all of which have been highly beneficial.

Addressing Financial Services Industry Pain Points

Once a well-constructed data catalog is in place, the industry can more effectively address critical concerns. A neutrally provided centralized data catalog provides a data registry where vendors can control the content and viewing of their own data. A common language and a data taxonomy provide data identifiers at a more granular level and can address a host of industry challenges.

A neutrally provided centralized data catalog, a data registry where vendors can control the content and viewing of their own data, a common language, and a data taxonomy that facilitates data identifiers at the most granular level will address a host of industry challenges.

Utilization of Multiple Data Sources

Consumers of market data require multiple data sources for a variety of reasons, including the fulfillment of regulatory requirements, business continuity, and compliance, among others. Vendors know they are often not the sole provider of a dataset to a given consumer. This may be for compliance reasons, e.g., Net Asset Value (NAV) pricing, where there must be at least two, if not more, pricing or valuation data sources. In many cases, particularly for the larger data consumers, it is appropriate and necessary to have more than one set of data, if for no other reason, to have comparative points of view. While data consumers want to avoid accidental



duplication of datasets, digitization and a common platform provide the tools to determine that if a dataset is duplicated, it is a deliberate and researched purchase.

Data Usage Rights

Market data consumers often struggle to track and comply with complex requirements around how they are contractually authorized to use data they have licensed. These requirements vary significantly by product and vendor and are often very specific. Even within the same business unit, usage rights differ based on business purpose, distribution, and delivery, e.g., publishing, on-screen only, creating derivative works, etc. Noncompliance, even when unintentional due to lack of data, can cause operational, legal, and financial issues, resulting in sizeable fines and penalties.

Tying contractual usage rights to products via a centralized catalog is invaluable in pinpointing the specific datasets impacted and precisely what may be done with the data.

Regulation, Compliance, and Risk Management

The challenge that market data consumers face in staying ahead of regulatory requirements is more complex than ever before. Regulatory data and resulting compliance and risk management solutions have seen significant expansion in the last decade, particularly in Europe and the UK. There are approximately 70 capital markets regulations globally, and almost 70% of them apply to Europe and the UK. In the US, the financial crisis of 2008 led to the creation of the Consumer Financial Protection Board (CFPB) and a wave of US financial regulations. All of these regulations require solutions to manage reporting and compliance. Asia, while generally accountable for the UK and European regulations as part of global entities, also has numerous very specific regional and country-level regulations.

A centralized data catalog, which provides common identifiers and the ability to describe capabilities and datasets at the most granular level, will enable data consumers to pinpoint and source the required datasets across multiple vendors.

Looking Ahead to a Digitized Future

The financial services industry is undergoing a transformation unlike any that has occurred in the past. With a sharp increase in the types of datasets and applications, both consumers and vendors have moved toward building their own data catalogs. This is proving to be a costly, unsustainable approach. Proprietary data catalogs all have differing technologies, definitions, and ways of identifying and describing the data details.

Bringing consumers and vendors together via a neutral, third party-provided centralized data catalog is a solid starting point. However, this solution must be constructed properly to address



the industry's fundamental issues to support an impactful solution. Digitization of the data through common technology, normalization, a common language, and a granular way of describing, identifying, tagging, and actioning datasets will address the financial industry's pain points in a scalable and cost-effective way.



The VendEx Data Ecosystem: Bringing Vendors and Consumers Together

VendEx Solutions is the centralized hub for the financial services market data industry. The VendEx Ecosystem is built on a foundation of industry-standard identifiers and a dynamic digital directory of vendor products and services, powering a suite of sophisticated integrated tools and workflows that eliminate industry frictions for both market data management and financial technology vendors. Leveraging deep industry knowledge, VendEx works with the world's largest market participants to create, design, and ultimately deploy the only centralized utility with tools built by the industry, for the industry.

Vendor-Friendly Approach

VendEx Solutions does not and will not benchmark or rate any entities. The objective is to facilitate connectivity through technology solutions that represent vendors and data consumers equally. Vendors have complete control over the content and distribution of information about their data.

VendEx has worked with the world's largest market participants, both vendors and consumers, to create, design, and ultimately deploy the only centralized utility with tools built by the industry, for the industry.

VSource: The Dynamic Digital Data Catalog

The **V**Source catalog is a digital, flexible, dynamic, and searchable database that enables data consumers to research and discover detailed information on over 4,000 vendors. It allows consumers to identify and compare products and request authentication and updates from vendors. Vendors contribute and authenticate company and product data through the VendEx Data Registry™.

VSource powers the VendEx Ecosystem, which includes contract management tools, inventory management dashboards, and regulatory products. The database incorporates rich sector directories, including alternative data, indices, energy, fixed income, equities, cryptocurrency, and derivatives, and allows for API integration of vendor and product descriptions into internal platforms.

VLink: Market Data Customer Connection

VLink is the vendor gateway to the VendEx market data ecosystem, enabling vendors to highlight and share detailed information about products and services with market data consumers.



Vendors can use VLink to review and update their profile information and corresponding product pages published in the VendEx digital data catalog (VSource.) Using VLink, vendors can showcase their critical product and service benefits, in addition to features, and target products to specific user categories. Clients and targets are able to access the searchable golden source of standardized data fields associated with different data products.

VLink also enables vendors to upload promotional marketing content, such as product brochures, webinar announcements, videos, white papers, and case studies. Vendors can then publish alerts that display only to specific targets that would benefit from the information and map discrete services to specific regulatory requirements.

VendEx Data Registry™: The Golden Source of Standardized Data

VendEx utilizes catalog information from vendors to standardize and categorize available market data. Vendors also provide granular data information, including data fields, methodologies, and governance docs, accessible upon request. Once permissioned, consumers can also view all vendors that provide access to that data.

Vendor products and services in the VendEx Data Registry™ are assigned VendEx Identifiers (VIDs), a common market standard for categorizing their products and services for contracts and digitized direct invoicing. The data originator controls who sees its data. Data information is published to the VendEx Data Registry™ with permissioned access to the vendor's product pages.

The VendEx Identifier (VID): Industry Standard Data Identifier

The VendEx Identifier (VID) is a patent-pending alphanumeric taxonomy standardizing the identification of data fields offered by financial services vendors in their various products. The VID captures vital information about vendors, products, and their specific features and capabilities via an alphanumeric code, essentially a bar code for datasets. The elements captured include vendor, product, delivery method, product and content update frequency, solution type (pricing, analytics, etc.) and classification type (asset class, industry sector), geography, language, data source, and data originator.

The VID also captures vendor hierarchy, which shows the parent/child relationships between vendor parent companies and subsidiaries. This is particularly valuable in gauging exposure to a given vendor in the event of a corporate action and the ability to roll up products and capabilities and look at the entire vendor organization.

Using the VID as a standard market identifier provides a host of benefits to data consumers, including faster and more granular product searches, relief of invoice friction for faster, more accurate billing cycles, a consistent linkage between contracts, invoices, usage declarations, and usage rights, improved reporting, and reduction of risk.



For more information, visit vendexsolutions.com or contact VendEx Solutions at info@vendexsolutions.com.

US: San Francisco | New York | Charleston

EMEA: London | Lund

APAC: Singapore