

CognitivePath™

Decoding the Data Discussion

A Glossary of Key Data Terms for AI Business Champions

In today's AI-driven business landscape, data is more than just a resource – it's a strategic asset. Yet, for many senior executives and line-of-business leaders, navigating data-related conversations with technology teams and vendors can feel like decoding a foreign language. This glossary is designed to bridge that gap, equipping you with a foundational understanding of key data concepts. With this knowledge, you'll be better prepared to engage in meaningful discussions, make informed decisions, and drive AI initiatives with confidence.

Data Strategy – A comprehensive plan that defines how an organization collects, manages, analyzes, and leverages data to achieve its business objectives while ensuring security, compliance, and governance.

Data Governance – A set of policies and procedures that identifies, protects and cultivates data as an asset that can be used by departments and individuals throughout the organization as their role dictates.

Data Management – The technical and operational processes for collecting, storing, organizing, and maintaining data throughout its lifecycle to ensure its accuracy, availability, security, and usability for business and analytical purposes.

Data Map – A visualization designed for business audiences that explains the flow of data between your organization's technology systems.

Data Lineage – Documentation of where data comes from, where it moves, and how it changes throughout systems. Like tracking the journey of member information from a web form through various databases.

Data Catalog – An inventory of data assets by system, which at a minimum includes data type, system(s) impacted, an estimate of data amount, metatag types associated with data, users, and more.

Metadata – Information that describes other data. Think of it as "data about data" - like tags on a file that tell you when it was created, who created it, what it contains, etc.

Data Classification – The process of categorizing data based on its sensitivity and importance. For example, marking certain member information as "Confidential" or "Public."

Master Data – Data that impacts, is used by or is of interest to multiple groups throughout the organization. For example, a member's profile data would be used by multiple groups.

Reference Data – Data that is used primarily by a singular line of business, for example a member's meal or session choice for a conference or meeting. This information would be

CognitivePath™

maintained by an individual system but could offer contextual information that may be useful to other departments if data is mined in an AI.

Data Quality – The measure of how accurate, complete, consistent, and reliable data is across systems. For example, having multiple versions of a member's email address across different systems would indicate poor data quality.

Role-Based Access – This provides a society employee access to data types based on their responsibilities. In essence, they get access only to the data that impacts their job. This is a security measure to protect data assets and systems.

Time-Based Access – Like role-based access, this security measure grants access to data only for the period of time a user needs to perform their task.

Data Steward – An employee responsible for maintaining and protecting specific types of data within their business area. They help ensure data is accurate, properly used, and follows governance policies.

Data Lifecycle – The stages data goes through from creation to deletion, including how it's collected, stored, used, archived, and eventually removed from systems.

Data Retention – Rules about how long different types of data should be kept before being archived or deleted, often based on legal requirements or business needs.

Data Integration – The process of combining data from different sources while maintaining its accuracy and usefulness. For example, merging member information from the website database with the email marketing system.

API or Application Programming Interface – A software application interface to access data and port it between systems.

Cloud – A network of remote servers hosted on the internet rather than on local computers. It's like storing your files in a secure digital warehouse instead of on your computer. Cloud storage allows authorized users to access data from anywhere while maintaining security protocols.

Data Lake – A large storage system that holds raw data in its original format until needed. Unlike a traditional database where data is organized before storage, a data lake is like a reservoir that collects all types of data (emails, documents, images, etc.) that can be analyzed later for different purposes.

Data Warehouse – A structured repository that stores current and historical data from various sources in an organized way for reporting and analysis. Unlike a data lake, data in a warehouse is processed and structured before storage, making it ready for specific business uses.

Data Mesh – A modern approach to managing data where instead of having one central team control all data, different business departments manage their own data while following shared standards. This allows teams to be more independent while still maintaining consistency across the organization.

CognitivePath™

Customer Data Platform (CDP) – A software system that collects, unifies, and organizes customer data from multiple sources to create a single, real-time customer profile that can be easily accessed and used across an organization.

Data Silo – When information is stored in isolated systems that don't communicate well with other systems, making it difficult for different departments to access or share important data. For example, when the CRM can't easily share information with the inventory management system.

Single Source of Truth (SSOT) – One authoritative, up-to-date version of data that everyone agrees to use. This prevents confusion from having different versions of the same information across various systems.

Personally Identifiable Information (PII) – Any data that could be used to identify a specific individual, such as name, email address, social security number, or phone number. This type of data requires special protection under various privacy laws.

Data Privacy – The practices and policies that ensure personal data is collected, used, and shared appropriately and with proper consent. This includes following regulations like GDPR or CCPA that protect individual privacy rights.

Data Cleansing – Fixing or removing incorrect, incomplete, duplicate, or improperly formatted data. For example, standardizing how addresses are written or removing duplicate member records.

Data Enrichment – Enhancing existing data by adding related information from other sources. For example, adding industry classification codes to member company records or appending demographic data to customer profiles.

About CognitivePath

CognitivePath is a specialist AI consultancy dedicated to helping enterprises, associations, and government agencies harness the full potential of artificial intelligence to enhance performance, drive innovation, and achieve key objectives. Our services include:

- AI Readiness, Strategy, and Roadmap Development
- AI Program Planning, Oversight, and Project Management
- Data Strategy and Governance
- Change Management, Executive Upskilling, and Workforce Training

Let's put your data work for your AI initiatives.
Contact us today at hello@cognitivepath.com or 1-877-222-9140.