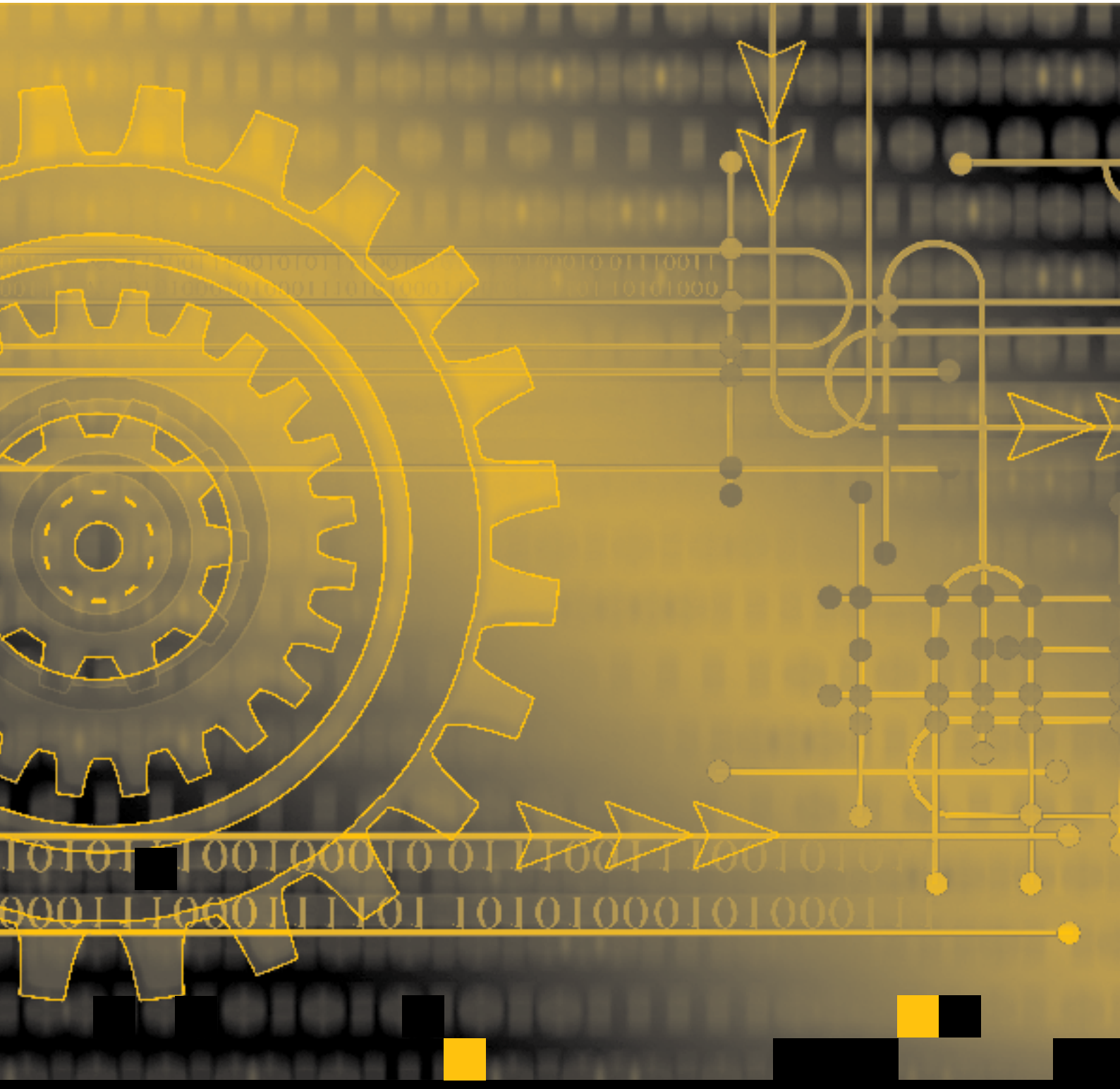
Data Management Review
FROM A-TEAMGROUP



Applying Open Data Principles to Financial Data Governance

Commissioned by

Bloomberg

Contents

Executive Summary	3
Why Current Data Governance Programs Are Ineffective	4
The First Step in Data Governance: Defining the Data Catalog	6
The Six Core Data Quality Checks	8
Data Governance for BCBS 239	10
Delivering Tangible Data Governance Results	11
How to Expedite Data Governance Progress with Bloomberg	12
Call to Action	14
About A-Team Group's Data Management Review	15
About Bloomberg for Enterprise	16

Executive Summary

Data quality and data management are the foundation for true data governance, but are often hampered by firms' focus on the human aspect rather than on identifying and understanding the full scope of the data used within the enterprise.

- Commercially available data management software can help. But generic software is essentially an empty vessel that lacks domain knowledge of financial data, specifically of common financial data vendor formats and industry standards for trade and transaction reporting.
- CDOs and their teams often get sidetracked dealing with governance committees and getting governance software to fit their firm's needs, thus achieving little in terms of improved data quality.
- What's needed is a centralized framework for understanding all data used and available within the enterprise, typically in the form of an online "Data Catalog" that lists all data sources and the metadata that describes data characteristics and provenance.
- Putting a Data Catalog in place is a necessary first step toward effective data governance. The second step is data quality, with the third step being data management. Getting started can be a challenge, but so-called Open Data initiatives by government and in other industries describe best practices in organizing metadata in support of an effective data governance program.

Why Current Data Governance Programs Are Ineffective

Effective data governance delivers data quality and controls across the enterprise and ensures that financial institutions can meet regulators' demands for transparency and full audit trails. But current approaches to data governance that rely on software and policies are ineffective without the right metadata flowing through them.

It is now widely recognized that to have effective controls over enterprise-wide data, financial institutions must first gain control of data governance. Once effective data governance is in place, firms can start to ensure that the highest quality data is used throughout the enterprise. Comfortable that their data is of the highest quality, CDOs and their teams can then set about instituting true data management and meet the needs of both the business and, increasingly, regulators.

But current approaches to data governance are proving ineffective, mostly because firms have focused on the human aspect. Financial institutions have been approaching data governance by forming committees, assigning data stewards, hiring expensive consultants and conducting detailed reviews of processes. Elsewhere, they have purchased generic data governance software to help provide a framework for managing the process.

Committees, however, can struggle to understand the full scope of data that needs to be managed. And the software they have purchased is essentially an empty vessel that comes without any domain knowledge of financial data and especially no knowledge of common financial data vendor formats and industry standards for trade and transaction reporting. Making the vessel fit the data sets in hand is a significant challenge, to say the least.

Often, the Chief Data Officer and the data management team get stuck in the weeds of database schemas with DBAs and data modelers as they try to figure out what needs to be governed and how to get it to fit the structure of the off-the-shelf governance solution.

Frequently what is missing is a central framework or superstructure that clearly defines data governance practices. That framework is metadata, or data about data.

Metadata describes all fields of data, such as the definition, data types and properties, but it can also provide descriptive information about the provenance of data, such as data sources and publishers, lineage, time stamps, categorization and other aspects.

Ensuring that you have robust metadata across all data sources, internal and external – establishing a so-called "Data Catalog" of all data sets used within business processes across the enterprise – should be the first step in any data governance program. The Catalog is typically a web portal that provides access to a list of the machine-readable data available within the organization. This enables everyone to be in agreement about what data is being managed and makes the enterprise data landscape understandable without having to delve too deeply into the complex web of data sources, applications, spreadsheets, workflows and more.

Only when you have clear definitions and a uniform approach to categorizing data can you move on to applying processes, metrics and quality checks to that data and deliver tangible results to the business.

This white paper argues that by applying widely accepted Open Data concepts to financial information, CDOs can rapidly create Data Catalogs, the first step on the road to effective, tangible data governance, enhanced data quality and – the Holy Grail – true data management. It explains what metadata is in the context of data governance and data quality, explores the approaches and challenges to sourcing effective metadata and outlines a best practice approach to achieving measurable results in data governance.

The First Step in Data Governance: Defining the Data Catalog

A strong definition of the metadata underpinning the information used by the business is fundamental to data governance. A robust Data Catalog needs to explain what data is available, its formats, what transformational rules are applied and what processes the data has undergone throughout its life cycle.

The Data Catalog is a single point of reference for all data attributes. Whatever the source – internal transactional data repositories like Murex or Calypso or any of the commercial data vendors – the Data Catalog ensures consistency in the metadata that describes the attributes of all data sets. With this in place, business users are able to consume data from multiple sources – internal or external – and apply it to solve business problems, safe in the knowledge that its provenance is good.

For example, if you have a Bond Duration field, the metadata relating to that field would determine that the value is a decimal, the units of measurement are years and that it's a kind of Macaulay Duration on the official closing Bid Price. From a provenance point of view, you would also track such attributes as the source of the data and a description and time stamp related to any changes.

You can then build rules to check that the data meets the standards defined in your Data Catalog and that any data not conforming to those rules can be flagged as an exception to be dealt with by data analysts, thus ensuring that quality standards and governance processes are met.

Six data definitions for developing a robust Data Catalog

The scope of data definitions should cover:

1. Products: data as it is bought or licensed
2. Data sets: the individual files
3. Records: the rows and structures in the files
4. Properties: individual fields in records
5. Enumerations: the list of valid values that act as a "lookup"
6. Provenance: the sources and transformations of the data

Complexity of financial markets

An in-depth knowledge of financial markets is required to meet the needs of the data governance process for a financial institution. This necessarily entails an understanding of data used across multiple asset classes, from vanilla listed equities through the most complex OTC derivatives.

On top of the financial markets' specific elements, the Catalog must define the attributes of the data in question. These extend to products, data sets, records, properties, enumerations and provenance (See Box, left).

Ensuring flexibility in the Data Catalog

Data governance for your financial institution is further complicated by the different uses for data across the enterprise. Users in the front office, for example, may make use of large volumes of "indicative" price quote information (perhaps consisting of buy and sell orders that are later cancelled), while middle- and back-office users may be more interested in

smaller volumes of "actual" trade data, specifying the price and volume at which specific deals are done.

To meet operational needs, the Data Catalog has to be flexible enough to accommodate many users and vary its rules to meet the varying needs of different groups within the enterprise. And, from a data governance point of view, it needs to be clear who owns the different data sets and what technology should be used for storing and managing that data.

The format and standards of the Data Catalog

Metadata should be readable to both people and machines. And it should be based on commonly accepted standards and conventions in order to make it easier to integrate with existing tools and industry expertise, further simplifying quality control. So-called Open Data standards have been widely adopted outside of financial services, particularly by government agencies and the technology industry, and can provide the bedrock for establishing a Data Catalog that can be used across a broad array of data types.

Examples of standards that can be used include:

- XML Schema for datatype standards
- Dublin Core, Schema.org, Simple Knowledge Organization System (SKOS), Friend of a Friend (FOAF), DBPedia for metadata standards
- International System of Units for units of measure
- ISO Standard 8601 for date formats
- W3C's Open Rights Statement Vocabulary for managing licenses related to each data set
- W3C's PROV format for provenance information related to each data set

Once you have defined the data itself, you can use it to drive rules and processes to deliver data quality.

The Six Core Data Quality Checks

Once you understand the full scope of data across the enterprise – and you are clear about its categorization through a well-defined Data Catalog – you can begin to apply meaningful data-quality processes and metrics.

Delivering high data quality is essential to enabling decision makers to make informed decisions and also to meeting the regulatory demands for the ability to track the provenance of data that have informed those decisions.

Six essential measures can help to ensure that data meets the right quality standards. Many financial institutions today might run one or two of these checks, but, to be truly effective, all six need to be implemented.

While there are many definitions of data quality, it is broadly accepted that quality data is fit for its intended uses in operations, decision making and planning. These data-quality checks can be used to ensure that the governance process is establishing the key foundations for data quality: accuracy, consistency, reliability, appropriateness, relevance and completeness.

Data Quality Process		Example of Quality Check
1 Lexical Validation	Checks that value is represented in correct lexical format	Price is represented as "2" versus "two"
2 Value Checks	Checks that value is permitted	Price > 0
3 Row Check	Checks that the right fields are complete based on the record type	If equity price, "Maturity Date" should be blank
4 Entity Resolution	Links common values across datasets, using lexical probabilistic modeling	If identified as "Michael," create a rule to match with "michael" but not "Michaela"
5 Similarity Coefficient	Compares value to other, similar data points to ensure that it is "reasonable"	EPS for equity is "reasonable" when compared with rest of sector
6 Explainable	Attribution technique to explain change	Change in price can be explained based on news, corporate actions, market, etc.

Increasing Complexity

Lexical Validation. Each data type belongs to a canonical lexical space, which refers to the accepted format for referring to the value space within the data type. A system may handle values as digits rather than words ("2" vs. "two"), for example.

Value Checks. A simple check to ensure that a given value is permitted for the data type in question. This may be a simple "greater than zero" check or a more complex business rule based on accepted norms.

Row Check. Another simple check based on domain knowledge of financial services. For example, some data fields do not apply to some asset classes. Thus, a table of data relating to equities markets would be subject to a rule expecting a "Maturity Date" field (applying to fixed income and/or derivatives) to be blank.

Entity Resolution. Business rules can be used to resolve discrepancies between records. For example, a business rule may be used to identify a misspelled identifier due to data capture error, while also differentiating it from a similarly spelled but discrete identifier.

Similarity Coefficient. Data-quality checks for individual records are only part of the picture. It is critical to look at changes in the data and to ascertain whether or not the change is within tolerated limits. Advanced data-quality measures use coefficients of similarity or similar techniques to assess value-space conformity.

Jaccard's coefficient (measures similarity) and Jaccard's distance (measures dissimilarity) are measurements of asymmetric information on binary (and non-binary) variables. For example, Jaccard can be used to assess the plausibility of a change to the ID_ISIN value of a single security where it might have been entered incorrectly: US0378331005 -> US0378331006.

Similarity coefficients can also be applied across datasets. For example, a change to 10% of all ISINs in a data set can be deemed implausible. Similarly, LAST_PX price data will be expected to change continually during a given time window, so "no change" across a wide range of price points would be implausible.

Explainable. Where changes in data values breach accepted norms, it is essential to understand whether the incidence is explainable. The "explanation" of a price change, for example, may be derived from a market event such as a news item or corporate action.

Key Statistics

- Fully 92% of financial services firms are unable to quantify the cost of the risk resulting from poor data quality
- Nearly 70% of financial services firms are developing or piloting new data-quality initiatives; 30% already have enterprise-wide programs
- Over 80% of firms are planning to spend greater than £5 million in the next 12-18 months to improve data governance

Sources: Informatica/WBR: Modernizing Data Quality & Governance survey of financial services companies; DTCC survey of financial services companies.

Data Governance for BCBS 239

From the start of January 2016, all globally systemic important banks (G-SIBs) and many domestically systemic important banks (D-SIBs) have had to comply with the Basel Committee's BCBS 239 risk guidelines. BCBS 239 establishes 14 principles requiring mandated institutions to have IT architecture and management information systems in place to support automated, timely and accurate risk aggregation and reporting practices firm-wide.

Given these goals – and the fact that the effectiveness of a risk management system is directly dependent upon the integrity of its underlying data – BCBS 239 also includes requirements for using metadata in establishing the integrity and lineage of data, calculations and reports arising out of risk systems. These requirements, primarily articulated through BCBS-239 Principles 2 and 3, apply even to data provided by external parties.

To meet BCBS 239 requirements, consumers of this data will require metadata relating to all data fields, including definitions, data types and properties, across all data received from third-party providers. Banks will need to understand the metadata underlying all content sets and their components, including the location and generation time of delivered content.

To meet the automated aspect of the requirement, ontology and provenance metadata should be machine process able and delivered in W3C standard formats. Additionally, metadata should be transformable into the standard of the bank's choosing, whether FIBO, SEDOL for security identifiers or a proprietary format.

What is BCBS 239?

The Bank of International Settlement's Basel Committee on Bank Supervision (BCBS) Regulation 239 comprises 14 principles focused on aggregating accurate risk positions and reporting to the regulator. These principles can be grouped into four distinct themes.

Completeness. GSIBs and DSIBs should be able to capture and aggregate all material risk data across business lines and be able to measure and monitor the comprehensiveness of that data.

Timeliness. Banks should be able to generate critical, aggregate risk data in a timely manner, including aggregate credit exposure to large borrowers, counterparty credit risk, trading exposures, liquidity risk indicators and time-critical operational risks.

Accuracy. To minimize errors in risk aggregation, banks should principally rely on automation and establish a dictionary that uniformly defines data across their organization.

Adaptability. GSIBs and DSIBs should have flexible data aggregation and reporting practices to meet a broad range of on-demand reporting requests, particularly during crisis or stress situations.

Delivering Tangible Data Governance Results

With a structured and uniform approach to categorizing every data element across the enterprise, the CDO or Head of Data Governance can then query across all these data sets in one stroke.

For example, a query can look at what data has been used since a certain date, the sources of the data or what changes have been made to that data. Another query could look at data usage to understand who used what data and via which application, whether the data is actually being used or not and where gaps exist in data.

In terms of delivering tangible results from this approach to data governance, a robust Data Catalog containing all requisite metadata offers the following benefits:

Metrics for data quality. It's often said "if you can't measure it, you can't improve it" and nowhere does this apply more than to data quality. The metadata underpinning a firm's data governance initiatives can support a set of metrics for data quality, which, in turn, can be used to benchmark current performance for future improvement.

Improved compliance. Many new and emerging regulations – among them BCBS 239, Dodd-Frank, CCAR, Solvency II and MiFID II – require firms to provide transparency into their risk and capital positions. Implementing a workable data governance program with solid metadata foundations can improve a firm's chance of compliance, thus reducing the risk of significant fines, capital charges or reputational damage.

Satisfying regulatory inquiries. This same foundation can help firms meet the on-demand aspect of regulators' ad hoc inquiries by drawing on data sets that have already been validated for quality, applicability and timeliness.

Faster results. More broadly, analysts and other data consumers will be able to efficiently access this high-quality data set, enabling faster and better decision making.

Saved costs/resources. Finally, improved data quality can streamline operations across the board, reducing the risk of mismatches and costly reconciliations, lowering the incidence of regulatory penalties and cutting the cost of reliable data for business purposes.

How to Expedite Data Governance Progress with Bloomberg

Achieving agreement on data governance can be politically and technically challenging within a financial organization. Working with data experts to build the required metadata foundation can speed up the process considerably, save significant costs and deliver higher quality.

For customers of Bloomberg Data License, full human- and machine-processable metadata updated daily is available for all:

- Data License fields
- Data License datasets and dataset layouts
- Data License products
- Data License notifications (beta)

Further, there is the availability of the hundreds of thousands of data-quality rules such as those assessing:

- Data encoding
- Conformance of data to its schema
- URL resolvability within the data
- Overreliance on string or token datatypes ("stringly-typed" test)
- Financial domain-specific rules, e.g., Maturity Date blank for Equities
- Etc.

Financial institutions can leverage Bloomberg's deep financial markets meta-model and data-quality rules to apply to their own internal data. The same meta-model and data-quality rules can also be applied to third-party data sources through Bloomberg PolarLake, enabling you to map the entirety of your enterprise data.

Leveraging these already developed meta-model and data-quality rules across internal and external data can save CDOs the time-consuming effort of building their own Data Catalog, much of the time spent in committee meetings on the data definition phase and the money spent on expensive consultants and internal resources to try to figure out the map of your data across the enterprise and applicable data quality rules.

Securing a strong foundation through the use of Bloomberg's meta-model enables CDOs to get their data governance programs delivering measurable results faster.

Depth of coverage

Bloomberg data coverage is unparalleled in the industry:

- Acquire bulk or per-security data with maximum granularity and more than 15,000 fields
- Our global coverage includes more than 10 million active instruments and a broad data model of nearly 5,000 fields

- Bloomberg's Corporate Action data contains more than 50 action types across capital changes, distributions, corporate events and fixed income-specific actions
- Bloomberg's Corporate Structure data provides basic reference data and hierarchical relationships (parent, subsidiary, ultimate parent) across nearly 4 million legal entities
- Bloomberg's Compliance data provides information about 20 sanctions programs at the legal-entity level across 10 jurisdictions worldwide
- Bloomberg's Capital Structure product covers approximately 2,700 companies and more than 25,000 unique entities

Example Metadata for a Field: Realized Equity Per Share	
Field ID	RR691
CleanName	realizedEquityPerShare
URI	http://ontology.bloomberg.com/fields/realizedEquityPerShare
Description	Realized Equity Per Share
Data License	Security Master
Category	Fundamentals – Financial Ratios
Definition	BANKS, FINANCIALS, & INSURANCE Realized equity attributed to each share outstanding. Figure is scaled in millions; the Scaling Format Override (DY339, SCALING_FORMAT) can be used to change the display units for the field. Calculated as: Realized Equity/Shares Outstanding Where: Realized Equity is RR690, REALIZED_EQUITY Shares Outstanding is BS081, BS_SH_OUT
Asset Classes	Corp Equity Pfd
xsd:type	decimal
xsd:fractionDigits	3

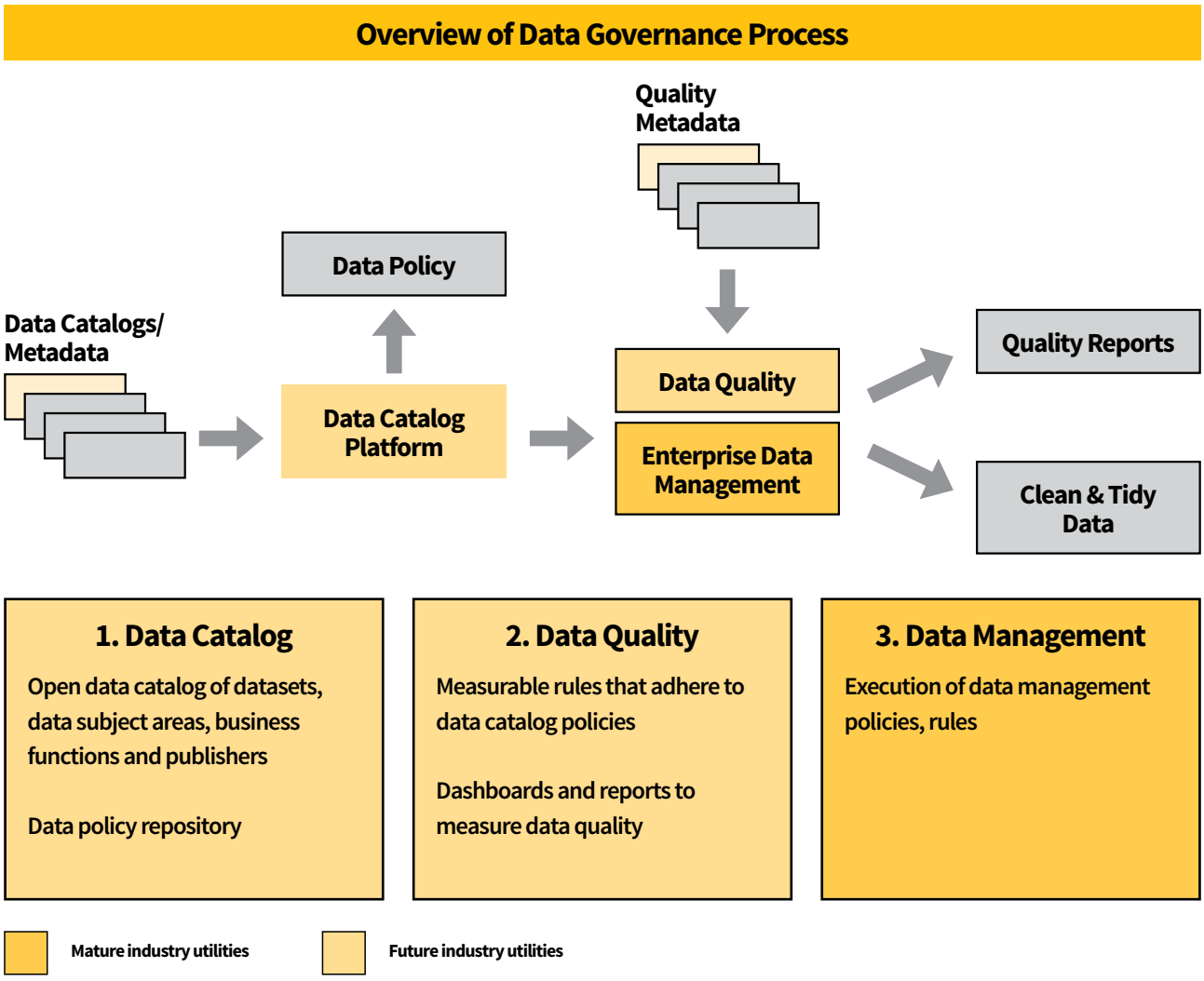
Example Metadata for a Dataset: Field Layout for corp_pfd_asia Family DatasetsDataset Family	**Field**	**Sequence**
corp_pfd_asia	TICKER	1
corp_pfd_asia	CPN	2
corp_pfd_asia	MATURITY	3
corp_pfd_asia	SERIES	4
…	…	…
corp_pfd_asia	ID_BB_GLOBAL_COMPANY_NAME	879

Call to Action

Listen to a recording of our recent webinar on "The missing piece of the data governance puzzle" where we delve deeper into methods for succeeding in data governance.

To learn more about how to speed up Data Governance with Bloomberg, visit **bloomberg.com/enterprise** or reach us at eprise@bloomberg.net.

Making Data Governance Real Relies on a Sequence of Three Underlying Solutions



About A-Team Group

A-Team Group provides news and analysis, white papers, webinars, events and more through our two online communities:

- Data Management Review
  www.datamanagementreview.com
- Intelligent Trading Technology
  www.intelligenttradingtechnology.com

Sign up as a member free, download recent white papers or look at our upcoming webinars and events and book your place today.

If you're a vendor and looking for high-quality content – like this white paper – to help articulate your message, take a look at **www.a-teamgroup.com**.

Or get in touch: **020 8090 2055** / **theteam@a-teamgroup.com**.

Data Management Review

Intelligent Trading Technology

A-TEAMGROUP

About Bloomberg for Enterprise

To succeed today, financial institutions must respond to challenges that are not addressed by traditional approaches. They require world-class solutions that integrate people, processes, information and technology for the front, middle and back office. Bloomberg partners with these institutions to protect and capitalize on data, manage risk, deliver transparency and control costs. Through enterprise-level expertise and three decades of deep industry experience, Bloomberg creates real value through the use of innovative technology that turns data into a strategic asset.

Get more information at **www.bloomberg.com/enterprise**

Bloomberg