CHRIST
(DEEMED TO BE UNIVERSITY)
BANGALORE, INDIA

# UNIT 3: RESEARCH DATA

Dr Sudhakar T
Assistant Professor
Department of Computer Science
CHRIST (Deemed to be University)
Bangalore

# Research Data Collection

- **Data collection** is a systematic process of gathering observations or measurements. Whether you are performing research for business, governmental or academic purposes, data collection allows you to gain first-hand knowledge and original insights into your research problem.

- Before you begin collecting data, you need to consider:

- The **aim** of the research
- The **type of data** that you will collect
- The **methods and procedures** you will use to collect, store, and process the data

- To collect high-quality data that is relevant to your purposes, follow these four steps.

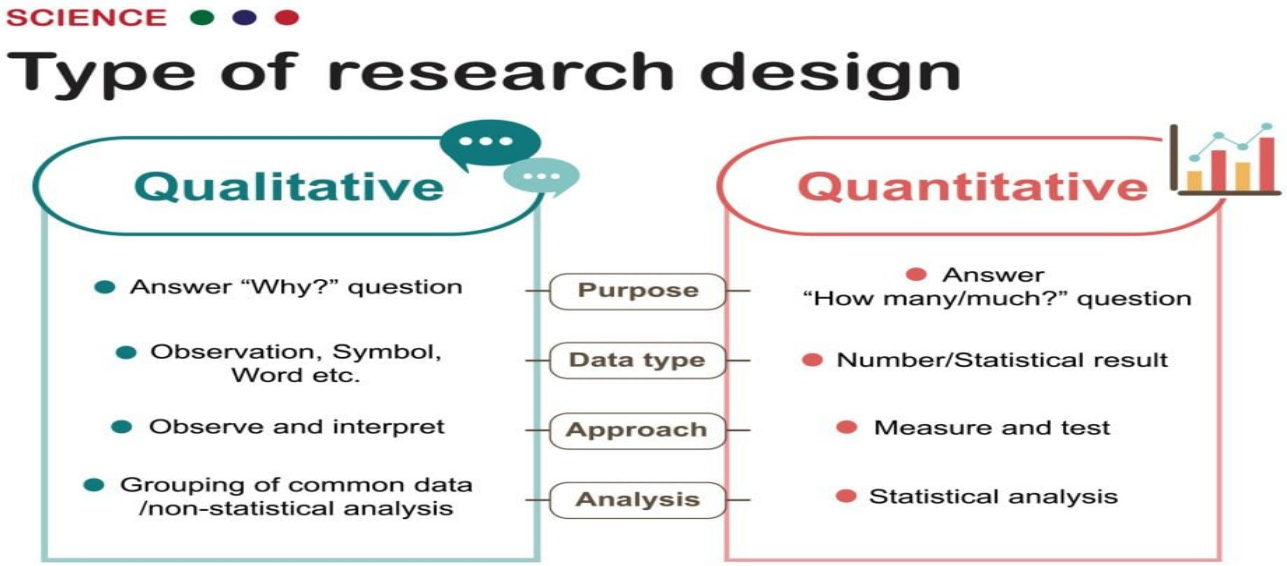
- **Step 1: Define the aim of your research**
- **Step 2: Choose your data collection method**
- **Step 3: Plan your data collection procedures**
- **Step 4: Collect the data**

# Step 1: Define the aim of your research

- Before you start the process of data collection, you need to identify exactly what you want to achieve.
- You can start by writing a problem statement: what is the practical or scientific issue that you want to address and why does it matter?
- Next, formulate one or more research questions that precisely define what you want to find out.
- Depending on your research questions, you might need to collect quantitative or qualitative data:

- **Quantitative data** is expressed in numbers and graphs and is analyzed through statistical methods.

- **Qualitative data** is expressed in words and analyzed through interpretations and categorizations.



- If your aim is to test a hypothesis, measure something precisely, or gain large-scale statistical insights, collect quantitative data.
- If your aim is to explore ideas, understand experiences, or gain detailed insights into a specific context, collect qualitative data.
- If you have several aims, you can use a mixed methods approach that collects both types of data.

# Step 2: Choose your data collection method

- Experimental research is primarily a quantitative method.
- Surveys, observations, archival research and secondary data collection can be quantitative or qualitative methods.
- Interviews, focus groups, and ethnographies are qualitative methods.

# Data collection methods

| Method | When to use | How to collect data |
|---|---|---|
| **Experiment** | To test a causal relationship. | Manipulate variables and measure their effects on others. |
| **Survey** | To understand the general characteristics or opinions of a group of people. | Distribute a list of questions to a sample online, in person or over-the-phone. |
| **Interview/focus group** | To gain an in-depth understanding of perceptions or opinions on a topic. | Verbally ask participants open-ended questions in individual interviews or focus group discussions. |
| **Observation** | To understand something in its natural setting. | Measure or survey a sample without trying to affect them. |
| **Ethnography** | To study the culture of a community or organization first-hand. | Join and participate in a community and record your observations and reflections. |
| **Archival research** | To understand current or historical events, conditions or practices. | Access manuscripts, documents or records from libraries, depositories or the internet. |

| | | |
|---|---|---|
| **Secondary data collection** | To analyze data from populations that you can't access first-hand. | Find existing datasets that have already been collected, from sources such as government agencies or research organizations. |

# Step 3: Plan your data collection procedures

## Operationalization

- Operationalization means turning abstract conceptual ideas into measurable observations.
- **Creating a data management plan**
- **Standardizing procedures**
- **Sampling**

# Step 4: Collect the data

To ensure that high quality data is recorded in a systematic way, here are some best practices:

- Record all relevant information as and when you obtain data. For example, note down whether or how lab equipment is recalibrated during an experimental study.
- Double-check manual data entry for errors.
- If you collect quantitative data, you can assess the reliability and validity to get an indication of your data quality.

# WHAT'S THE DIFFERENCE BETWEEN QUANTITATIVE AND QUALITATIVE DATA?

## Quantitative Data

- Countable or measurable, relating to numbers.
- Tells us how many, how much, or how often.
- Fixed and universal, "factual."
- Gathered by measuring and counting things.
- Analyzed using statistical analysis.

## Qualitative Data

- Descriptive, relating to words and language.
- Describes certain attributes, and helps us to understand the "why" or "how" behind certain behaviors.
- Dynamic and subjective, open to interpretation.
- Gathered through observations and interviews.
- Analyzed by grouping the data into meaningful themes or categories.

# What is quantitative data?

- Quantitative data refers to any information that can be quantified.

- If it can be **counted or measured**, and given a numerical value, it's quantitative data.
- Quantitative data can tell you **"how many," "how much," or "how often"**—for example, how many people attended last week's webinar? How much revenue did the company make in 2019? How often does a certain customer group use online banking?

- To analyze and make sense of quantitative data, you'll conduct **statistical analyses.**

# What is qualitative data?

- Unlike quantitative data, qualitative data cannot be measured or counted. It's **descriptive**, expressed in terms of language rather than numerical values.

- Researchers will often turn to qualitative data to answer "Why?" or "How?" questions. For example, if your quantitative data tells you that a certain website visitor abandoned their shopping cart three times in one week, you'd probably want to investigate why—and this might involve collecting some form of qualitative data from the user.

- Qualitative data also refers to the **words or labels used to describe certain characteristics** or traits—for example, describing the sky as blue or labeling a particular ice cream flavor as vanilla.

# Example

## Quantitative data:

- My best friend is 5 feet and 7 inches tall
- They have size 6 feet
- They weigh 63 kilograms
- My best friend has one older sibling and two younger siblings
- They have two cats
- My best friend lives twenty miles away from me
- They go swimming four times a week

## Qualitative data:

- My best friend has curly brown hair
- They have green eyes
- My best friend is funny, loud, and a good listener
- They can also be quite impatient and impulsive at times
- My best friend drives a red car
- They have a very friendly face and a contagious laugh

# Different types of quantitative and qualitative data

- **Discrete quantitative data** takes on fixed numerical values and cannot be broken down further.

- **Continuous quantitative data** can be placed on a continuum and infinitely broken down into smaller units.

- **examples of quantitative data:**
  - Measurements such as height, length, and weight
  - Counts, such as the number of website visitors, sales, or email sign-ups
  - Calculations, such as revenue

# Types of qualitative data

- **Nominal data** is used to label or categorize certain variables without giving them any type of quantitative value.

- **Ordinal data** is when the categories used to classify your qualitative data fall into a natural order or hierarchy.

- Example:
- Interview transcripts or audio recordings
- The text included in an email or social media post
- Product reviews and customer testimonials

# How are quantitative and qualitative data collected?

# How is quantitative data generated?

- Surveys and questionnaires
- Analytics tools
- Environmental sensors
- Manipulation of pre-existing quantitative data

# How is qualitative data generated?

- Interviews
- Surveys and questionnaires
- Observations

# Quantitative vs qualitative data: methods of analysis

# Quantitative data analysis

- Some of the most popular methods used by data analysts include:

- Regression analysis
- **Monte Carlo simulation**
- Factor analysis
- Cohort analysis
- Cluster analysis
- **Time series analysis**

# Qualitative data analysis

- With qualitative data analysis, the focus is on making sense of unstructured data (such as large bodies of text).
- Given that qualitative data cannot be measured objectively, it is open to subjective interpretation and therefore requires a different approach to analysis.
- Another type of analysis is **sentiment analysis**, which seeks to classify and interpret the emotions conveyed within textual data.

# When should I use qualitative or quantitative data?

- Generally, you can use the following criteria to determine whether to go with qualitative data, quantitative data, or a mixed methods approach to collecting data for your project.

- Do you want to understand something, such as a concept, experience, or opinions? Use qualitative data.

- Do you want to confirm or test something, such as a theory or hypothesis? Use quantitative data.

- Are you taking on research? You may benefit from a mixed methods approach to data collection.

# Measurement Scales in Research Methodology

## Measurement– types of scales

- Nominal scale
  - Frequency in each category
  - Percentage in each category
- Ordinal scale (rank order)
  - Median
  - Range
  - Percentile ranking
- Interval scale (rank order in equal intervals)
  - Mean
  - Standard deviation
  - Variance
- Ratio scale (arithmetic operations on actual quantities)
  - Geometric mean
  - Coefficient of variation

# Data Collection

- The process of gathering and analyzing accurate data from various sources to find answers to research problems, trends and probabilities, etc., to evaluate possible outcomes is Known as Data Collection.

- During data collection, the researchers must identify the data types, the sources of data, and what methods are being used.

Before an analyst begins collecting data, they must answer three questions first:

- What's the goal or purpose of this research?
- What kinds of data are they planning on gathering?
- What methods and procedures will be used to collect, store, and process the information?

# Why Do We Need Data Collection?

- You need data collection to help you make better choices.

# What Are the Different Methods of Data Collection?

1. Primary Data Collection

- Primary data collection involves the collection of original data directly from the source or through direct interaction with the respondents.

- Surveys and Questionnaires
- Interviews
- Observations
- Experiments

## 2. Secondary Data Collection

- Secondary data collection involves using existing data collected by someone else for a purpose different from the original intent. Researchers analyze and interpret this data to extract relevant information.

- Published Sources
- Online Databases
- Government and Institutional Records
- Publicly Available Data
- Past Research Studies

# Data Collection Tools

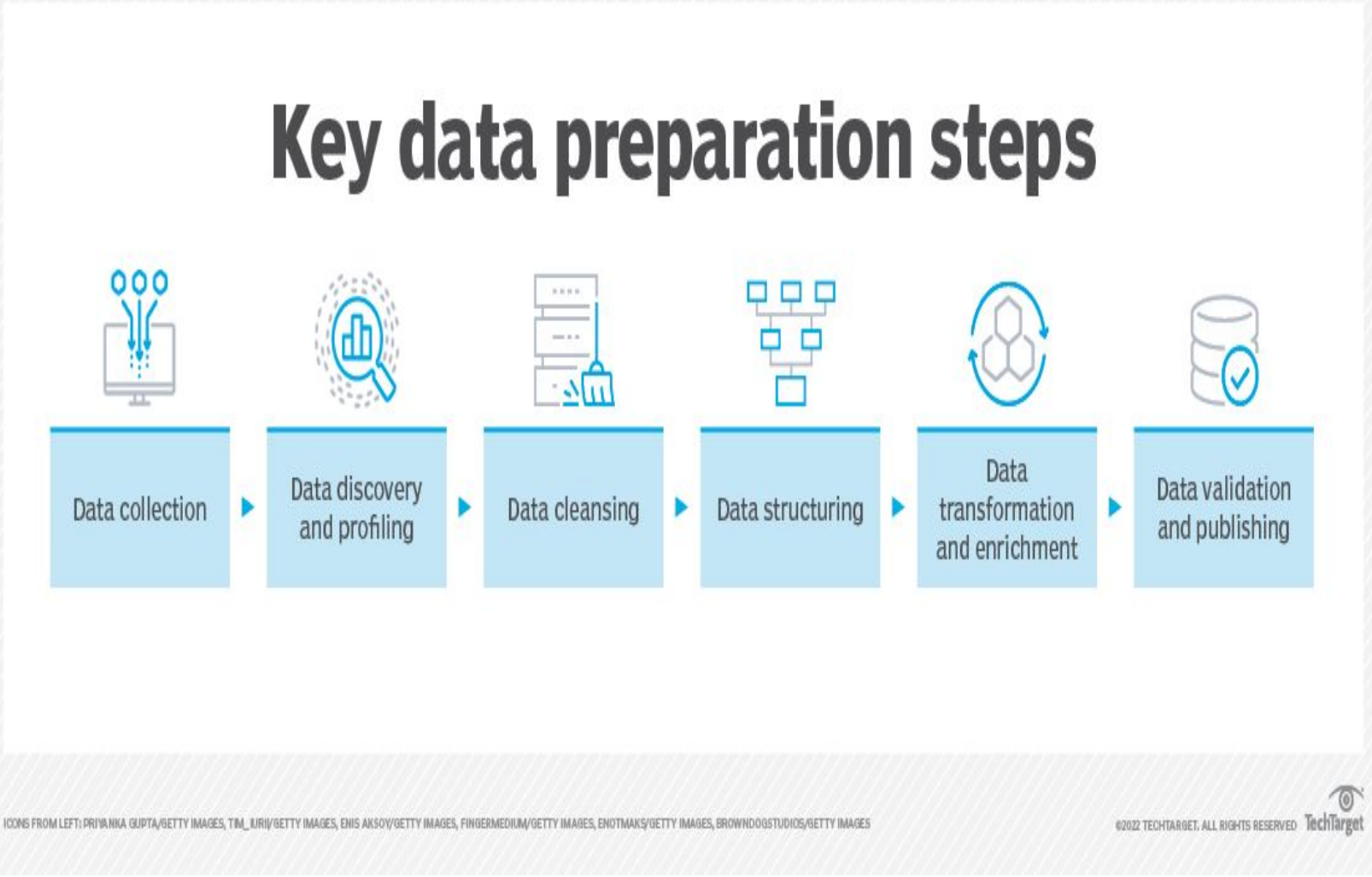
- Word Association
- In-Person Surveys
- Online/Web Surveys
- Mobile Surveys
- Observation

# What are Common Challenges in Data Collection?

- Data Quality Issues
- Inconsistent Data
- Data Downtime
- Ambiguous Data
- Duplicate Data
- Inaccurate Data
- Dealing With Big Data

# What are the Key Steps in the Data Collection Process?

1. Decide What Data You Want to Gather
2. Establish a Deadline for Data Collection
3. Select a Data Collection Approach
4. Gather Information
5. Examine the Information and Apply Your Findings


Key data preparation steps
Data collection
Data discovery and profiling
Data cleansing
Data structuring
Data transformation and enrichment
Data validation and publishing
ICONS FROM LEFT: PRIYANKA GUPTA/GETTY IMAGES, TIM_KIRIL/GETTY IMAGES, ENIS AKSOY/GETTY IMAGES, FINGERMEDIUM/GETTY IMAGES, ENOTMAKS/GETTY IMAGES, BROWNDOGSTUDIOS/GETTY IMAGES
©2022 TECHTARGET. ALL RIGHTS RESERVED
TechTarget


Publishing
Discovering
Validating
Structuring
Enriching
Cleaning

# Steps in the data preparation process

- **Data collection**
- Relevant data is gathered from operational systems, data warehouses, data lakes and other data sources.
- **Data discovery and profiling**
- The next step is to explore the collected data to better understand what it contains and what needs to be done to prepare it for the intended uses.
- To help with that, data profiling identifies patterns, relationships and other attributes in the data, as well as inconsistencies, anomalies, missing values and other issues so they can be addressed.

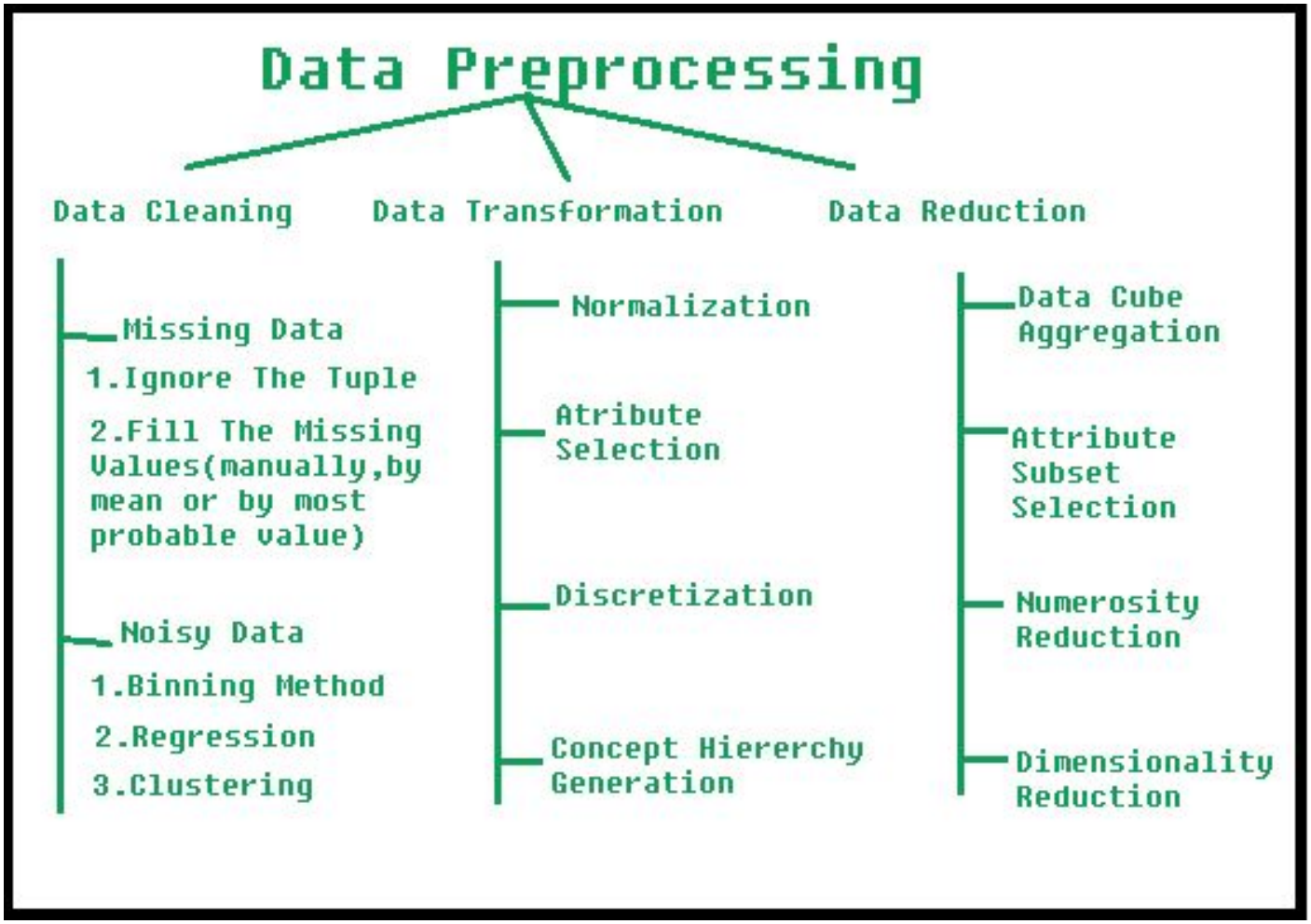
- Data cleansing

The identified data errors and issues are corrected to create complete and accurate data sets.

For example, as part of cleansing data sets, faulty data is removed or fixed, missing values are filled in and inconsistent entries are harmonized.

## Data structuring

At this point, the data needs to be modeled and organized to meet the analytics requirements. For example, data stored in comma-separated values (CSV) files or other file formats has to be converted into tables to make it accessible to BI and analytics tools.


Data Preprocessing
Data Cleaning
Data Transformation
Data Reduction
Missing Data
1.Ignore The Tuple
2.Fill The Missing Values(manually,by mean or by most probable value)
Noisy Data
1.Binning Method
2.Regression
3.Clustering
Normalization
Atribute Selection
Discretization
Concept Hierarchy Generation
Data Cube Aggregation
Attribute Subset Selection
Numerosity Reduction
Dimensionality Reduction

| person_name | Salary | Year_of_ experience | Expected Position Level |
|---|---|---|---|
| Aman | 100000 | 10 | 2 |
| Abhinav | 78000 | 7 | 4 |
| Ashutosh | 32000 | 5 | 8 |
| Dishi | 55000 | 6 | 7 |
| Abhishek | 92000 | 8 | 3 |
| Avantika | 120000 | 15 | 1 |
| Ayushi | 65750 | 7 | 5 |

The attributes salary and year_of_experience are on different scale and hence attribute salary can take high priority over attribute year_of_experience in the model.

## Data transformation and enrichment

- In addition to being structured, the data typically must be transformed into a unified and usable format.
- For example, data transformation may involve creating new fields or columns that aggregate values from existing ones.
- Data enrichment further enhances and optimizes data sets as needed, through measures such as augmenting and adding data.

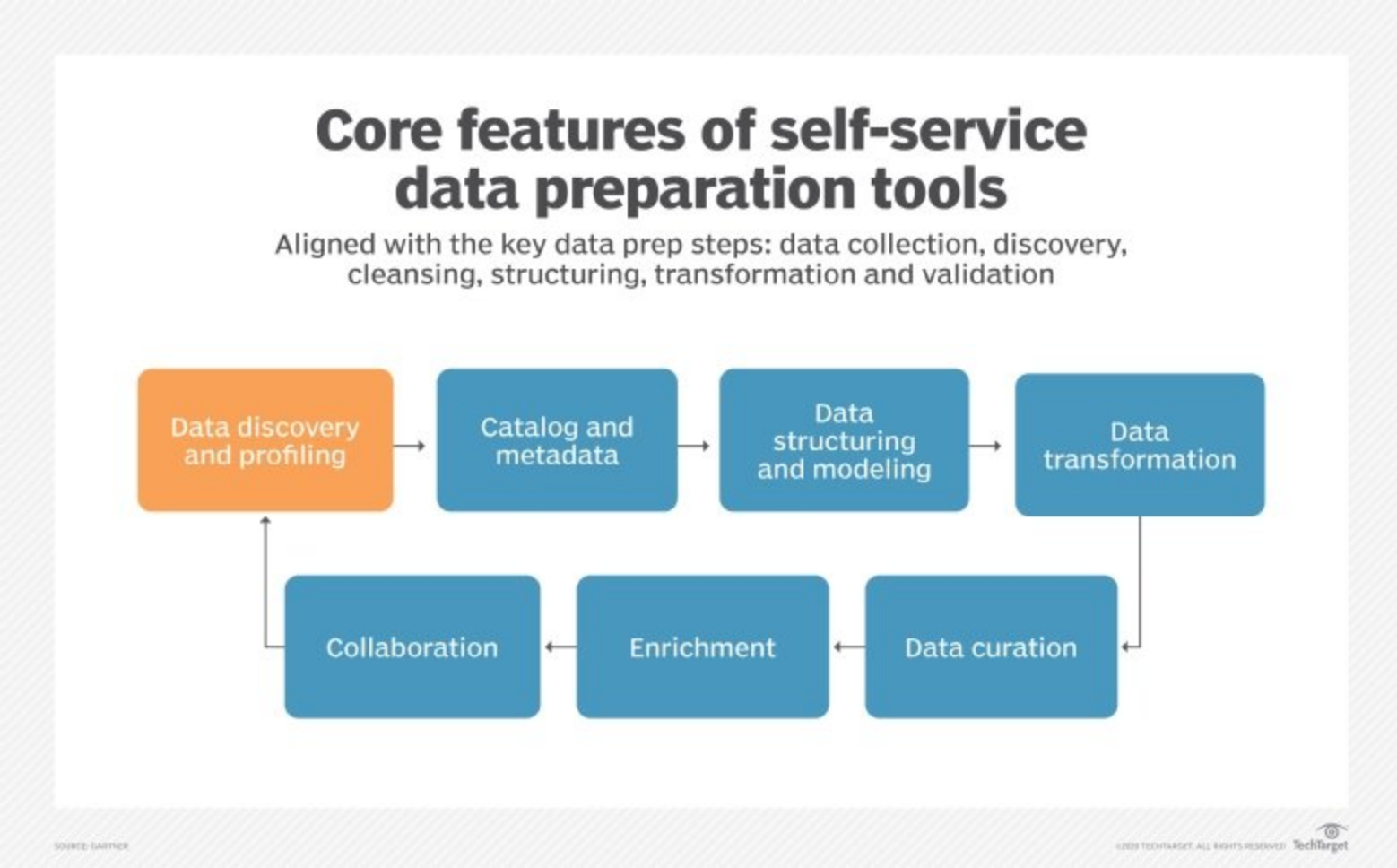
- **Data validation and publishing**

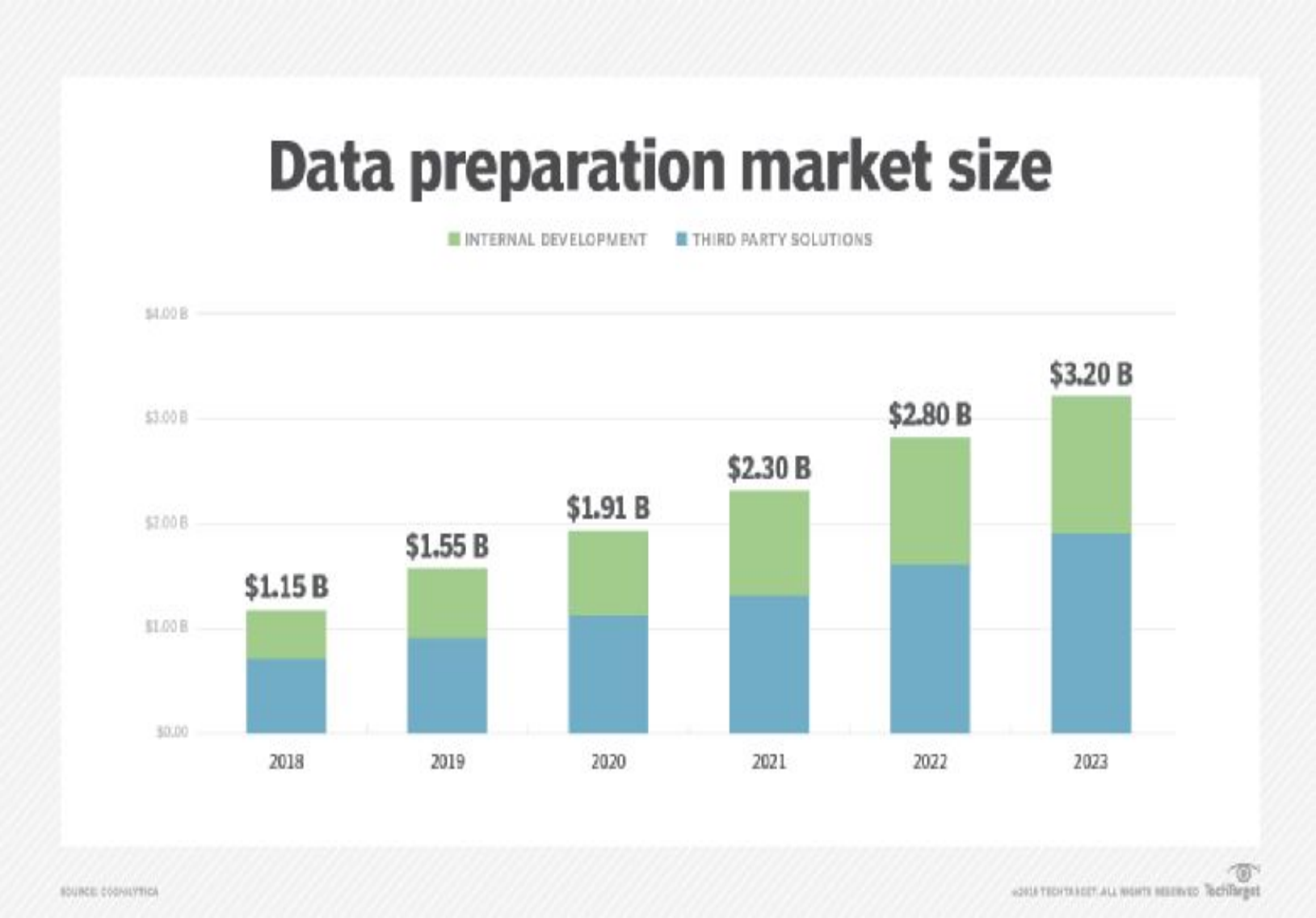
- In this last step, automated routines are run against the data to validate its consistency, completeness and accuracy.

- The prepared data is then stored in a data warehouse, a data lake or another repository and either used directly by whoever prepared it or made available for other users to access.

# Top challenges on data preparation

- Inadequate or nonexistent data profiling
- Missing or incomplete data
- Invalid data values
- Name and address standardization
- Inconsistent data across enterprise systems
- Data enrichment
- Maintaining and expanding data prep processes


Core features of self-service data preparation tools
Aligned with the key data prep steps: data collection, discovery, cleansing, structuring, transformation and validation
Data discovery and profiling
Catalog and metadata
Data structuring and modeling
Data transformation
Data curation
Enrichment
Collaboration
TechTarget


Data preparation market size
INTERNAL DEVELOPMENT
THIRD PARTY SOLUTIONS
$4.00 B
$3.00 B
$2.00 B
$1.00 B
$0.00
$1.15 B
$1.55 B
$1.91 B
$2.30 B
$2.80 B
$3.20 B
2018
2019
2020
2021
2022
2023

# References


- https://careerfoundry.com/en/blog/data-analytics/difference-between-quantitative-and-qualitative-data/
- https://www.scribbr.com/methodology/data-collection/
- https://www.mbaknol.com/research-methodology/measurement-scales-in-research-methodology/#:~:text=There%20are%20four%20different%20scales%20of%20measurement%20used,Nominal%20scale%20is%20the%20simplest%20form%20of%20measurement.