GOVERNANCE

RETHINKING DATA GOVERNANCE AND MANAGEMENT

A Practical Approach for Data-Driven Enterprises

ISACA®

CONTENTS

ABSTRACT

It is never too late for any enterprise to start a data governance program, but it cannot be successful without a sound implementation strategy that is fully aligned with the enterprise business objectives.

This white paper uses a hypothetical use case to propose an approach for data governance and standardization, and improved strategic planning and decision making for data-driven digital transformation.

This white paper offers focused guidance on the following topics:

- Data governance foundation
- Data architecture
- Data quality and data cleansing
- Data democratization
- Data analytics

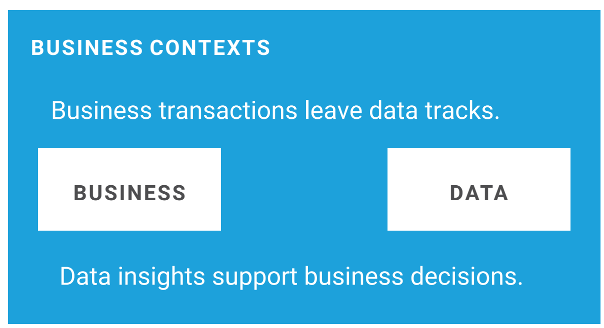
The Business and Its Data

Many enterprises strive to realize the potential value of the data they possess or that can legally be obtained from outside sources. For example, digital retailers use data for precision marketing, and city management uses data to dynamically optimize traffic controls. Harnessing data effectively can bring new value to businesses in the form of improved strategic planning and decision making.

According to McKinsey & Company, data governance is a key enabler when an enterprise begins the kind of data transformation that is necessary to use big data and advanced analytics for competitive differentiation.[1] McKinsey & Company states, "Before embarking on digital transformation, data governance needs to be developed and embedded throughout the process. Leaders need to clarify the policies and standards required to ensure effective data management, and they must define dedicated roles and responsibilities across the organization."[2]

It is meaningful to rethink the relationship between the business and data in the context of digital transformation (**figure 1**). In the modern digital environment, all business transactions or activities leave data tracks. The method in which the business intends to use the data ultimately provides the mechanism to describe the purpose of the data. Conversely, the data created while conducting business has its own intended business purpose, which supports decision making. In China, a recent hot topic is the business middle-platform and data middle-platform architecture approach to data governance, which was introduced by Alibaba.[3] This approach recognizes the important relationship that is established between two paradigms: the business paradigm, which can benefit from data insights, and the data paradigm, which accrues from and shapes the business.[4] This architecture allows Alibaba to gain data insights that powers the business to be a differentiator and helps to ensure that the business creates, analyzes and consumes premium-quality data.

FIGURE 1: Reciprocal Relationship Between the Business and Data



The architecture approach is supported by the growing understanding of the important inter-relationship between business and data. This white paper does not include a detailed discussion on the definition and details of this architecture, but the concept is important to understand.

Data Governance: Common Challenges

Although the importance of data governance and management has been emphasized, many enterprises still face challenges in this area. Some common challenges include:

- Enterprises often cannot easily perceive the value of data governance, which results in a lack of management commitment, because the benefits are difficult to quantify.

[1] "Getting your data house in order.," July 2018, https://www.mckinsey.com/industries/consumer-packaged-goods/our-insights/digital-and-analytics-in-consumer/capabilities/getting-your-data-house-in-order
[2] *Ibid.*
[3] Alibaba is a leader in data governance practices globally and especially in China. They developed the business middle-platform and data middle-platform architecture approach. The business uses data and data must be managed in business context.
[4] Zhong Hua; **企业*IT*架构转型之道: 阿里巴巴中台战略思想与架构实战** *(Enterprise IT Architecture Transformation Approach: Alibaba's Strategy and Philosophy in Middle-platform and its Architecture Practices)*, Alibaba, China, 2017

- Data ownership is often not clearly defined due to the misconception that data management is technical work and, therefore, the responsibility of the IT department.
- Marketing or other departments can obtain the data needed by their business units without considering the costs or risk associated with using the data.
- Evolving regulatory requirements impact data governance. The number of data-related regulations is trending upward (e.g., California Consumer Privacy Act [CCPA], General Data Protection Regulation [GDPR] on privacy protection and China Cybersecurity Law on data localization). The lack of a clear data governance structure makes it harder to respond to emerging or changing international data laws.
- Enterprises often identify a need for workflow changes, which has an impact on enterprise architecture. Opportunities to standardize, interoperate and re-architect current processes and systems must be taken. Proper change management, data protection and executive sponsorship are necessary to make data governance a reality, particularly in heavily regulated environments.
- Siloed department and organizational structures result in disaggregated datasets and data analytics challenges.
- Data deluge can result in quality problems. Big data is of no value if it cannot be harnessed for better business decision making. Effective decisions cannot be made based on poor-quality data. In addition, more complex and unstructured data are being introduced, which can strain data governance efforts.
- Lack of skilled staff to conduct data analysis creates a barrier to effective data governance. A good working knowledge of data analytics becomes critical to the success of an enterprise. The best resources should be able to identify hidden patterns and unknown correlations, draw conclusions, make inferences and predict future trends.

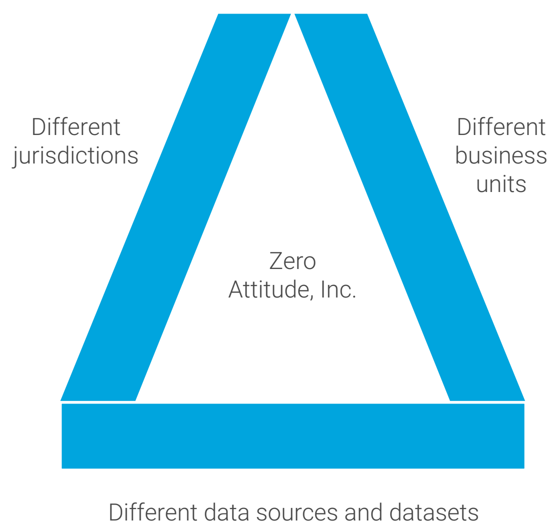
To address these challenges, successful enterprises can employ a phased, multistage approach to data management, with a well-designed data governance foundation, appropriately implemented data architecture, data quality and data cleansing efforts, and data analytics that improve decision making. In addition, strong data security and privacy processes bolster the overall data management approach by helping to reduce risk that is introduced when these challenges go unaddressed.

A Hypothetical Case Study

Zero Attitude, Inc. (ZA) has a global footprint and grew from acquisitions made over the past several years, which has resulted in a hybrid of system and data environments.[5] As a result, ZA operates in various jurisdictions and currently uses a variety of data sources and datasets.

ZA designs and produces consumer products (busines-to-consumer) and intermediate products to other enterprises (business-to-business). ZA defined aggressive digital transformation goals to grow its business. ZA executives have committed to the investment, knowing IT must play an important role in achieving these goals. One of the foundational objectives is to make better use of data to support business decision making. This case study focuses on helping ZA design and deploy an effective data governance and management program (**figure 2**).

FIGURE 2: A Hypothetical Data Governance Approach



[5] It was difficult to find a case that included all aspects of data governance, so this hypothetical case study is based upon several real-life enterprises.

A Practical Data Governance and Management Strategy

There is no one-size-fits-all strategy for developing a data governance program. For ZA, a two-phase process is proposed to achieve its objectives, using a practical data framework. First, ZA will establish a data governance framework and subsequently cleanse the data against quality standards. Next, ZA will implement a data management platform, which provides better insights for business decision making.

Five adoption stages are to be executed across the two phases (**figure 3**). As these stages are completed, they will serve to mature the program in each phase.[6] These five stages continuously turn raw data into business value.

Data security, privacy and operations (including requirements definition, data life cycle management, and internal/external data provider management) should be considered and aligned at each of the five stages. Since the scope of work with ZA is focused on establishing a foundation for data governance and data architecture to enable data analytics and data sciences, data security, privacy and operations are not covered in this case study.

Various data management frameworks or disciplines may have unique requirements for each of these stages, which can be tailored based on the enterprise (see Appendices B through D for more information). In addition, it can be difficult to depict the interdependencies of each stage linearly, so ZA executives are advised to expect these stages of adoption to follow a natural, sequential logic with some degree of interaction. For example, data governance cannot be successfully achieved if data quality requirements are not considered early.

FIGURE 3: Data Governance and Management Strategy: Five Stages of Adoption

Phase 1—Design and Deploy a Data Governance Strategy
Stage 1. Establish a data governance foundation to define beneficial data uses and the legal requirements of using that data.
Stage 2: Establish and evolve the data architecture.
Stage 3: Define, execute and assure data quality, and conduct data cleansing.
Phase 2—Implement a Data Management Platform
Stage 4: Realize data democratization.
Stage 5: Focus on data analytics.

Stage 1. Establish a Data Governance Foundation

A good data governance foundation sets the groundwork to collect and use data. This foundation includes addressing legal, business intellectual property and customer sensitivity considerations. Laws and regulations require that certain data are not collected or stored, and that data use and storage are well controlled. Customer sensitivity to collection and storage and difficulty in collection also need to be understood. For example, the enterprise can employ mechanisms that include culture, roles and responsibilities, organizational structure, skills and knowledge, and performance measures.[7]

The ZA executive leadership team expects that the following benefits will be achieved with the implementation of the data governance structure:

[6] Here we referred to Huawei and some other organizations' data governance practices: phase 1 – establish a data governance foundation, realize data cleansing, improve the accuracy of financial statements, interoperate with business flows; and phase 2 – implement a data pedestal, provide data services, support the digital transformation.

[7] ISACA; *COBIT® 2019 Framework: Governance and Management Objectives,* USA, 2019, https://www.isaca.org/COBIT/Pages/COBIT-2019-Framework-Governance-and-Management-Objectives.aspx

- Manage data as an asset and capture value from the data.
- Define data ownership, stewardship and roles and responsibilities.
- Implement data governance as a practice area that will mature over time.

The holistic data governance foundation to be designed for ZA should answer four questions:

1. What data does ZA have and need to use?
    - The data classification and data taxonomy need to be defined.
2. When are data governance practices taking place through the ZA data life cycle?
    - A consistent data life cycle model, mapped to data management activities, needs to be tailored to ZA.
3. Who is responsible for ZA governance?
    - A data governance structure, accountability and responsibility of ownership, stewardship, and custodianship roles need to be defined.
4. How will data be managed in ZA?
    - Data management policies, standards for implementation, and required technology and metrics need to be documented.

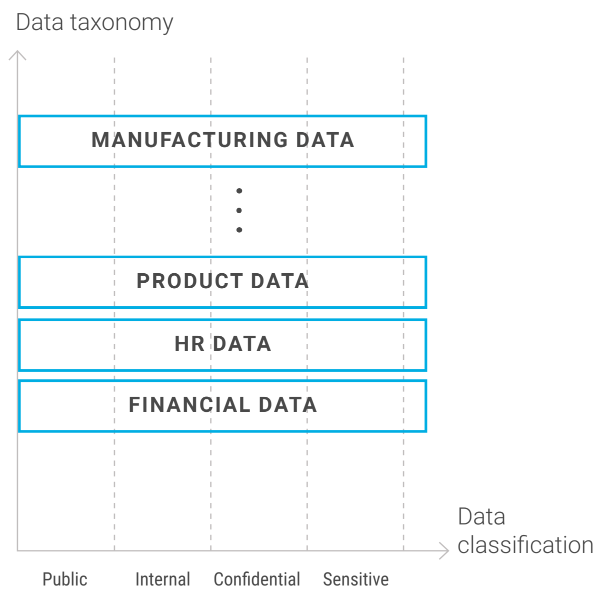
WHAT—Data Classification and Data Taxonomy

Defining the taxonomy, the amount of data collected and the classification of the information is extremely important for ZA to consider to ensure that it is in compliance with the applicable laws and regulations of the countries in which it operates. Data classification[8] is important from a security and privacy perspective, because it identifies sensitive and confidential data categories to promote effective data protection requirements. The data taxonomy provides enterprisewide data categories and hierarchy, which lay down a foundation for defining the business glossary and the data dictionary.[9]

According to GDPR, personal data refers to information relating to an identified or identifiable natural person (data subject). A natural person is an individual who can be identified, directly or indirectly, in particular by reference to an identifier, such as a name, identification number, location data, electronic identifiers or one or more specific elements of the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.[10] Data can be categorized using any taxonomy as long as it does not conflict with the principles of the GDPR (e.g., biometric, health and/or genetic data). Consideration should be made regarding the safety of the data, because the same data, after being processed will generate information that ZA will use to profile their customers.[11]

Because ZA is a multinational enterprise, compliance with *ISO/IEC 27001: Information Security Management*, *ISO/IEC 27002:2013: Information technology—Security techniques—Code of practice for information security controls* and *ISO 27701:2019: Security techniques—Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management—Requirements and guidelines* is advised to help ensure that ZA data are secure according to its taxonomy. GDPR requires data subjects to take the effective and necessary measures to ensure data security through its Article 32, Security in Treatment. An example of data classification and taxonomy is shown in **figure 4.**

FIGURE 4: Illustrative Data Classification/Taxonomy

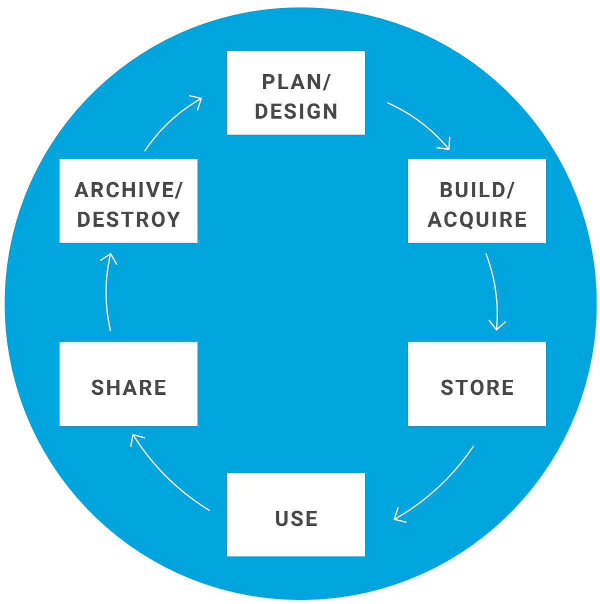


[8] Data classification is commonly known in information security management as the process of organizing data assets.
[9] Data scientists or academics may prefer "business terminology" or "data dictionary." Easy-to-understand terms were intentionally chosen for business background professionals and IT generalists.
[10] European Commission; "What is considered personal data under the EU GDPR?" https://gdpr.eu/eu-gdpr-personal-data/
[11] Additional resources that may assist ZA in ensuring the safety and correct use of data include NIST Special Publication 800-37, *Guide for Applying the Risk Management Framework to Federal Information Systems: A Security Life Cycle Approach,* https://csrc.nist.gov/publications/detail/sp/800-37/ rev-1/archive/2014-06-05, and NIST Special Publication 800-53: *National Vulnerability Database,* https://nvd.nist.gov/800-53.

WHEN—The Data Life Cycle and Mapping with Data Governance Activities

The data life cycle provides a high-level overview of the steps involved in the successful management and preservation of data for use/reuse. The proposed data life cycle for ZA is based on the model described in *COBIT 5: Enabling Processes* (**figure 5**).

FIGURE 5: The Data Life Cycle Model



Source: Adapted from ISACA, *COBIT 5: Enabling Information*, USA, 2013

ZA needs to standardize its data life cycle model across all business functions to ensure that a single taxonomy is consistently followed and defined, to create insights and value for the business. The activities vary across enterprises and cannot be easily elaborated using one common model. Enterprises need to define their own data governance activities and these activities must be addressed at each stage of the data life cycle (**figure 6**).

Figure 7 illustrates a possible mapping between ZA's data life cycle and key security/privacy management activities.

ZA should look at all stages of the data collection process and describe a model of the type of data that it is collecting. Security requirements are required to ensure data confidentiality, integrity and availability, according to data usage. For example, data used in application development should follow principles, such as privacy by design and by default. A framework that validates data from the point of its collection, transformation, analysis and reporting ensures that data do not contain information that can identify or be used against customers. To ensure confidentiality, development environments should mask production data during validation and apply privacy-by-design and privacy-by-default principles.

ZA must take the necessary measures to ensure data storage security, such as data encryption through hash tables, K-Anonymity systems and using encrypted and mostly separate databases using pointers (i.e., the name is not in the same table as the address, credit card number or tax document, etc.).

Subcontractors must also adopt security measures for the storage of data to help ensure that they are not accessed by or transferred to third parties. For guidance, ZA can refer to NIST 800-37, NIST 800-57, ISO 27001 and GDPR.

FIGURE 6: Data Life Cycle Mapping with Data Governance Activities

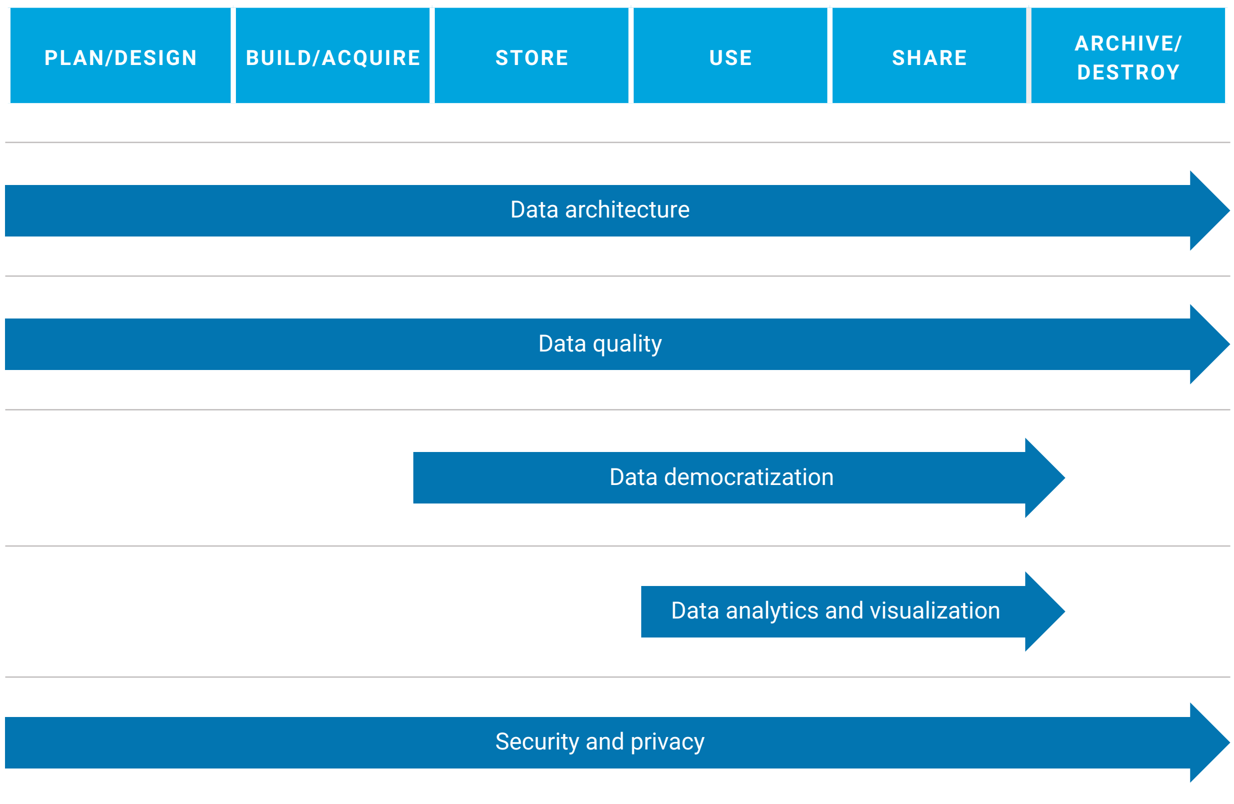


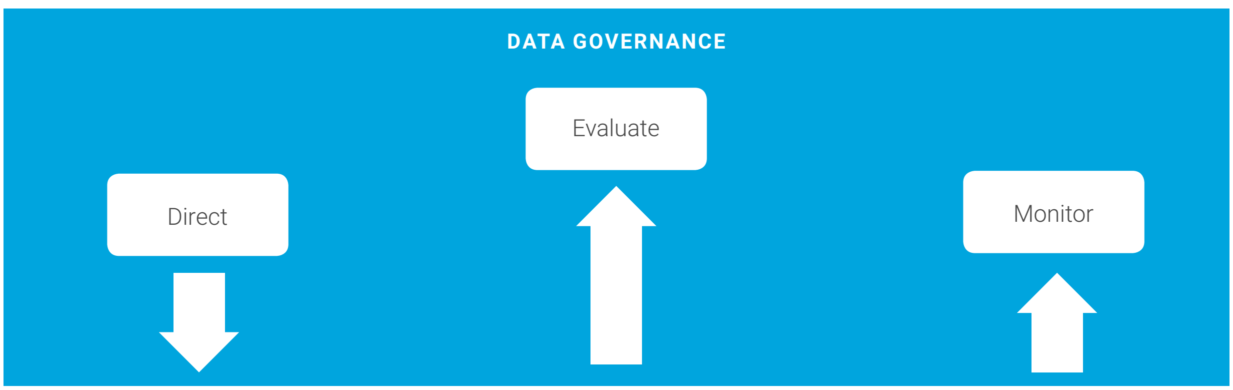
FIGURE 7: Data Life Cycle Mapping with Security/Privacy Management Activities

PLAN/DESIGN	BUILD/ACQUIRE	STORE	USE	SHARE	ARCHIVE/DESTROY
Identify security/privacy requirements	Validate data	Implement data protection	Monitor and control data access	Ensure compliance with laws/regulations	Data wiping
Identify data classification		Implement data-retention rules			
Design security/privacy framework					

WHO—The Data Governance Structure and Data Stewardship

The ZA data governance structure should be created to align business priorities across the enterprise to increase trust in data. Inconsistent business process definition and misalignment across various business units adversely impacts data. For ZA, a history of acquisitions resulted in inconsistent business processes and varying datasets. As a result, the ZA main business and IT department installed a joint data governance structure to review and approve data governance initiatives, and resolve outstanding issues, especially for data quality (**figure 8**). Then they used the data governance meetings/committees to prioritize data-related decisions. According to COBIT,[12] such a governance structure provides a regular meeting cadence for evaluating, directing and monitoring data governance activities. This governance structure ensures that ZA data assets are complete and accurate, and are compliant with ZA data governance policies, processes and applicable standards.

FIGURE 8: Data Governance and Data Management Model (Based on the COBIT Core Model)



DATA MANAGEMENT ACTIVITIES

APO14.01 Define and communicate the organization's data management strategy and roles and responsibilities.
APO14.02 Define and maintain a consistent business glossary.
APO14.03 Establish the processes and infrastructure for metadata management.
APO14.04 Define a data quality strategy.
APO14.05 Establish data profiling methodologies, processes and tools.
APO14.06 Ensure a data quality assessment approach.
APO14.07 Define the data cleansing approach.
APO14.08 Manage the life cycle of data assets.
APO14.09 Support data archiving and retention.
APO14.10 Manage data backup and restore arrangements.

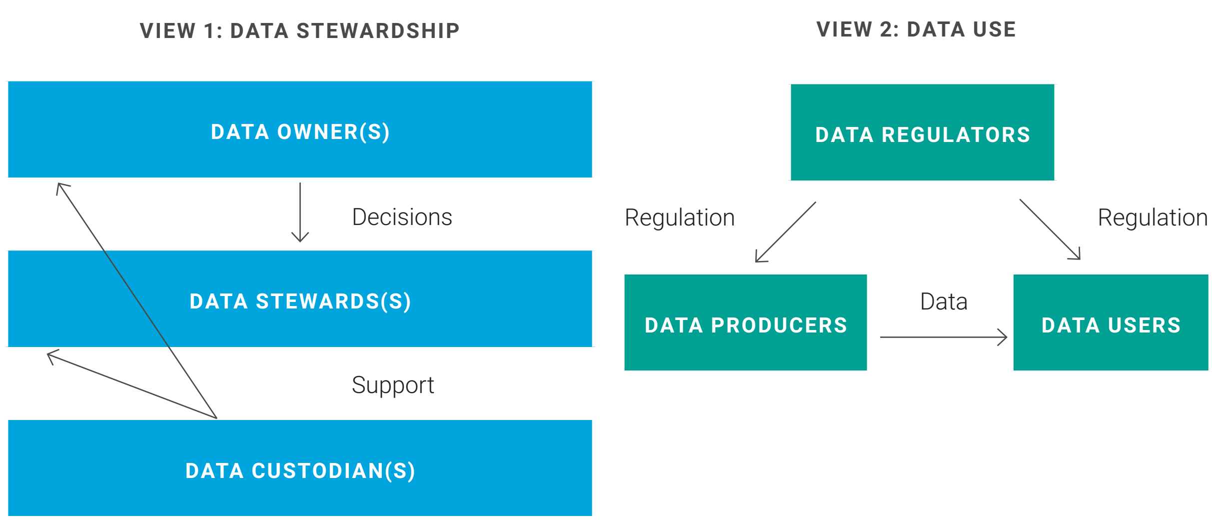
Next, ZA established a data stewardship structure that provided a central point of contact for data knowledge, visibility, accessibility and services. A good data stewardship structure defines roles and responsibilities for data management activities. An example of a suggested RACI chart and conceptual data stewardship structure are shown in **figure 9** and **figure 10**. The roles in the figures are not intended to be all-inclusive.

[12] *Op cit* ISACA, *COBIT 2019 Framework: Governance and Management Objectives*

FIGURE 9: Data Stewardship Structure–A Conceptual Model (Part 1)

ACTIVITIES	LIFE CYCLE PHASE	DATA OWNER	DATA STEWARD	DATA CUSTODIAN	...
Identify security/privacy requirements	Plan/Design	A	R	R	
Determine data classification	Plan/Design	A	R	I	
Design security/privacy framework	Plan/Design	A	C	R	
Validate data	Build/Acquire	C	A/R	C	
Implement data protection	Store	A	R	R	
Implement data retention rules	Store	A	C	R	
Monitor and control data access	Use	C	R	A/R	
Ensure compliance with laws and regulations	Share	A	R	R	
Data wiping	Archive/Destroy	I	C	A/R	

FIGURE 10: Data stewardship structure–A Conceptual Model (Part 2)



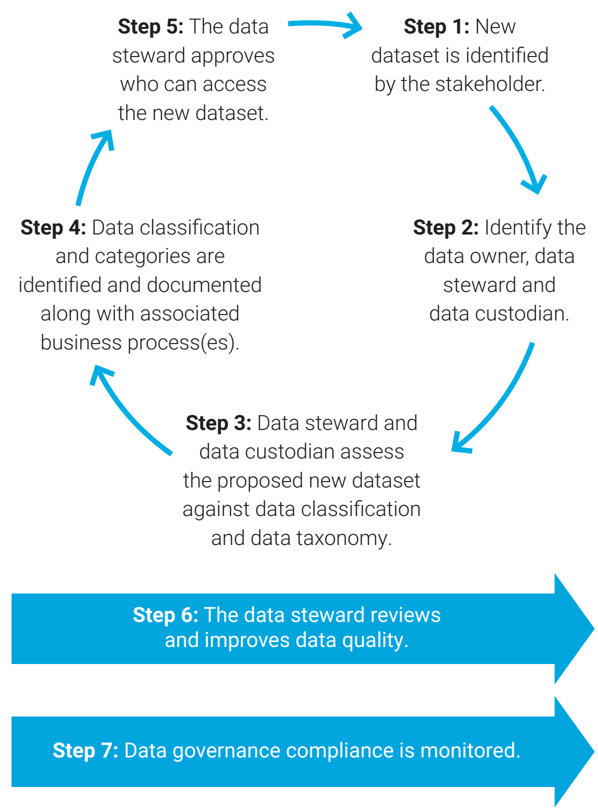
HOW—Data Governance Policies and Standards

A risk assessment is a helpful step to take for data governance and helps to identify the technology gaps, which address the security element in the data, as detailed in *COBIT 5: Enabling Information*.[13] It is also critical to identify the regulations and laws that require certain controls to be incorporated into the data governance approach, such as GDPR. After these are determined, ZA should look at the technology elements, the ability to collect data through the system or data warehouse, etc. The output of this analysis are data governance policies, which include data stewardship, data quality, security/privacy, data architecture, etc.

The data governance processes should be tightly based on IT tools. For example, marketing or other business units may introduce external datasets or data sources with dynamic data at ZA. In this situation, the following data stewardship steps (**figure 11**) should be taken:

- **Step 1**: The stakeholder (producer, user, or anyone else) identifies new datasets.
- **Step 2**: The data owner, data steward, and data custodian are identified.
- **Step 3**: The data steward and data custodian assess the proposed new dataset against the data classification and the data taxonomy.
- **Step 4**: Data classification and category are identified, documented and associated.
- **Step 5**: The data steward approves who can access the new dataset.
- **Step 6**: The data steward reviews and improves data quality.
- **Step 7**: Data governance compliance is monitored.

FIGURE 11: Data Stewardship Cycle–A Conceptional Model



Some key considerations to this governance process are:

1. Ensuring that taxonomy and classification align with those previously defined by the enterprise
2. Identifying the actors responsible for data storage, data processing and data classification
3. Storing and processing data access and classifying the new data according to the taxonomy defined by the enterprise
4. Allotting sufficient time for the enterprise to document the new data according to taxonomy and classification
5. Defining access rules according to NIST 800-53 and ISO 27001 and creating a properly documented CRUD (create, read, update and delete) matrix
6. Reviewing and correcting the data as necessary, so they can be put in place
7. Establishing governance, risk and compliance (GRC) processes

Stage 2. Establish and Evolve the Data Architecture

According to The Open Group Architecture Framework (TOGAF), data architecture is a description of the structure and interaction of the major types and sources of data, logical data assets, physical data assets and data management resources of the enterprise.[14] Stage 2 discusses the data standardization considerations that are important in the establishment and evolution of the data architecture.

Data Standardization Requirements

ZA has a lot of polluted data that may inhibit its ability to properly analyze the data. To cleanse the data, ZA must first standardize the data rules. This stage can be time consuming because there are many areas to consider when establishing data rules. Data rules should include data modeling, master data management, data dictionary, etc., but should be customized to meet the needs of the enterprise. Careful consideration must be given to this step, because it is cheaper to define good rules at the beginning than to have to make changes later. ZA must analyze this standardization effort from a cost-benefit perspective and should start with the most critical systems and datasets so as not to spend time in areas where value will not be derived.

Without a proper definition of the desired data architecture, it is unclear as to what ZA should do in regard to achieving systems, processes and data optimization and consolidation. Data cleansing terms are used here to highlight the focus on data itself. However, system re-architecting, integration and optimization should also be considered. According to TOGAF[15], there are five tactics to achieve architecture optimization:

- Re-platform: Migration of systems/data to virtualized or a cloud platform without significantly changing the business features and functions
- Refactor: Code optimization to improve the run-time efficiency of an application
- Replace: Provide a framework to replace existing systems/datasets with standard ones
- Retire: Decommission legacy systems, historical data from current architecture landscape
- Re-architect: Extracting business objects from current systems and using to develop new ones

ZA will see two main benefits after this new data architecture is established — standardized data management for the future and cleansed datasets for the historical data.

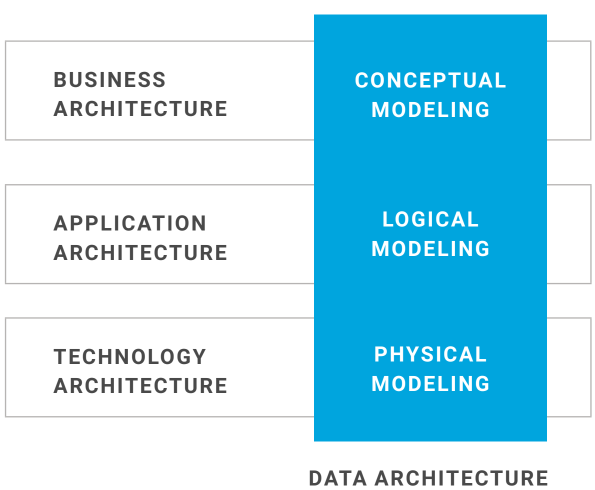
Data Models to be Standardized

Data objects encompass entities and attributes. They are different relationships among the entities. ZA uses data object populations to group different data objects, and it uses subject areas to group different entities. Data modeling schemes assume data object populations are strictly disjoined, which means an individual member of a data object population cannot be a member of another one. However, this is not always true; for example, a Person can be Employee, Customer and Stakeholder.

In terms of data modeling, ZA identifies three distinct levels (**figure 12**):

- **Conceptual**—This is the highest level, encompassing the enterprisewide or business functions view of ZA and the abstract model(s) that target the business architecture layer.
- **Logical**—This level adds details to the conceptual level, free of physical implementation details, which do not contribute to the use of the logical level, and targets the application architecture layer.
- **Physical**—This level addresses how data are going to be encoded and stored in a database (e.g., SQL and NoSQL), and deals with processing performance (denormalization of the database) and its partitioning and distribution (e.g., cloud storage), which targets the technology architecture layer.

FIGURE 12: Data Modeling Across Enterprise Architecture



ZA starts with conceptual modeling, which refers to the industrial data reference model for the data encoding standard. There is a range of choices for industries on the data reference model, such as the ARTS reference model[16] for the retail industry. The reference model provides a standard means by which data may be described, categorized and shared.[17] These are reflected within the three standardization areas of the data reference model:

1. Data context: Facilitates the discovery of data through an approach to the categorization of data according to taxonomies, and enables the definition of authoritative data assets within a community of interest.
2. Data description: Provides a means to uniformly describe data, thereby supporting its discovery and sharing.
3. Data sharing: Supports the access and exchange of data where access consists of *ad hoc* requests (such as a query of a data asset), and exchange consists of fixed, re-occurring transactions between parties.

Data sharing is enabled by capabilities provided by data context and data description. For details of data modeling and design, refer to the Data Management Body of Knowledge V2.[18]

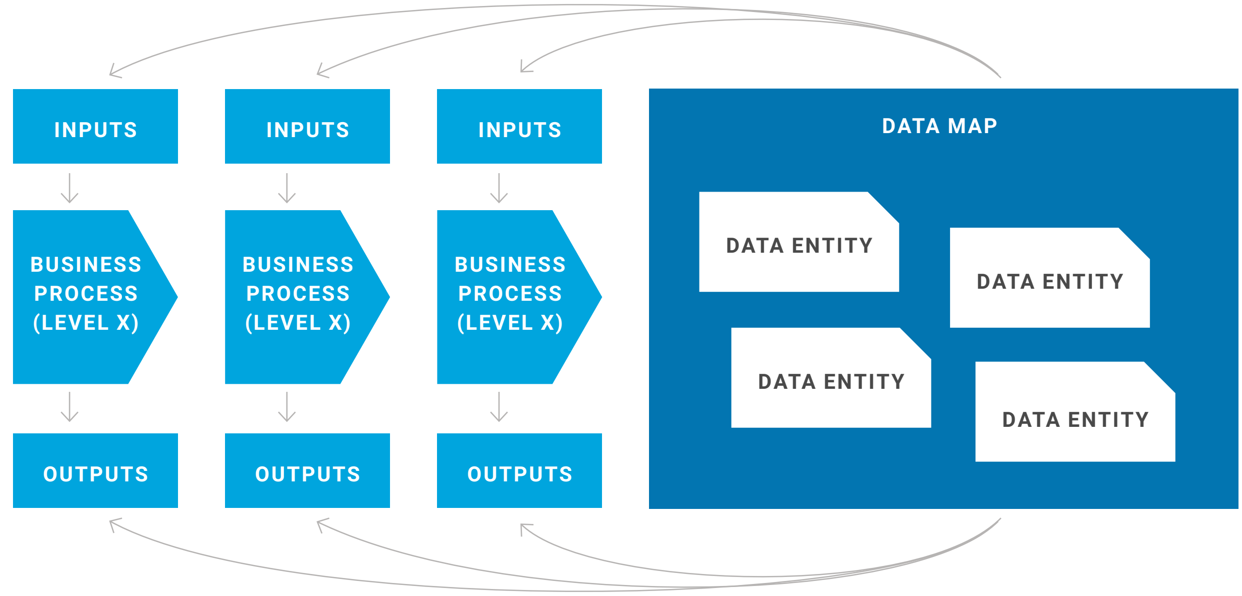
For ZA, the inputs to data modeling are (**figure 13**):

- As-is and To-be business processes
- Redundant data objects existing among current business applications
- Data objects from manual business processing without applications' support
- Data objects to be added for future business requirements
- Available common data reference model for ZA

The outputs from data modeling are:

- Data map (enterprisewide key data entities)
- The mapping of business process model (BPM) and data map
- Data models

FIGURE 13: Mapping Relationship Between the Business Process Model and the Data Map

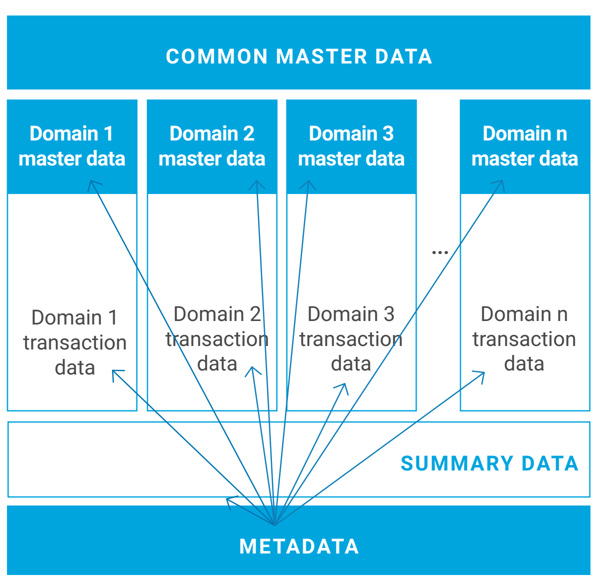


Data quality management entails the establishment and deployment of roles, responsibilities, policies and procedures concerning the acquisition, maintenance, dissemination and disposition of data.

Establish and Standardize Metadata and Master Data

The data map can be divided into five categories: common master data; domain master data; transaction data; summary data, which is dedicated for Business Intelligence (BI); and metadata (**figure 14**).

FIGURE 14: Data Map, Master Data and Metadata



Master data are the core data needed to uniquely define objects (e.g., Customer, Partner, Supplier, Person, Product or Service, Ledger and Asset). It is infrequently changed and is often referenced by business processes and associated with other data types. Data sharing is why the master data exists. **Metadata** are data that describe various facets of data assets to improve its usability throughout the data life cycle.

Benefits from implementing master data management include:

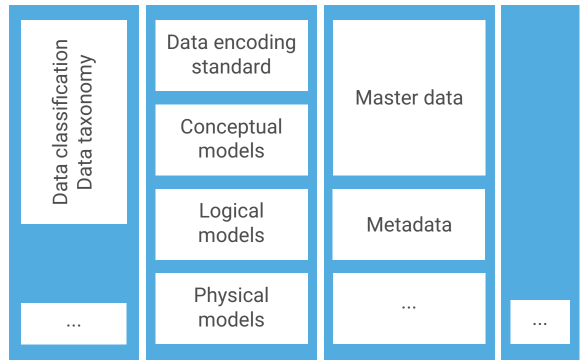
- Enhanced data sharing across business lines
- Centralized single source of key data for item definition and enrichment
- Improved item setup workflows, transparency in status and approval process
- Agility to support new business models
- Enhanced capabilities to support global growth
- Process consolidation to improve cycle time and go-to-market time
- Hierarchy management and association with items

Customer master data serve as the single source of key data about ZA's customers. It validates and enhances internal data with external sources to provide the most precise customer view. It assigns customers to the right customer hierarchy using the name provided on the customer record. Customer name inputs are validated against a repository of standardized company names furnished by trusted external data providers. It should also be checked against the up-to-date customer's legal entities.

Publish and Apply the Data Standards

The data dictionary is created based on already developed data taxonomy, business terminology and data models. It is a central repository where detailed data definitions can be found as the single source of trust. The data dictionary varies in its components across organizations, but it is commonly agreed that the data dictionary is the cornerstone for data standardization (**figure 15**).

FIGURE 15: Data Dictionary Structure



Stage 3. Define, Execute, Assure Data Quality and Clean Polluted Data

After the standards around data classification and taxonomy are established, ZA can build on that foundation by developing and executing a solid metadata strategy.

Good Metadata Strategy Leads to Good Data Quality

Metadata is closely related to data quality. Metadata summarizes basic data about data, which can make finding and working with particular instances of data easier. Author, date created, date modified and file size are examples of basic ZA document metadata. Having the ability to filter through that metadata makes it much easier for someone to locate a specific document.

A good metadata strategy includes the following:

- Understand metadata requirements
- Define the metadata architecture
- Define and maintain metadata standards
- Implement a managed metadata environment
- Create and maintain metadata
- Integrate metadata
- Manage metadata repositories
- Distribute metadata
- Query, report and analyze metadata
- Data sensitivity in metadata

In most cases, the metadata strategy focuses on the data warehouse or data lake because it is where the most data needs to be shared. The development of a data warehouse/data lake helps reduce the amount of metadata to be managed, which removes the complexity and narrows the work scope.

Define Data Quality Criteria

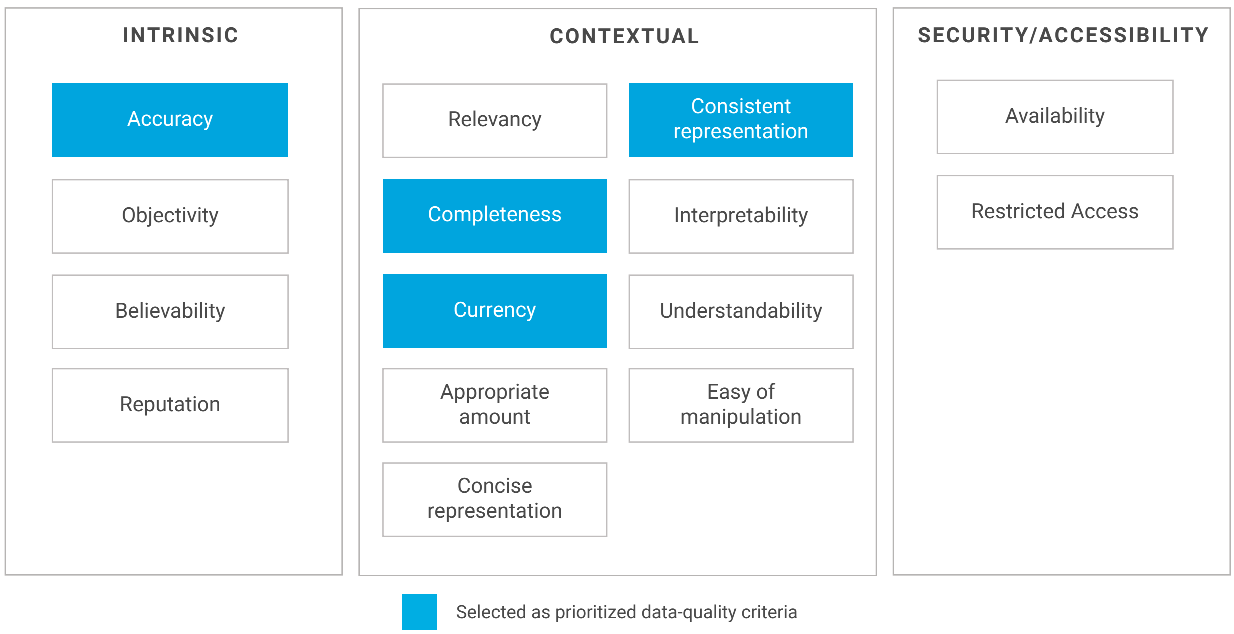
Following COBIT, a well-developed data quality model[19] defines three main quality criteria and 15 subcriteria. For ZA, four criteria are prioritized: accuracy, completeness, currency and consistent representation **(figure 16)**:

- Accuracy: Does the data reflect the dataset or data standard? Check for specified value(s) to examine the degree to which the data mirror the characteristics of the real-world object or objects it represents.

[19] We used aspects from both *COBIT® 5: Enabling Information* and COBIT 2019.

- Completeness: Are all datasets and data items recorded? Do the attributes associated with the measure contain the expected valid values? Are the attributes complete, meaning do they contain values or are there NULL values in the data set?
- Currency: Are the data available from the required point in time according to defined service level agreement?
- Consistent representation: Can the dataset be matched across data stores? Is a data attribute used consistently in the system of record, or can it have multiple meanings in the system of record? Are the data used as intended in external applications, or does is it have a different meaning externally?

FIGURE 16: Quality Model



Source: Adapted from ISACA, *COBIT 5: Enabling Information*, USA, 2013

Data quality criteria and the thresholds are selected based on the business context, requirements, levels of risk, etc. Each dimension is likely to have a different weighting in order to obtain an accurate data quality measure.

Execute Data Quality

ZA business and IT parties should jointly manage the data quality on an ongoing basis. The business parties are responsible for establishing the business rules that govern the data and are ultimately responsible for verifying the data quality. IT is responsible for establishing and managing the technical environment.

Key data quality roles and responsibilities include:

- Data steward: is responsible for managing data as a corporate asset.
- Business analyst: conveys the business requirements, including detailed data quality requirements. In addition, the business analyst reflects the requirements in the data models and the requirements for new dataset acquisition processes.
- Project manager: is responsible for the program or individual projects and ensures that the program/project is consistent with data quality requirements during system design, development and implementation. The project manager sets the tone with respect to data quality and interacts with data stewards to establish program/project level data quality requirements.

A successful data quality program has both proactive and reactive components. The proactive components diminish the potential for new problems to arise, and the reactive components address problems that already exist. That is why a regular meeting cadence should be set up to report and discuss data quality issues at the data stewardship level.

Regular Data Quality Assessment

Data quality issues can be identified and resolved as part of an ongoing assessment program. A regular assessment program helps to ensure that data clean up efforts stay evergreen. A typical data quality assessment approach includes the following steps:

1. Identify the data owner(s) for the data items.
2. Work with the data owner(s) to identify which data items are deemed critical and need to be assessed for data quality.
3. Establish business thresholds identified by the data owners/users.
4. Assess which data quality aspects to be used and their associated weighting.
5. Define data quality assurance rules.
6. Define values or ranges representing good and bad quality data, for each data quality aspect.
7. Define and agree on quality assessment results reporting.
8. Apply the assessment criteria to the data items.
9. Review the results and determine if data quality is acceptable or not based on the above-established business thresholds.
10. Take corrective actions, such as cleaning the data and improving data handling processes to prevent future recurrences.
11. Repeat the above steps on a periodic basis to monitor trends in data quality.
12. Make any needed update to business data requirements.

Ad hoc Data Quality Issue Management

To ensure that data quality issues do not negatively affect the data architecture, it is necessary to:

- Formalize an approach for identifying data quality expectations and defining data quality rules against which the data can be validated.
- Have individuals/teams identified to support the data quality measurement and improvement and define the RACI for ongoing maintenance of the quality.
- Baseline the levels of data quality and provide a mechanism to identify leakages as well as analyze root causes of data failures.

Work on Data Cleansing Against the Data Standard

After finalizing the data standard, the biggest challenge is full implementation of the data standard. A possible approach for data cleansing follows:

- At the business/data architecture layer, business processes and terminology need to be updated, reflecting the defined standards.
- At the application/data architecture layer, data encoding, data models, master data and metadata standards need to be built into application systems. Possible integration opportunities need to be reviewed and considered. Use extract, transform, and load (ETL) to integrate different data sources with the data warehouse or data lake to form a basis for data ingestion prepared for subsequent data analytics.
- At the technology/data architecture layer, ensure data storage, operations, and security/privacy are incorporated into re-architecting work.

Executive sponsorship and practical change management strategies throughout the ZA organization are critical to the success of the data cleansing efforts.

Stage 4. Realize Data Democratization

During the ZA implementation, there is a lack of visibility into where datasets exist, what data objects exist, who owns them and where they are located. Furthermore, it is difficult to understand how to access them. This gap reduces the business productivity and decision-making capability. Therefore, it is recommended that ZA create an enterprisewide platform, where permitted users can access the data. This effort is called data democratization, which facilitates the sharing of data and insights across the enterpreise, providing a single source of reference to search curated data and data-related expertise.[20] It can be achieved through a data catalog platform built in-house or support by a purchased commercial product.

ZA wants to use such a platform to achieve these objectives:

- Enable the data users to easily search but with proper access to the trusted data
- Enable a better understanding of the data in a related and more friendly data format

ZA uses the platform to show to its data users:

- Object name
- Title, Description
- Top users, Stewards
- Quality flags/ Trust check
- Tags, Article references
- Listing of sub-object pages with contextual information[21]

Individual data owners are established to create accountability for fulfilling best practices and maintaining data catalog content. Data can be searched, providing self-service capabilities. Access to the catalog and user experience is persona driven.

Data security and privacy is fundamental to data democratization. Data democratization relies heavily on the ability to secure data properly for end-user access. In some cases, users are not allowed to see certain data. Another aspect is securing rows of data in the database for specific user groups. This requires the development of reporting solutions that only allow the users to see specific rows of data based on their credentials.[22]

Stage 5. Focus on Data Analytics

Overview

Data analytics adds much value to the business; for example, a financial institution can leverage data analytics and data visualization in different capacities, from targeted marketing for financial products to detecting credit card fraud. Data analytics is used to examine data and apply statistical methods to identify hidden patterns and unknown correlations, draw conclusions and predict the likelihood of future events and trends. Data visualization is the graphical representation of data by using visual elements. There are various data analytics and data visualization tools available in the market for an enterpriseto assist in this area.

[20] Marr, B.; "What Is Data Democratization? A Super Simple Explanation and the Key Pros and Cons," *Forbes,* 24 July 2017, https://www.forbes.com/sites/bernardmarr/2017/07/24/what-is-data-democratization-a-super-simple-explanation-and-the-key-pros-and-cons/#646484476013
[21] Alation, https://www.alation.com
[22] Wilson, M.; "What to Consider When Building Your Data Democratization Strategy," *Ironside Group,* January 15, 2019, www.ironsidegroup.com/2019/01/15/what-to-consider-when-building-your-data-democratization-strategy/

The data analytics team works across multiple functions to accelerate business transformation powered by data-driven decision making. It is a problem-solving capability that combines business, applied math and technologies.

There are three types of data analytics: descriptive, diagnostic, predictive and prescriptive. Data analytics has an iterative cycle to seize the value. The main focus areas of data analytics are:

- Descriptive statistics: Enables data-driven decision making with interactive data services
- Statistical modeling and AI: Enables business outcomes through diagnostic, predictive and prescriptive analytics delivered at scale

Prototyping of Data Analytics Capability

Prototyping delivers a quick impact assessment proof-of-concept for targeted use cases. This enables measurable business impacts as well as scaled services to operationalize decision support tools that leverage and contribute to data analytics within the enterprise.

Develop Data Services for Business Purposes

Data services deliver analytical solutions that impact the critical needs of the enterprise. Standardization should be focused around delivering analytics at-scale using the latest analytical paradigms (e.g., natural language processing/natural language generation) across all consumption channels. This also included developing and deploying application programming interfaces (APIs) for other organization applications to consume analytics at-scale.

The data services can be requested through a process. The business is requested to identify their data service requirements. Then the data analytics team acknowledges the request and develops an appropriate solution possibly through a data warehouse or data lake.

Create Data Labels for Better Use of The Data

Data science drives diagnostic, predictive and prescriptive projects working with structured and unstructured datasets and leverages artificial intelligence (AI)/machine learning (ML) methods. It focuses on high-impact business problems that deliver measurable value to the enterprise and enable the acceleration towards a data-driven business.[23]

Typically, unlabeled data consists of samples of natural or human-created artifacts that can be obtained relatively easily. Labeled data typically takes a set of unlabeled data and augments each piece with some sort of meaningful label (or tag) that is somehow informative or desirable to know.

Labels are often obtained by asking humans to make judgments about a given piece of unlabeled data. After obtaining a labeled dataset, ML models can be applied to the dataset.

Data Visualization and Storytelling for Better Business Results

In business, a good story can grab people's attention and makes it easier to inspire, motivate or persuade them. Some considerations for using storytelling include:[24]

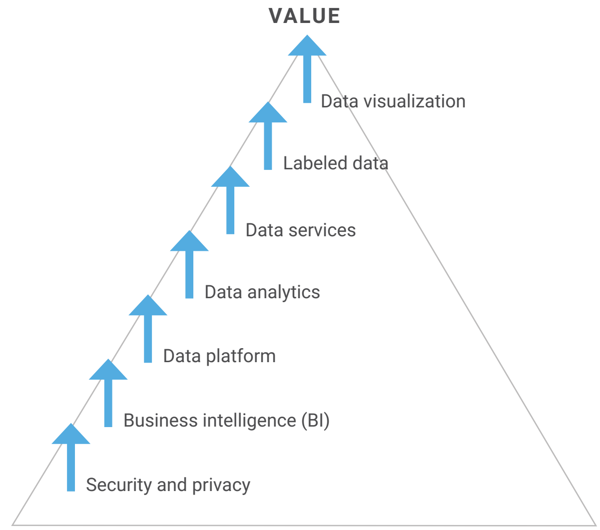
1. What is the business context? There are too much data, so the scope of data gathering needs to be narrowed down.
2. Know your audience: what do they really care about?
3. Construct a story from your data and make an emotional connection between your story and the audience.
4. Visualize the story.
5. Tell a story with real examples; make impacts specific to solving the audience's problem(s) using everyday language.

[23] Google Cloud, "AI Platform Data Labeling Service," 22 November 2019, https://cloud.google.com/data-labeling/docs/
[24] Qlik International AB, *5 Steps for Effective Data Storytelling,* 2017, https://www.qlik.com/us/-/media/files/resource-library/global-us/register/ebooks/eb-5-steps-for-effective-data-storytelling-en.pdf

Value Comes from Business Insights by Using the Data

Classifying the data in a business-oriented manner rather than via taxonomy can improve data value. However, it is important that there are sound criteria for classification, such as the mission criticality of the data. The value of the data will always be measured in accordance with ZA business context. IT has several ways of providing value to the business. **Figure 17** shows the simple model acknowledged in the ZA case.

FIGURE 17: Data Maturity Model for Value-Based Decision Making



Conclusion

Data are ubiquitous, making data governance a challenge. Adopting many of the best practices employed by the hypothetical enterprise highlighted in this white paper can contribute to the success of any enterprise data governance program. Data governance and data management are important for businesses that want to make use of data to create value for their stakeholders while also minimizing risk. For an enterprise to gain meaningful insights from data, strong data governance strategies and practices need to be in place.

Appendix A: Data Stewardship—Three Key Roles

Data Owner

A data owner has the responsibility and related authority to make decisions about data as well as its business definitions. A good data owner knows the data, knows the business, and understands the regulations and policies related to the data. A data owner is primarily concerned with value, risk, quality and utility of data. The data owner should be a senior-level business role.

At the data-owner level, the governance structure can be named the strategic data governance committee, which serves as the strategic level focal point with accountability for any given data subject. It is recommended to form this committee with senior stakeholders coming from both business units and IT.

Key data owner responsibilities include:

- Approves vision, objectives, strategies and data governance policies
- Defines, documents and communicates data governance policies
- Accountable for the data taxonomy and the definitions, data quality criteria, security and privacy
- Accountable for policy compliance of the data subjects
- Provides mandate and serves as the final authority in the escalation chain
- Provides organizationwide oversight of the data subjects
- Reviews and analyzes available reporting regarding compliance with data governance policies

Data Steward

A data steward is an individual or group who ensure data assets are used and adopted properly. They serve as the primary point of contact and understand the day-to-day use of a data domain as well as the value derived from the data. Stewards facilitate consensus about data definitions, quality and usage. Stewards guide the work needed to complete metadata, improve data quality, ensure regulatory compliance and ensure that data is fit for the specific business purpose. Stewards are also responsible for making recommendations about data access security, distribution and retention to data owners and custodians.

At the data steward level, the governance structure can be named the tactical data governance committee, which serves as the tactical level focal point with accountability and responsibility to drive process, data governance policy and DOC recommendations in the respective business unit and/or cross-functionally. A data steward should be nominated and named for prioritized data domains. These individuals are accountable for maintaining data quality and have the decision rights to help people enforce agreed-upon data governance policy. And these data stewards should also work collaboratively on cross-functional issues.

Key data steward responsibilities include:

- Provides governing body for organizationwide data subjects to the other stakeholders, including data producers and data users
- Aligns each business unit with organizationwide unified data governance framework
- Influences digital transformation and data services creation
- Reports data quality and data governance policy compliance
- Assesses security/privacy
- Identifies opportunities to improve data quality
- Provides change management of data governance policies
- Monitors and controls data governance by using metrics and providing feedback

Data Custodian

A data custodian is an individual or group who is responsible for ensuring the IT controls and safeguards for the data, and providing guidance and insight into the technical environment, the structure of the data and the architecture of the environment. Data custodians are the visible, action-oriented engine of an information governance effort. Data custodianship is ideally a technical role. It is the primary point of responsibility, accountability and activity for assessing, improving and evaluating our critical data sources. At data custodian level, the governance structure can be named the operational data governance committee, which serves as the operational level focal point with accountability and responsibility for IT tools.

Key data custodian responsibilities include:

- Implements IT capabilities with applications, data tools and technologies to support data governance policy
- Ensures the optimal and simplified IT architecture across the organization
- Ensures availability, continuity, capacity and performance levels and well-managed access
- Documents/logs all data handling activities
- Assists in diagnosing data related issues and inquiries

Appendix B: Mapping to COBIT 2019

Rethinking Data Governance and Management	COBIT 2019 Governance and Management Practices
1. Introduction	APO14 - Managed Data
2. Data Governance Foundation	APO14 - Managed Data EDM01 - Ensured Governance Framework Setting and Maintenance
3. Data Standardization	APO14 - Managed Data APO03 - Managed Enterprise Architecture
4. Data Quality	APO14 - Managed Data APO11 - Managed Quality
5. Data Democratization	BAI09 - Managed Assets
6. Data, Analytics & Visualization	APO04 - Managed Innovation
7. Conclusion	APO14 - Managed Data

Appendix C: Mapping to DMM 2.0

Rethinking Data Governance and Management	DMM 2.0 Process Areas
1. Introduction	
2. Data Governance Foundation	Data Management Function Governance Management Data Requirements Definition Data Life Cycle Management
3. Data Standardization	Business Glossary Meta-data Management Data Cleansing Architectural Approach Architectural Standards
4. Data Quality	Data Quality Strategy Data Quality Assessment
5. Data Democratization	Data Management Platform
6. Data, Analytics & Visualization	
7. Conclusion	Risk Management

Appendix D: Mapping to DAMA-DMBOK 2.0

Rethinking Data Governance and Management	DMBOK 2.0 Knowledge Areas
1. Introduction	Data Security
2. Data Governance Foundation	Data Governance
3. Data Standardization	Data Architecture Data Modeling & Design Reference & Master Data Meta-Data Data Integration & Interoperability
4. Data Quality	Data Quality
5. Data Democratization	
6. Data, Analytics & Visualization	
7. Conclusion	

Suggestions for Further Reading

ARTS, "ARTS Data Model Version 7.3," 1 December 2016, https://www.omg.org/retail-depository/arts-odm-73/

CEB Enterprise Architecture Leadership Council; *Data Governance: Step-By-Step Guide*, USA, 2015

CMMI Institute; *Data Management Maturity (DMM) Model*, USA, 2019

DAMA International, *Data Management Body of Knowledge (DMBOK) Version 2.* USA, 2017

"Data Democratization," www.techopedia.com/definition/32637/data-democratization

European Commission; "Data Protection in the EU," https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en

Eve, Bob; "Data Preparation Plus Data Governance Equals Better Analysis," *Cisco Blogs,* 30 November 2015, https://blogs.cisco.com/analytics-automation/data-preparation-plus-data-governance-equals-better-analysis

ISACA; China Construction Bank; *银行业IT治理最佳实践 (IT Governance Best Practice in Banking Industry)*, USA, 2019

Morgan, Lisa; "3 Data Governance Challenges Today's Companies Face," *Information Week,* 21 March 2017, www.informationweek.com/big-data/3-data-governance-challenges-todays-companies-face/a/d-id/1328449

Import. io, "Data Discovery Explained" December 31, 2019, https://www.import.io/post/data-discovery-explained/

Acknowledgments

ISACA would like to acknowledge:

Lead Developer

Guodong Zou
CGEIT, CRISC, CISA, CISM, CBRM, TOGAF Certified, EBDP, PMP, PgMP, PfMP
Manager & Sr. Consultant
Shanghai, China

Expert Reviewers

Mais Barouqa
CISA, CRISC, COBIT 5 FL, GRCP ITIL, ISO 27001 LA
Manager, IT Risk & Assurance
Deloitte & Touche, Jordan

Veronique Barrotteaux
PMP, CAMS, CSPO
Operational Risk Consultant
Wells Fargo, USA

Expert Reviewers *(cont.)*

Junlei Cai
CISA, CISM, CRISC, CGEIT, Cybersecurity Audit, CIPT, CIPM, CISSP, ISO 27001 LA
Vice President
China International Fund Management Co, Ltd., China

Daniel Ferreira
CISA, CISSP
CEO
PFCJ, Portugal

Ping (Jack) Gao, Ph.D.
Director of DMM Program, Data Scientist
Shanghai Data Exchange Corp, China

Ron Lear
CHMLA, CMQ/OE
Director of IP Development, Chief Architect
CMMI Institute, USA

Kevin Schaaff
CHMLA, PMP-ACP, ICP, LSSBB, CSQE
Principal Engineer
CMMI Institute, USA

Beverly Thomas
CISA, CITP, CPA, CGMA, ISO9001
Senior Manager
UMWA H&R, USA

Kevin Wegryn
PMP, Security+, PfMP
Technology Operational Risk Manager
Wells Fargo, USA

James Sun ZhenHua
CISA, CIPM, CIPT
Information Security and Compliance Manager
Universal Beijing Theme Park and Resort, China

About ISACA

For more than 50 years, ISACA® (www.isaca.org) has advanced the best talent, expertise and learning in technology. ISACA equips individuals with knowledge, credentials, education and community to progress their careers and transform their organizations, and enables enterprises to train and build quality teams. ISACA is a global professional association and learning organization that leverages the expertise of its 145,000 members who work in information security, governance, assurance, risk and privacy to drive innovation through technology. It has a presence in 188 countries, including more than 220 chapters worldwide.

DISCLAIMER

ISACA has designed and created *Rethinking Data Governance and Management: A Practical Approach for Data-Driven Organizations* (the "Work") primarily as an educational resource for professionals. ISACA makes no claim that use of any of the Work will assure a successful outcome. The Work should not be considered inclusive of all proper information, procedures and tests or exclusive of other information, procedures and tests that are reasonably directed to obtaining the same results. In determining the propriety of any specific information, procedure or test, professionals should apply their own professional judgment to the specific circumstances presented by the particular systems or information technology environment.

RESERVATION OF RIGHTS

ISACA®

1700 E. Golf Road, Suite 400
Schaumburg, IL 60173, USA

Phone: +1.847.660.5505

Fax: +1.847.253.1755

Support: support.isaca.org

Website: www.isaca.org

Provide Feedback:
www.isaca.org/rethinking-data-governance

Participate in the ISACA Online Forums:
https://engage.isaca.org/onlineforums

Twitter:
www.twitter.com/ISACANews

LinkedIn:
www.linkedin.com/company/isaca

Facebook:
www.facebook.com/ISACAGlobal

Instagram:
www.instagram.com/isacanews/