

Scientific Methodology in Computer Science

MO430

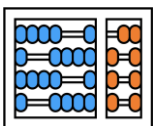
Prof. Dr. Bruno B. P. Cafeo

Institute of Computing
University of Campinas



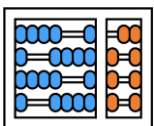
Agenda

- Overview of Recording and Organizing Data
- Importance in Scientific Methodology
- Examples in Computer Science
- Best Practices



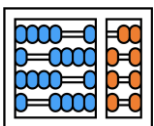
Importance in Scientific Methodology

- Why it matters:
 - Reproducibility of Experiments: Ensures others can replicate and verify results.
 - Accuracy in Analysis: Organized data leads to precise conclusions.
 - Meaningful Conclusions: The foundation for drawing insightful inferences.



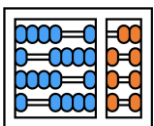
Role in Computer Science

- Data in Computer Science:
 - Drives Decision-Making: Informed choices based on data-driven insights.
 - Essential for Algorithm Development: Data structures and patterns.
 - Key to Machine Learning Models: Training requires quality, organized data.



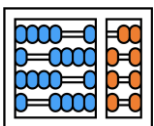
Manual Recording

- Traditional Methods:
 - Notebooks, Spreadsheets: Manual entry for experiments and observations.
- Example:
 - Imagine a researcher recording experimental results in a physical lab notebook, ensuring a tangible record of each step.



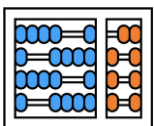
Automated Recording

- Modern Tools:
 - Sensors, Data Loggers: Automated collection of data in real-time.
- Example:
 - Consider IoT devices collecting environmental data, automatically logging temperature, humidity, and other metrics.



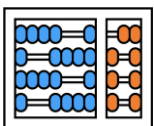
Database Systems

- Structured Storage:
 - Relational Databases, NoSQL Databases: Organized, efficient data storage.
- Example:
 - In a web application, user data is stored in a relational database, allowing for structured retrieval and manipulation.



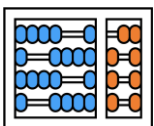
Logging

- Capturing Events:
 - Application Logs, Server Logs: Recording events and errors.
- Example:
 - When developing software, logs are crucial for debugging, providing insights into the program's behavior.



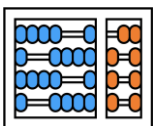
Version Control

- Tracking Changes:
 - Git, SVN: Managing and versioning code collaboratively.
- Example:
 - Git enables developers to track changes, collaborate seamlessly, and revert to previous versions when needed.



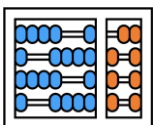
Data Structures

- Organizing Within Programs:
 - Arrays, Linked Lists, Trees: Structures for efficient data handling.
- Example:
 - Imagine a program using arrays and linked lists to organize and manipulate large datasets.



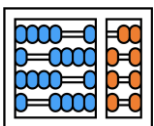
Databases and Tables

- Structured Database Design:
 - Normalization, Indexing: Techniques for efficient data organization.
- Example:
 - Designing a relational database schema involves normalization and indexing for optimal data storage and retrieval.



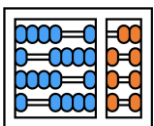
File Organization

- Managing Information:
 - Directories, Folders: Structuring files for easy access.
- Example:
 - In software development, organizing project files into directories ensures a coherent and accessible file structure.



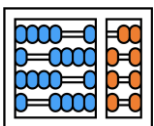
Metadata

- Descriptive Information:
 - Tagging Files with Additional Details: Enhancing data context.
- Example:
 - Adding metadata to files, like tags and descriptions, simplifies searching and understanding.



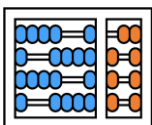
Data Cleaning

- Ensuring Quality:
 - Removing Duplicates, Handling Missing Values: Preprocessing for accurate analysis.
- Example:
 - Before training machine learning models, data cleaning involves tasks like removing duplicates and handling missing values.



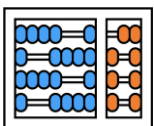
Machine Learning Dataset

- Preparing Data for Models:
 - Collection and Organization: Crucial steps before model training.
- Example:
 - Consider preparing an image dataset with labeled data for training a computer vision model.



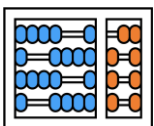
Software Development

- Version Control and Logging:
 - Ensuring Code Integrity: Versioning for collaborative projects.
- Example:
 - In software development, using Git for version control ensures the integrity and traceability of the codebase.



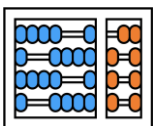
Web Analytics

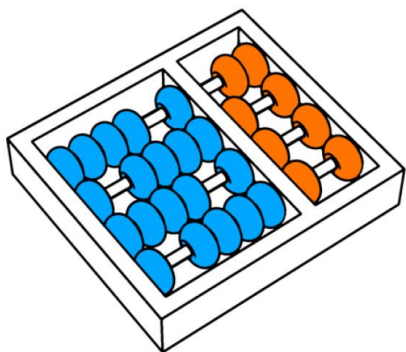
- User Behavior Analysis:
 - Tracking Interactions: Insights into user engagement.
- Example:
 - Analytics tools like Google Analytics track user interactions on a website, providing valuable insights for optimization.



Best Practices

- Guidelines for Success:
 - Consistent Naming Conventions: Uniformity for easy identification.
 - Regular Backups: Safeguarding against data loss.
 - Documenting Sources and Transformations: Enhancing reproducibility.
 - Collaborative Tools for Teamwork: Ensuring synchronized efforts.





**INSTITUTO DE
COMPUTAÇÃO**



Prof. Dr. Bruno B. P. Cafeo

Sala 04

Instituto de Computação Unicamp

Av. Albert Einstein, 1251

Cidade Universitária

Campinas – SP

13083-852

<https://ic.unicamp.br/~cafeo/>
cafeo@ic.unicamp.br