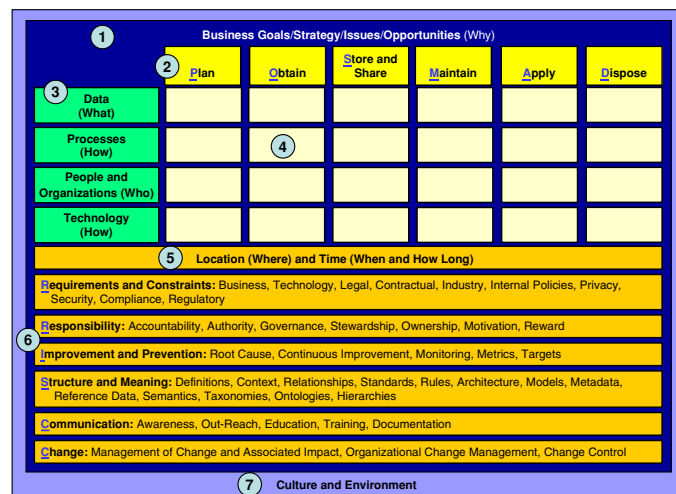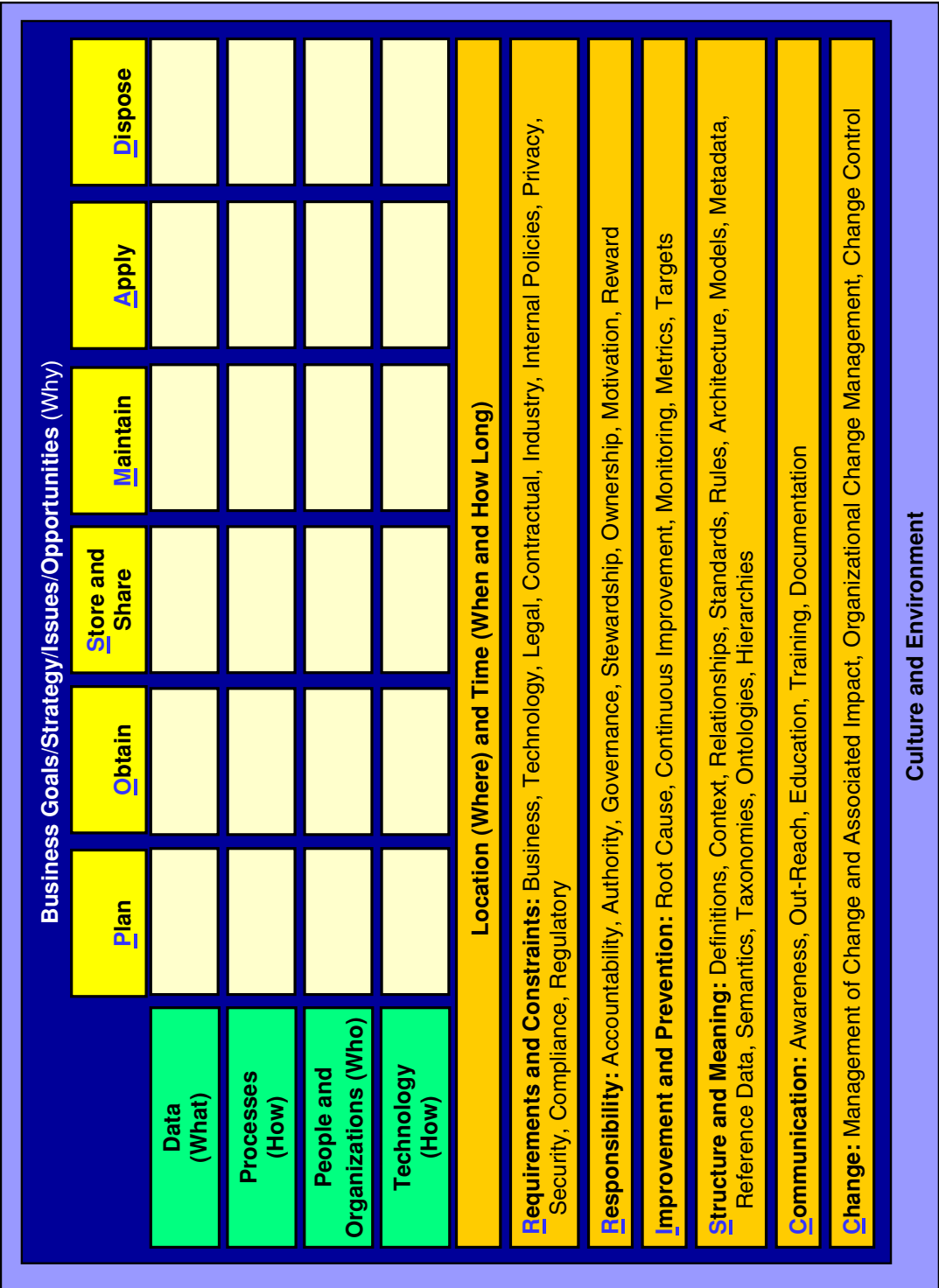# The Framework for Information Quality

1. **Business Goals/Strategy/Issues/Opportunities.** The "Why." Anything done with information should help the business meet its goals.

2. **Information Life Cycle.** Use POSMAD to help remember the information life cycle.
   - **Plan**—Identify objectives, plan information architecture, and develop standards and definitions; many activities associated with modeling, designing, and developing applications, databases, processes, organizations, and the like.
   - **Obtain**—Data or information is acquired in some way; for example, by creating records, purchasing data, or loading external files.
   - **Store and Share**—Data are stored and made available for use.
   - **Maintain**—Update, change, manipulate data; cleanse and transform data, match and merge records; and so forth.
   - **Apply**—Retrieve data; use information. Includes all information usage such as completing a transaction, writing a report, making a management decision, and completing automated processes.
   - **Dispose**—Archive information or delete data or records.

3. **Key Components.** Four key components affect information quality.
   - **Data (What)**—Known facts or other items of interest to the business.
   - **Processes (How)**—Functions, activities, actions, tasks, or procedures that touch the data or information (business processes, data management processes, processes external to the company, etc.).
   - **People and Organizations (Who)**—Organizations, teams, roles, responsibilities, or individuals.
   - **Technology (How)**—Forms, applications, databases, files, programs, code, or media that store, share, or manipulate the data, are involved with the processes, or are used by the people and organizations.

4. **Interaction Matrix.** Interaction between the Information Life Cycle phases (POSMAD) and the four Key Components.

5. **Location (Where) and Time (When and How Long)** *Note:* The top half of the framework, along with the first long bar, answers the interrogatives of who, what, how, why, where, when, and how long.

6. **Broad-Impact Components.** Additional factors that affect information quality. Lower your risk by ensuring that components have been discussed and appropriately addressed. If they are *not* addressed, you are still at risk (RRISCC) as far as information quality is concerned.
   - Requirements and Constraints
   - Responsibility
   - Improvement and Prevention
   - Structure and Meaning
   - Communication
   - Change

7. **Culture and Environment.** Take into account to better accomplish your goals.

Business Goals/Strategy/Issues/Opportunities (Why)

	Plan	Obtain	Store and Share	Maintain	Apply	Dispose
Data (What)						
Processes (How)						
People and Organizations (Who)						
Technology (How)						

Location (Where) and Time (When and How Long)

Requirements and Constraints: Business, Technology, Legal, Contractual, Industry, Internal Policies, Privacy, Security, Compliance, Regulatory

Responsibility: Accountability, Authority, Governance, Stewardship, Ownership, Motivation, Reward

Improvement and Prevention: Root Cause, Continuous Improvement, Monitoring, Metrics, Targets

Structure and Meaning: Definitions, Context, Relationships, Standards, Rules, Architecture, Models, Metadata, Reference Data, Semantics, Taxonomies, Ontologies, Hierarchies

Communication: Awareness, Out-Reach, Education, Training, Documentation

Change: Management of Change and Associated Impact, Organizational Change Management, Change Control

Culture and Environment

- The Framework for Information Quality (FIQ).

The POSMAD Interaction Matrix in Detail

The POSMAD Interaction Matrix is part of the Framework for Information Quality. This figure contains sample questions in each cell of the matrix to indicate the interaction between the phases of the POSMAD Information Life Cycle and the four Key Components—data, processes, people/organizations, and technology—that impact information quality.

	Plan	Obtain	Store and Share	Maintain	Apply	Dispose
Data (What)	What are the business objectives? Which data supports the business needs? What are the business rules? What are the data standards?	Which data are acquired? Which data are entered into the system—individual data elements or new records?	Which data are stored? Which data are shared? What is the key data to be backed up for rapid recovery?	Which data are updated and changed in the system? Which data will be transformed prior to migration, integration, or sharing? Which data are aggregated to support metrics or reporting?	What information is needed by the business to support transactions, metrics, compliance, requirements, decision making, automated processes, and other objectives? What information is available for use by the business?	Which data need to be archived? Which data need to be deleted?
Processes (How)	What are the high-level processes? What is the training and communication strategy?	How are data acquired from sources (internal and external)? How are data entered into the system? What are the triggers for creating new records?	What is the process for storing data? What is the process for sharing data?	How are data updated? How are data monitored to detect change? How are standards maintained? How are data change managed and impact assessed? Triggers for maintenance?	How are data used? How is information accessed and secured? How is information made available for those using it? What are the triggers for use?	How are data archived? How are data deleted? How are archive locations and processes managed? Triggers for archival? For disposition?
People and Organizations (Who)	Who identifies business objectives and assigns priorities and resources? Who develops processes, business rules, and standards? Who manages those involved in this phase?	Who acquires information from sources? Who enters new data and creates records in the system? Who manages those involved in this phase?	Who supports the storing technology? Who supports the sharing technology? Who manages those involved in this phase?	Who decides what should be updated? Who makes actual changes in the system? Who is responsible for quality? Who needs to know about changes? Who manages those involved in this phase?	Who directly accesses the data? Who uses the information? Who manages those involved in this phase?	Who sets the retention policy? Who archives the data? Who deletes the data? Who needs to be informed? Who manages those involved?
Technology (How)	What is the high-level architecture and the technology that support the business?	How is the application used to create new information and create records in the system?	What is the technology for storing the data? What is the technology for sharing the data?	How are data maintained and updated in the system?	How is information accessed to meet various business needs? How are business rules applied in the application architecture?	How are information and records deleted from system and/or archived? How are information and records archived from system?

● POSMAD interaction matrix detail—sample questions.

POSMAD Phases and Activities

The acronym POSMAD is used to help remember the six phases—Plan, Obtain, Store and Share, Maintain, Apply, Dispose—in the Information Life Cycle. This table describes the activities and provides examples of them within each of the life cycle's phases as they apply to information.

● POSMAD Information Life Cycle Phases and Activities

INFORMATION LIFE CYCLE* PHASE (POSMAD)	DEFINITION	EXAMPLE ACTIVITIES FOR INFORMATION
Plan	Prepare for the resource.	Identify objectives, plan information architecture, develop standards and definitions. When modeling, designing, and developing applications, databases, processes, organizations, etc., many activities could be considered part of the Plan phase for information.
Obtain	Acquire the resource.	Create records, purchase data, load external files, etc.
Store and Share	Hold information about the resource electronically or in hardcopy, and make it available for use through a distribution method.	Store data electronically in databases or some type of file, or store as hardcopy such as a paper application form. Share information about the resource through networks, an enterprise service bus, or email.
Maintain	Ensure that the resource continues to work properly.	Update, change, manipulate, parse, standardize, validate, or verify data; enhance or augment data; cleanse, scrub, or transform data; de-duplicate, link, or match records; merge or consolidate records, etc.
Apply	Use the resource to accomplish your goals.	Retrieve data; use information. This includes all information usage: completing a transaction, writing a report, making a management decision from information in those reports, running automated processes, etc.
Dispose	Discard the resource when it is no longer of use.	Archive information; delete data or records.

*Note: The Information Life Cycle may also be referred to as the Information Resource Life Cycle, the Data Life Cycle, the Information Value Chain, or the Information Chain.

Data Quality Dimensions

A Data Quality Dimension is an aspect or feature of information and a way to classify information and data quality needs. Dimensions are used to define, measure, and manage the quality of the data and information. This table contains a quick reference list of the 12 data quality dimensions used in The Ten Steps process.
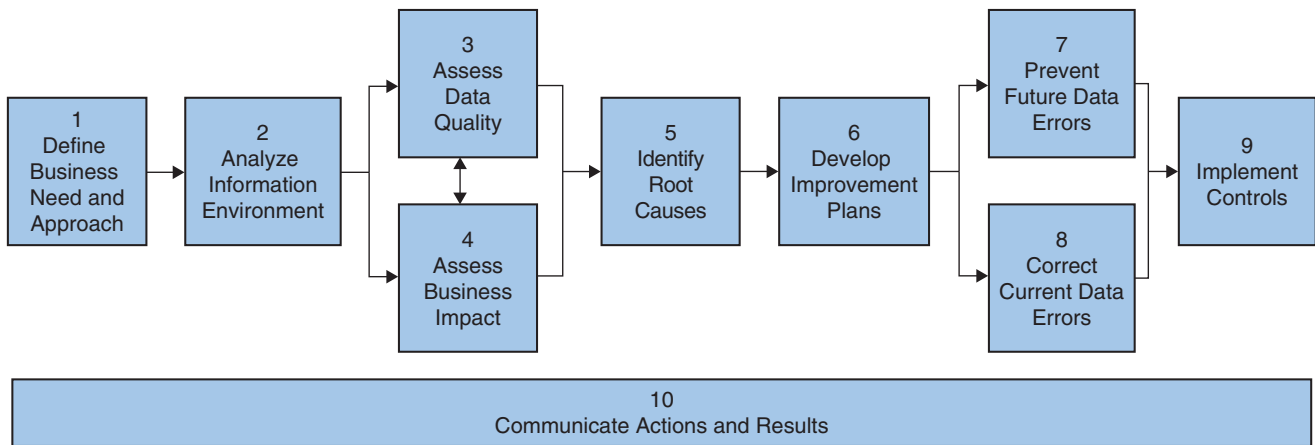
● Data Quality Dimensions

No.	Dimension	Definition
1	Data Specifications	A measure of the existence, completeness, quality, and documentation of data standards, data models, business rules, metadata, and reference data
2	Data Integrity Fundamentals	A measure of the existence, validity, structure, content, and other basic characteristics of the data
3	Duplication	A measure of unwanted duplication existing within or across systems for a particular field, record, or data set
4	Accuracy	A measure of the correctness of the content of the data (which requires an authoritative source of reference to be identified and accessible)
5	Consistency and Synchronization	A measure of the equivalence of information stored or used in various data stores, applications, and systems, and the processes for making data equivalent
6	Timeliness and Availability	A measure of the degree to which data are current and available for use as specified and in the time frame in which they are expected
7	Ease of Use and Maintainability	A measure of the degree to which data can be accessed and used and the degree to which data can be updated, maintained, and managed
8	Data Coverage	A measure of the availability and comprehensiveness of data compared to the total data universe or population of interest
9	Presentation Quality	A measure of how information is presented to and collected from those who utilize it. Format and appearance support appropriate use of information.
10	Perception, Relevance, and Trust	A measure of the perception of and confidence in the quality of the data; the importance, value, and relevance of the data to business needs
11	Data Decay	A measure of the rate of negative change to the data
12	Transactability	A measure of the degree to which data will produce the desired business transaction or outcome

Business Impact Techniques

Business Impact Techniques use qualitative and quantitative measures for determining the effects of data quality on the business. The table contains a quick reference list of the eight Business Impact Techniques used in the methodology—Ten Steps to Quality Data and Trusted Information™.

The figure below the table shows a continuum of the relative time and effort to determine business impact for each technique from generally less complex and taking less time (technique 1) to more complex and taking more time (technique 8).

● Business Impact Techniques

No.	Business Impact Technique	Definition
1	Anecdotes	Collect examples or stories about the impact of poor data quality.
2	Usage	Inventory the current and/or future uses of the data.
3	Five "Whys" for Business Impact	Ask "Why" five times to get to the real business impact.
4	Benefit versus Cost Matrix	Analyze and rate the relationship between benefits and costs of issues, recommendations, or improvements.
5	Ranking and Prioritization	Rank the impact of missing and incorrect data on specific business processes.
6	Process Impact	Illustrate the effects of poor-quality data on business processes.
7	Cost of Low-Quality Data	Quantify the costs and revenue impact of poor-quality data.
8	Cost–Benefit Analysis	Compare potential benefits of investing in data quality with anticipated costs, through an in-depth evaluation. Includes return on investment (ROI)*—profit from an investment as a percentage of the amount invested.

*The phrases ROI or return on investment are often used in a general sense to indicate any means of showing some type of return on an investment. ROI in technique 8 refers to the formula for calculating return on investment.

Less Time/Less Complex   →   More Time/More Complex

1 2 3   4   5 6   7   8

● Business impact techniques relative to time and effort.

Overview of The Ten Steps Process

The Ten Steps process is the approach for assessing, improving, and creating information and data quality. The steps are shown in the figure and described in the box.

The Ten Steps Process diagram:
1. Define Business Need and Approach → 2. Analyze Information Environment → 3. Assess Data Quality / 4. Assess Business Impact → 5. Identify Root Causes → 6. Develop Improvement Plans → 7. Prevent Future Data Errors / 8. Correct Current Data Errors → 9. Implement Controls
10. Communicate Actions and Results

The Ten Steps Process—Assessing, Improving, and Creating Information and Data Quality

1. **Define Business Need and Approach**—Define and agree on the issue, the opportunity, or the goal to guide all work done throughout the project. Refer to this step throughout the other steps in order to keep the goal at the forefront of all activities.

2. **Analyze Information Environment**—Gather, compile, and analyze information about the current situation and the information environment. Document and verify the information life cycle, which provides a basis for future steps, ensures that relevant data are being assessed, and helps discover root causes. Design the data capture and assessment plan.

3. **Assess Data Quality**—Evaluate data quality for the data quality dimensions applicable to the issue. The assessment results provide a basis for future steps, such as identifying root causes and needed improvements and data corrections.

4. **Assess Business Impact**—Using a variety of techniques, determine the impact of poor-quality data on the business. This step provides input to establish the business case for improvement, to gain support for information

quality, and to determine appropriate investments in your information resource.

5. **Identify Root Causes**—Identify and prioritize the true causes of the data quality problems and develop specific recommendations for addressing them.

6. **Develop Improvement Plans**—Finalize specific recommendations for action. Develop and execute improvement plans based on recommendations.

7. **Prevent Future Data Errors**—Implement solutions that address the root causes of the data quality problems.

8. **Correct Current Data Errors**—Implement steps to make appropriate data corrections.

9. **Implement Controls**—Monitor and verify the improvements that were implemented. Maintain improved results by standardizing, documenting, and continuously monitoring successful improvements.

10. **Communicate Actions and Results**—Document and communicate the results of quality tests, improvements made, and results of those improvements. Communication is so important that it is part of every step.

Definitions of Data Categories

Data categories are groupings of data with common characteristics or features. The table that starts here and continues on the next page includes definitions and examples for major data categories. These definitions were jointly created by Danette McGilvray and Gwen Thomas, president of the Data Governance Institute.

● Definitions of Data Categories

Data Category	Definition
Master Data	Master data describe the people, places, and things that are involved in an organization's business. *Examples* include people (e.g., customers, employees, vendors, suppliers), places (e.g., locations, sales territories, offices), and things (e.g., accounts, products, assets, document sets). Because these data tend to be used by multiple business processes and IT systems, standardizing master data formats and synchronizing values are critical for successful system integration. Master data tend to be grouped into master records, which may include associated reference data. An example of associated reference data is a state field within an address in a customer master record.
Transactional Data	Transactional data describe an internal or external event or transaction that takes place as an organization conducts its business. *Examples* include sales orders, invoices, purchase orders, shipping documents, passport applications, credit card payments, and insurance claims. These data are typically grouped into transactional records, which include associated master and reference data.
Reference Data	Reference data are sets of values or classification schemas that are referred to by systems, applications, data stores, processes, and reports, as well as by transactional and master records. *Examples* include lists of valid values, code lists, status codes, state abbreviations, demographic fields, flags, product types, gender, chart of accounts, and product hierarchy. Standardized reference data are key to data integration and interoperability and facilitate the sharing and reporting of information. Reference data may be used to differentiate one type of record from another for categorization and analysis, or they may be a significant fact such as country, which appears within a larger information set such as address. Organizations often create internal reference data to characterize or standardize their own information. Reference data sets are also defined by external groups, such as government or regulatory bodies, to be used by multiple organizations. For example, currency codes are defined and maintained by ISO.

(Continued)

DATA CATEGORY	DEFINITION
Metadata	Metadata literally means "data about data." Metadata label, describe, or characterize other data and make it easier to retrieve, interpret, or use information.
	Technical metadata are metadata used to describe technology and data structures. Examples of technical metadata are field names, length, type, lineage, and database table layouts.
	Business metadata describe the nontechnical aspects of data and their usage. Examples are field definitions, report names, headings in reports and on Web pages, application screen names, data quality statistics, and the parties accountable for data quality for a particular field. Some organizations would classify ETL (Extract–Transform–Load) transformations as business metadata.
	Audit trail metadata are a specific type of metadata, typically stored in a record and protected from alteration, that capture how, when, and by whom the data were created, accessed, updated, or deleted. Audit trail metadata are used for security, compliance, or forensic purposes. Examples include timestamp, creator, create date, and update date. Although audit trail metadata are typically stored in a record, technical metadata and business metadata are usually stored separately from the data they describe.
	These are the most common types of metadata, but it could be argued that there are other types of metadata that make it easier to retrieve, interpret, or use information. The label for any metadata may not be as important as the fact that it is being deliberately used to support data goals. Any discipline or activity that uses data is likely to have associated metadata.

Additional data categories that impact how systems and databases are designed and data are used:

Historical Data	Historical data contain significant facts, as of a certain point in time, that should not be altered except to correct an error. They are important to security and compliance. Operational systems can also contain history tables for reporting or analysis purposes. Examples include point-in-time reports, database snapshots, and version information.
Temporary Data	Temporary data are kept in memory to speed up processing. They are not viewed by humans and are used for technical purposes. Examples include a copy of a table that is created during a processing session to speed up lookups.