

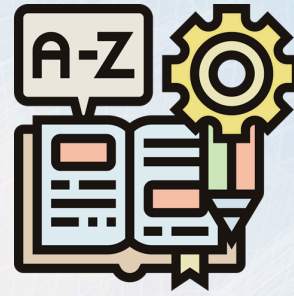
iTHRIV

Data Lexicon

Definitions of the most common data terms



Data Terms



Algorithms: Procedures for problem-solving or computation, providing step-by-step instructions for actions.

Attribute: Descriptors used to label columns in spreadsheets or databases for organization and analysis.

Changelog: Documenting steps taken when working with data for future reference.

Common Data Model: Standardizes structure and sometimes coding systems of disparate datasets for consistent analysis.

CSV (Comma-Separated Values) File: Text file separating data pieces with commas, commonly used for data analysis.

Dashboard: Tool for live data monitoring with visualizations connected to databases.

Data: Information like facts and numbers used for analysis and decision-making, computable by computers.

Data Bias: Inaccurate data due to omitted variables or human bias.

Data Governance: Formal plan for managing organizational data, including access, use, and compliance rules.

Data Science: Scientific study of data, involving questions, data transformation, analysis, and predictive modeling.

Data Source: Origin of specific information used for measuring business success and strategic decision-making.

Database: Organized collection of searchable, sortable, and updatable information stored electronically.

Numeric Files: Filing system arranging records by numbers.

Structured Query Language (SQL): Programming language for managing relational databases.

Text Files: Documents in plain text format, often denoted by .txt extension.

Data Management



Artificial Intelligence (AI): Computers can sometimes perform tasks as well as or better than humans, like identifying cells in a tissue sample.

Data Analytics: This involves collecting, organizing, and analyzing data to make informed decisions. It includes descriptive (what happened), diagnostic (why it happened), predictive (what will likely happen), and prescriptive (how to act) analytics.

Data Architecture: This is the plan for managing an organization's data, covering its lifecycle from collection to disposal.

Data Cleaning: Getting raw data ready for analysis by ensuring it's accurate, complete, consistent, and unbiased.

Data Engineering: Making data accessible for analysis by building systems to collect, manage, and convert raw data.

Data Enrichment: Adding more data to your existing dataset during analysis if needed.

Data Governance: Formal rules for how an organization manages its data, including access, use, accountability, and compliance.

Data Integrity: Maintaining data accuracy, reliability, and consistency over time.

Data Mining: Examining data closely to identify patterns and gain insights.

Data Visualization: Representing information using visual tools like charts and graphs to make data easier to understand.

Data Wrangling: Converting raw data into a usable form through stages like discovery, transformation, validation, and publishing.

Generative AI: Algorithms that create new text, images, etc., based on examples they've learned from.

Harmonizing data: Combining datasets to make them compatible.

Machine Learning (ML): Algorithms that predict outcomes and adapt to new data, like predicting surgery risks.

Query: A request for information from a database, formulated using a query language like SQL.

Types of Data



Aggregate data: Data that has been combined or averaged by geographic area, year, service agency, or other means.

Clean data: Accurate, complete, and ready-to-analyze data that has undergone thorough cleaning to address inaccuracies, inconsistencies, irregularities, and biases.

Data interpretation: Reviewing and drawing relevant conclusions from data using various analytical research methods.

Delimited data: Data organized in a flat file format where records are separated by a specified delimiter, such as a comma, and fields within each record are delineated.

Deidentified data: Data from which identifiers unique to individuals have been removed to ensure that it cannot be linked to a single person, particularly in health information.

Limited data: Protected health information (PHI) that doesn't have things like names or other personal details, to keep it private and safe, as defined by Privacy Regulations under HIPAA.

Metadata: Information about data, including how it was collected, stored, file type, or creation date, useful for verification and tracking purposes.

Missing data: Missing data occurs when information for a certain variable isn't recorded in an observation. It's a common issue in research and can strongly impact the conclusions drawn from the data.

Open data: Publicly available data that is accessible for anyone to use, enabling practice of data analysis skills.

Qualitative data: Descriptive data about qualities or characteristics, and subjective, like eye color or emotions and typically not numbers.

Quantitative data: Objective data is all about numbers, often countable or measurable, such as height or speed.

Raw data: Data that has not undergone thorough processing and remains in its original form.

Relational database: Database containing multiple tables with related information, allowing access to related data across tables with a single query.

Structured data: Formatted data organized into rows and columns. It's easy to work with because it's so organized.

Synthetic data: Artificially generated information rather than derived from real-world events.



Research Terms

Case study is like a story about a person or a group of people who have something in common, like a certain illness or experience. Researchers study these cases closely to learn more about the condition or situation and how it affects people.

Clinical trials are research studies that check if medical treatments are safe and work well for people. They test new drugs, therapies, devices, and procedures, gathering data on their effectiveness and side effects. These trials have different phases and follow strict rules to ensure safety and accuracy. They're important for developing new treatments or improving existing ones.

Community engaged research is collaborations between academics and non-academic groups to address community needs. They work together to do research and to meet the needs of those communities through programs and services.

Consent is voluntary agreement to participate in research or treatment. **Assent** is a minor or incapacitated individual's agreement to participate in research or treatment.

Efficacy measures how well a treatment works in ideal conditions, while **effectiveness** measures its real-world impact.

Electronic Health Record (EHR) (also called Electronic Medical Record) is an electronic version of a patient's medical history. They include all kinds of information, like diagnoses, procedures, lab results, vital signs, and medications.

HIPAA is a law protecting patient health information.

Institutional Review Boards (IRB) ensure research with human participants follows ethical guidelines. They review and approve research plans to protect the rights and welfare of the people involved.

Randomized trials assign participants randomly to different groups to ensure fairness.

Social drivers of health, or things that affect how healthy people are, are based on where they live, learn, work, and have fun. Social drivers of health are typically grouped in a few general categories including: income, education, healthcare access and quality, social and community connections, and neighborhood and home environment.

Vulnerable populations need special protections in research and healthcare, including minors, elderly, pregnant women, disabled, economically disadvantaged, minorities, and those with limited decision-making capacity.