Module 4: How is Data Organized and Stored?

Background

This module introduces the concept of data storage by exploring different tools and systems used for data storage. The examples provided are based on DataWorks' systems—facilitators should adapt the slides to reflect their organization's specific tools and practices.

Additional Reading for Facilitators

- Pietsch, Wolfgang. *On the Epistemology of Data Science: Conceptual Tools for a New Inductivism*. Springer, 2022.
- Zhang, Amy X., et al. "How Do Data Science Workers Collaborate? Roles, Workflows, and Tools." *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. CSCW1, May 2020, pp. 1–23, https://doi.org/10.1145/3392826.

Module Motivation

Organizations use a variety of data storage systems and tools. Some are specific to particular client requests, while others are based on specific project needs. This module helps learners recognize key differences between these systems, understanding not only how they process data, but also what they mean for the security and privacy of respective datasets.

Learning Goals

Students will be able to:

1. Identify and delineate key characteristics between data processing tools (such as a local Excel document vs. an Excel document saved to the cloud vs. Google Sheets)
2. Develop and apply key organizational principles for a database schema, including appropriate columns (or "tabs" or "fields")

3. Recognize and identify various formats for working with and saving data tables

Warm-Up: What is Data Storage? (Slide 2)

Purpose:

Check in with learners' associations with data storage concepts.

Discussion Topics:

- **Storage systems**: Cloud storage vs. local storage
- **Product variations**: Different companies' solutions
- **Security concepts**: Encryption and access controls
- **Access considerations**: Who can potentially see file contents?
  - Co-workers
  - Clients
  - Product companies
- **Practical factors**:
  - Licensing fees
  - Availability across different operating systems or devices

Activity 1: Introducing Data Tables (Slides 3-11)

Purpose:

Introduce different formats in which learners might experience a data table.

Activity Structure:

Part 1: Table Format Progression

Starting point: Familiar table formats (as seen in polished handouts or newspaper articles) **Ending point**: How computers actually process tables (as a series of characters)

Part 2: Format Utility Discussion

Key question: What is the utility of these different kinds of tables?

Answer framework:

- Benefits are specific to each organization
- Different formats represent different ways of creating and using tables in software
- Focus on benefits and interaction methods specific to each application type

Part 3: Table Composition and Relationships

Focus areas:

- How tables can be composed to show relationships between different data items
- Understanding dimensions of data
- Practicing how to "pull" (or "query") particular dimensions of data based on known information about entities

Facilitation Notes:

- Use examples relevant to your organization's specific software and tools
- Emphasize hands-on exploration of different table formats
- Encourage discussion about practical applications in learners' work contexts

Reflection & Journal Entry

Assignment:

Design a small database and practice identifying related data items via functions or queries.

Guidelines:

- Can be based on real-life events or connections
- Focus on translating real-world relationships into data table environments
- Practice query techniques learned during the activity

Key Skills to Demonstrate:

- Database design principles
- Relationship identification between data items

- Query formulation and execution