Università degli Studi di Padova

Introduction to Databases

Basi di Dati

Bachelor's Degree in Computer Engineering
Academic Year 2024/2025

Stefano Marchesin

Intelligent Interactive Information Access (IIIA) Hub
Department of Information Engineering
University of Padua

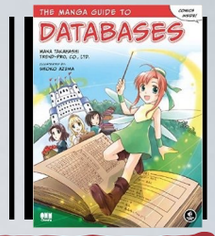
Dipartimento di Ingegneria dell'Informazione

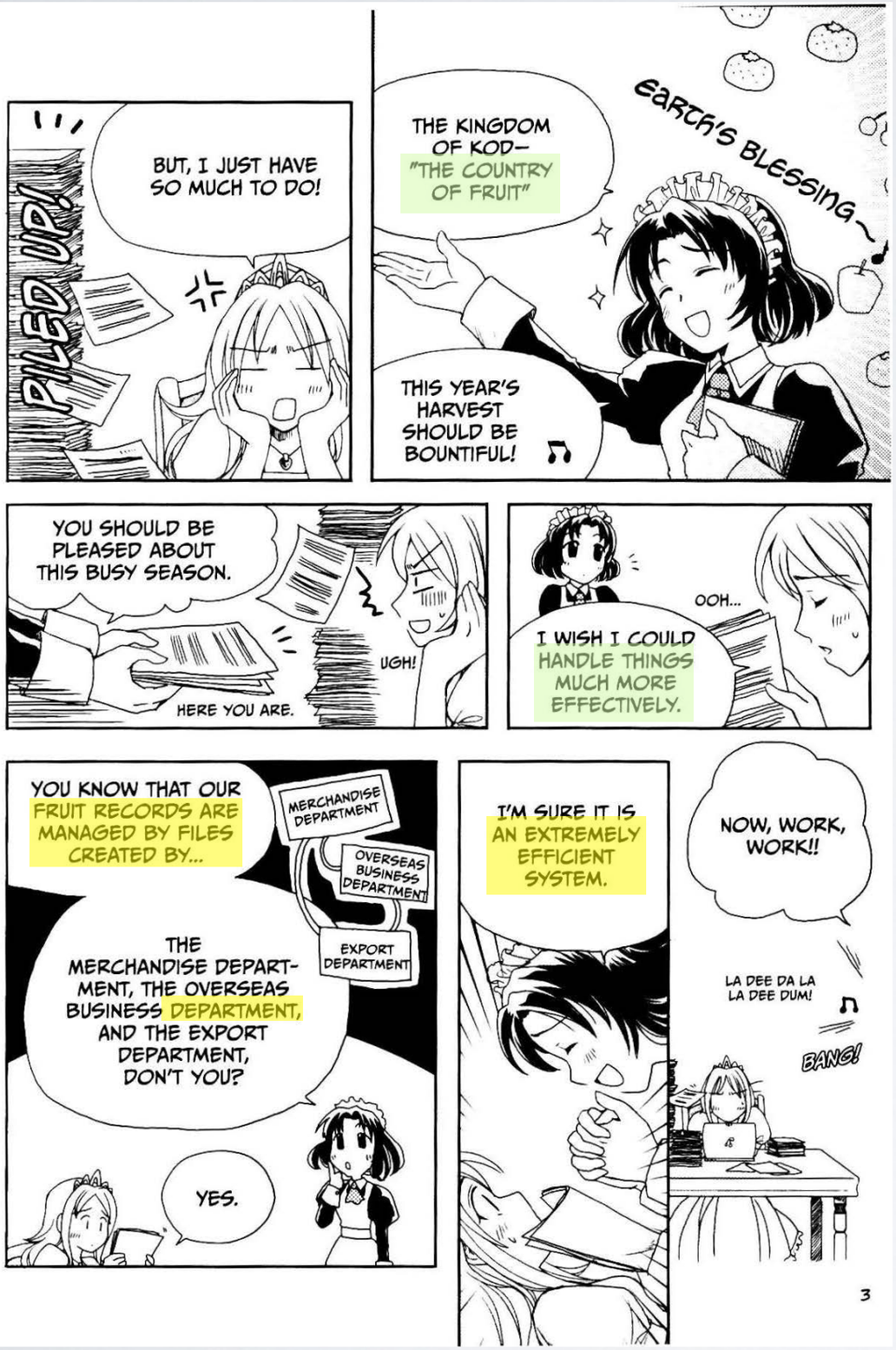
Outline

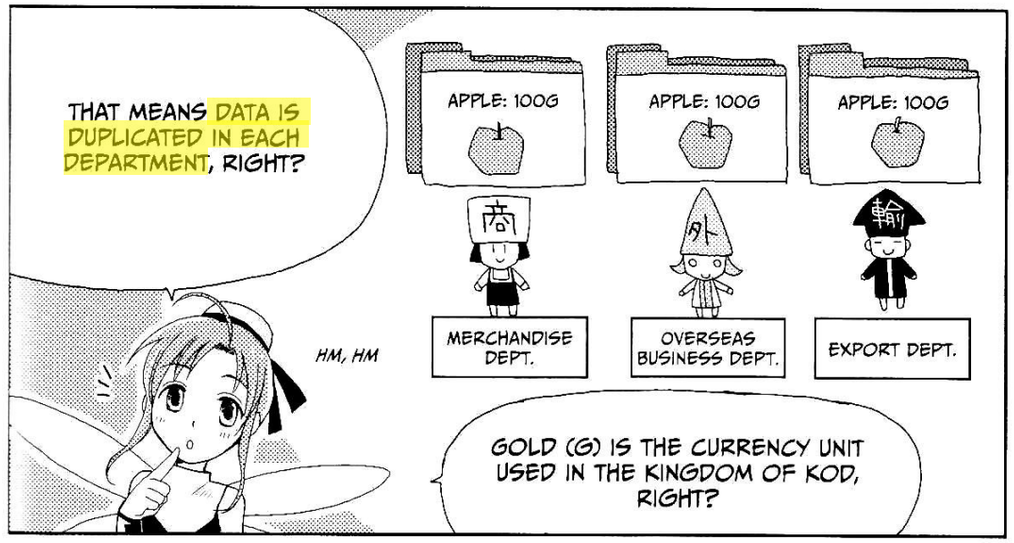
- Introduction to databases
- Advantages of using the DBMS approach
- Data-independence and the ANSI/SPARC architecture

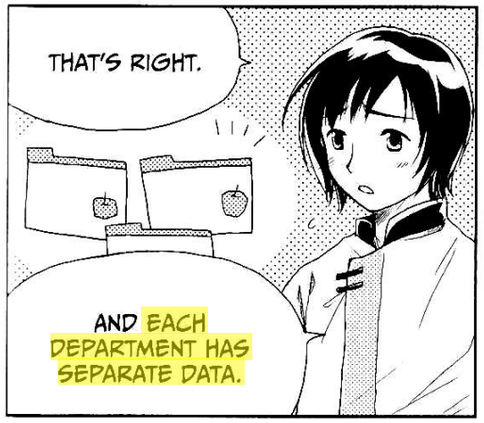
Introduction to Databases

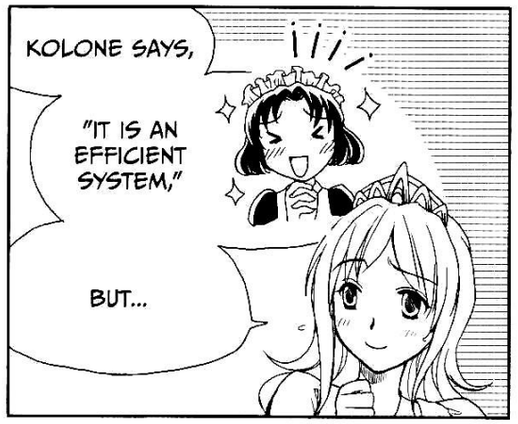
Once upon a time…

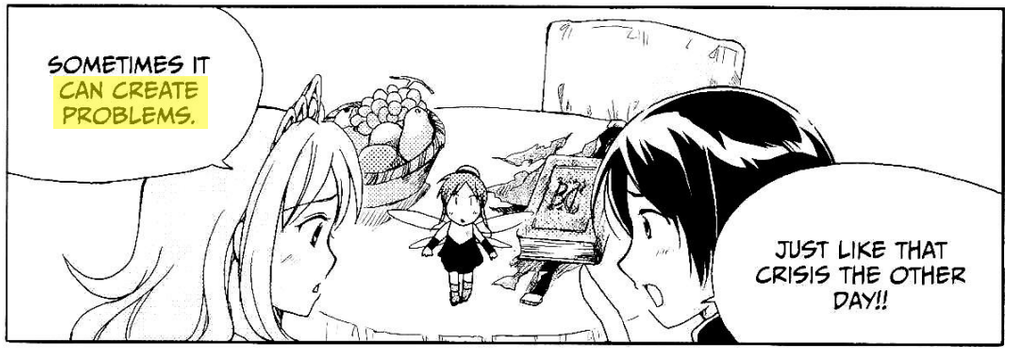






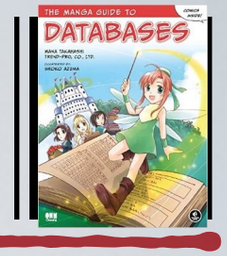


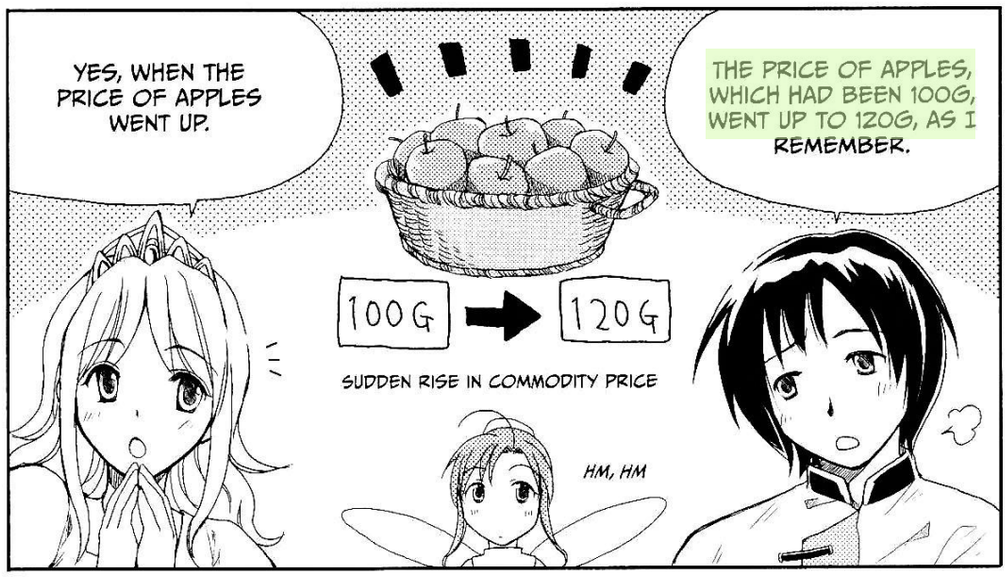


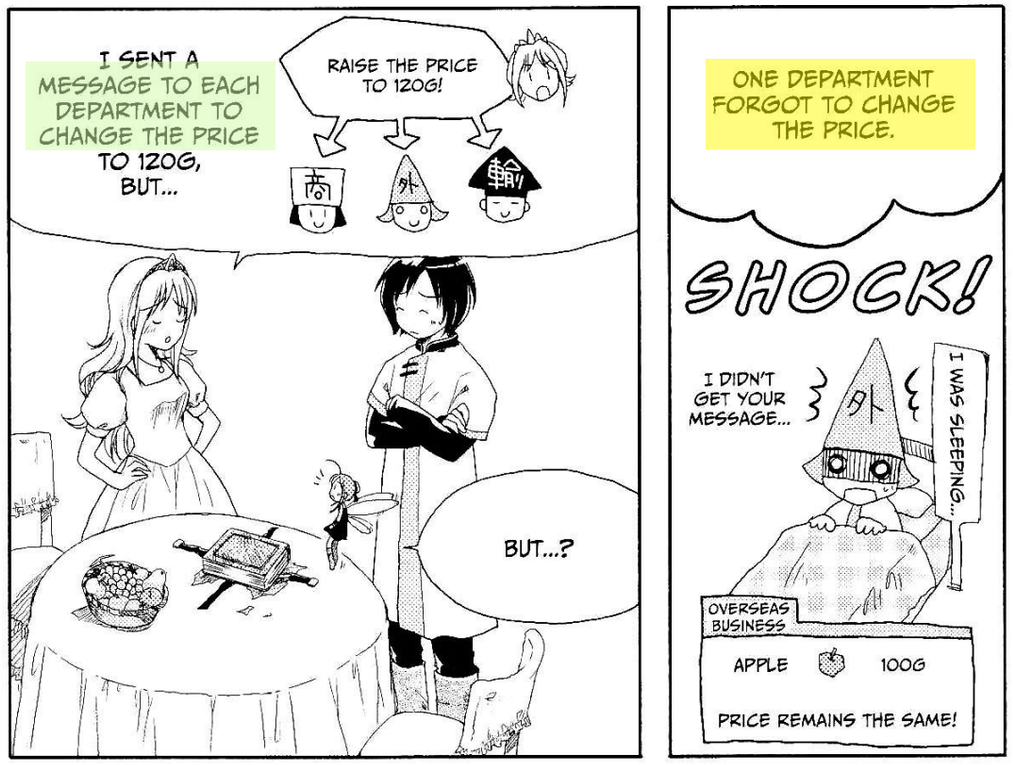


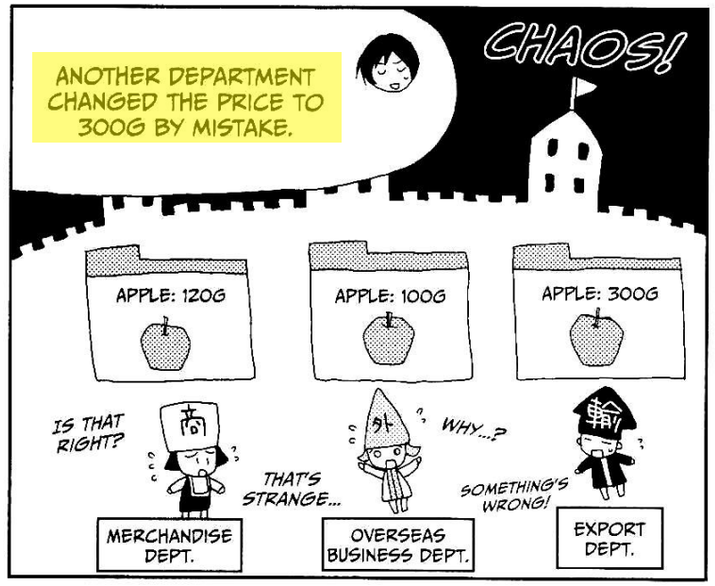
WHAT IS A DATABASE? 11

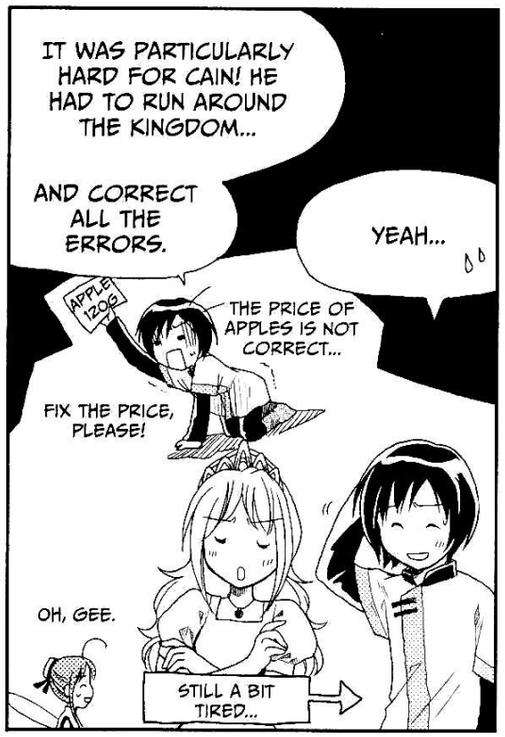
Once upon a time…



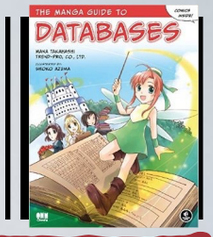


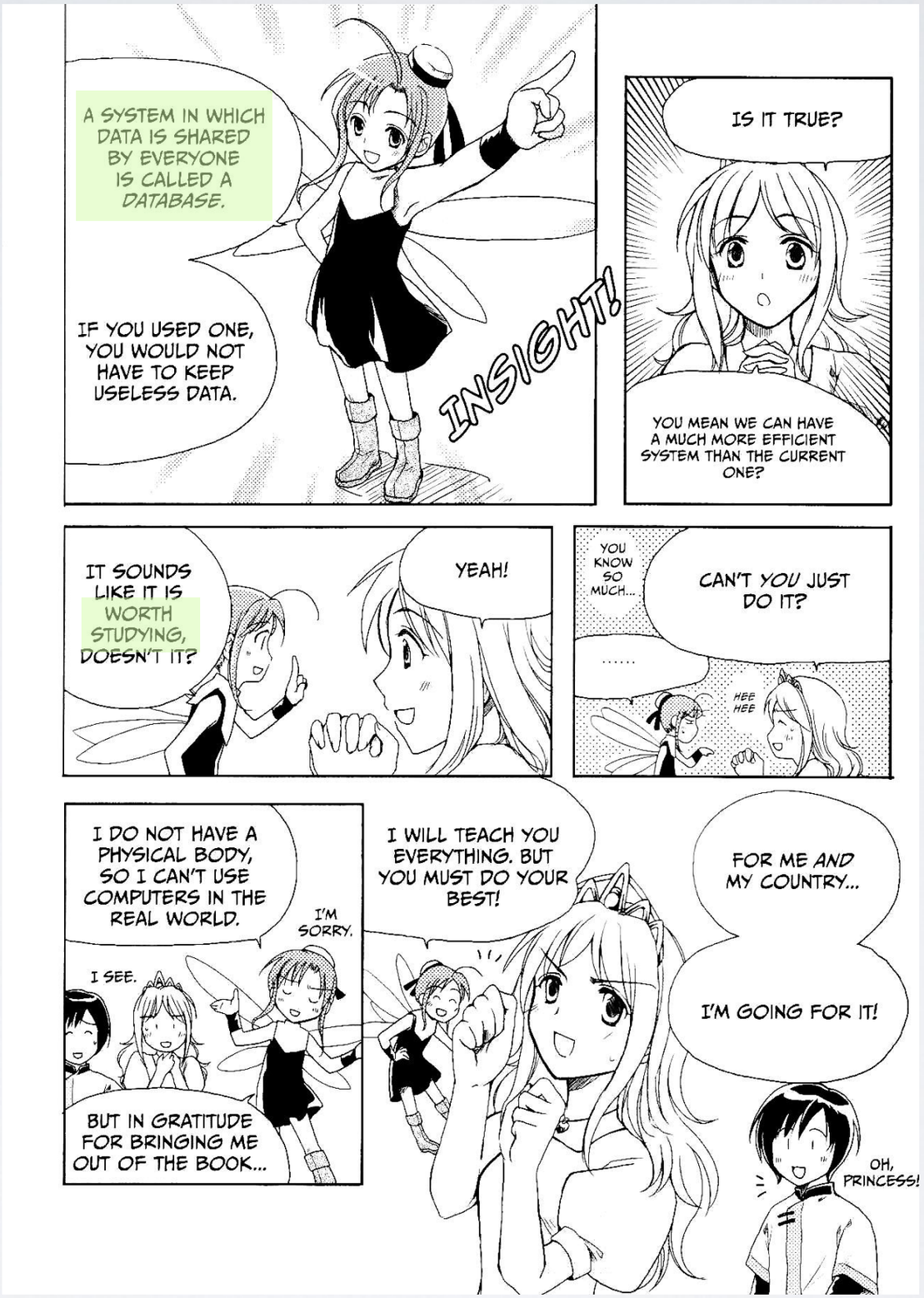






Once upon a time…





Database

A **database (DB)** is a **permanent collection of related data**, where **data** indicates **known facts** which can be recorded and have an intrinsic meaning

- It represents some aspects of the real world, also called **mini-world** or **universe of discourse (UoD)**
- It is a **coherent and integrated collection** of data which is kept for a long period, also tens of years
- A database is designed for a specific purpose and to satisfy the needs of well identified **users**

Database Management System

A **database management system (DBMS)** is a **general purpose software system** which allows users to create, manage, and update a database

- Definition of a database
  - **schema** of the database (data types, structure, constraints) stored in the system **catalog** as **metadata**
- Construction of a database
  - storing of the data themselves
- Manipulation of a database
  - querying and updating the data

Features of Databases

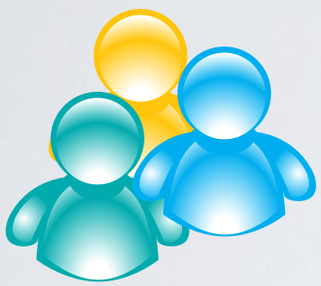
- Databases are **big**
  - their dimension is much bigger than central memory available, i.e. terabyte or more and more frequently petabyte
- Databases are **shared** among applications and users
  - reduction of **redundancy** and **inconsistency**
  - **concurrency control** to avoid undesired interaction between users/applications (**isolation**) and partial or incomplete operations on the data (**atomicity**)
- DBMS are **durable**, i.e. they keep the data also in the case of hardware/software malfunctioning
- DBMS guarantee the **security** of the data
  - users and applications can access data only upon **authentication** and **authorisation**
- DBMS are **efficient**
  - they optimise the use of resources, both in space and in time

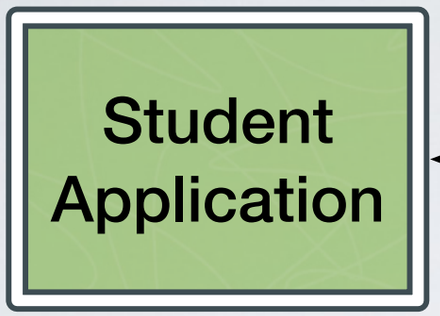
Why the File System is not a Database?

The file system

- allows us to store **big** amounts of data
  - the support for data schema is fairly limited (directory trees)
  - it does not guarantee efficient access to data whose exact position is not already known
- provides a **permanent** data storage
- does not necessarily guarantee the **durability** of the data
  - you need additional functionalities on top of it, such as checksums or backup policies
- does not avoid data **redundancy** and **inconsistency**
- does not provide concurrency control in terms of **isolation** and **atomicity**
- requires ad-hoc application to provide DBMS-like functionalities

From Archives to Databases





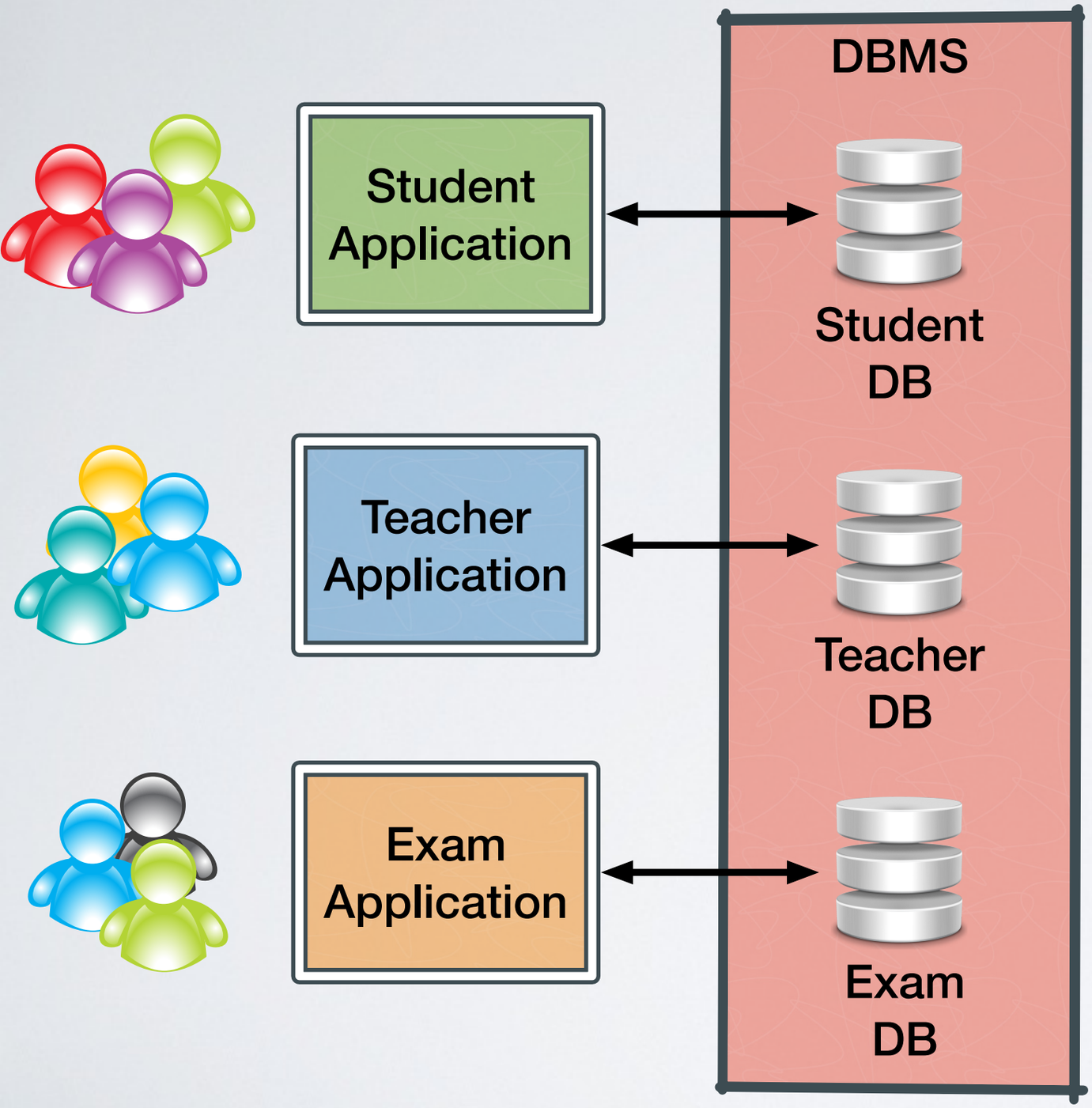
Student Archive (Files)

Teacher Archive (Files)

Exam Archive (Files)

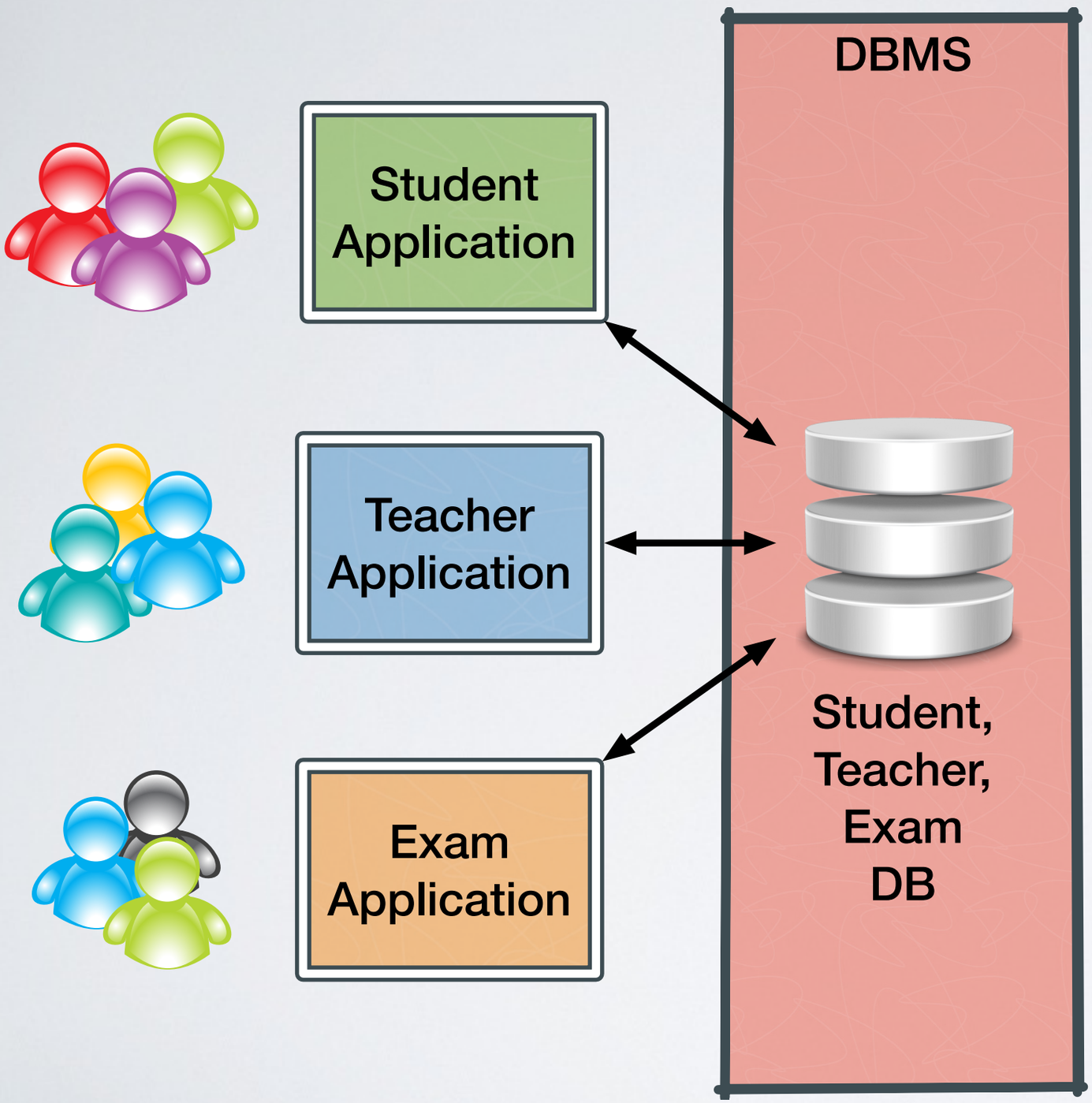
- Duplication and high data redundancy
  - different representations of the data in different archives (files)
  - inconsistencies possible
  - waste of resources
- Different life-cycles for the same data
  - inconsistencies possible
- Different physical features
  - local and not global optimisation
- Different integrity constraints
  - inconsistencies possible

From Archives to Databases



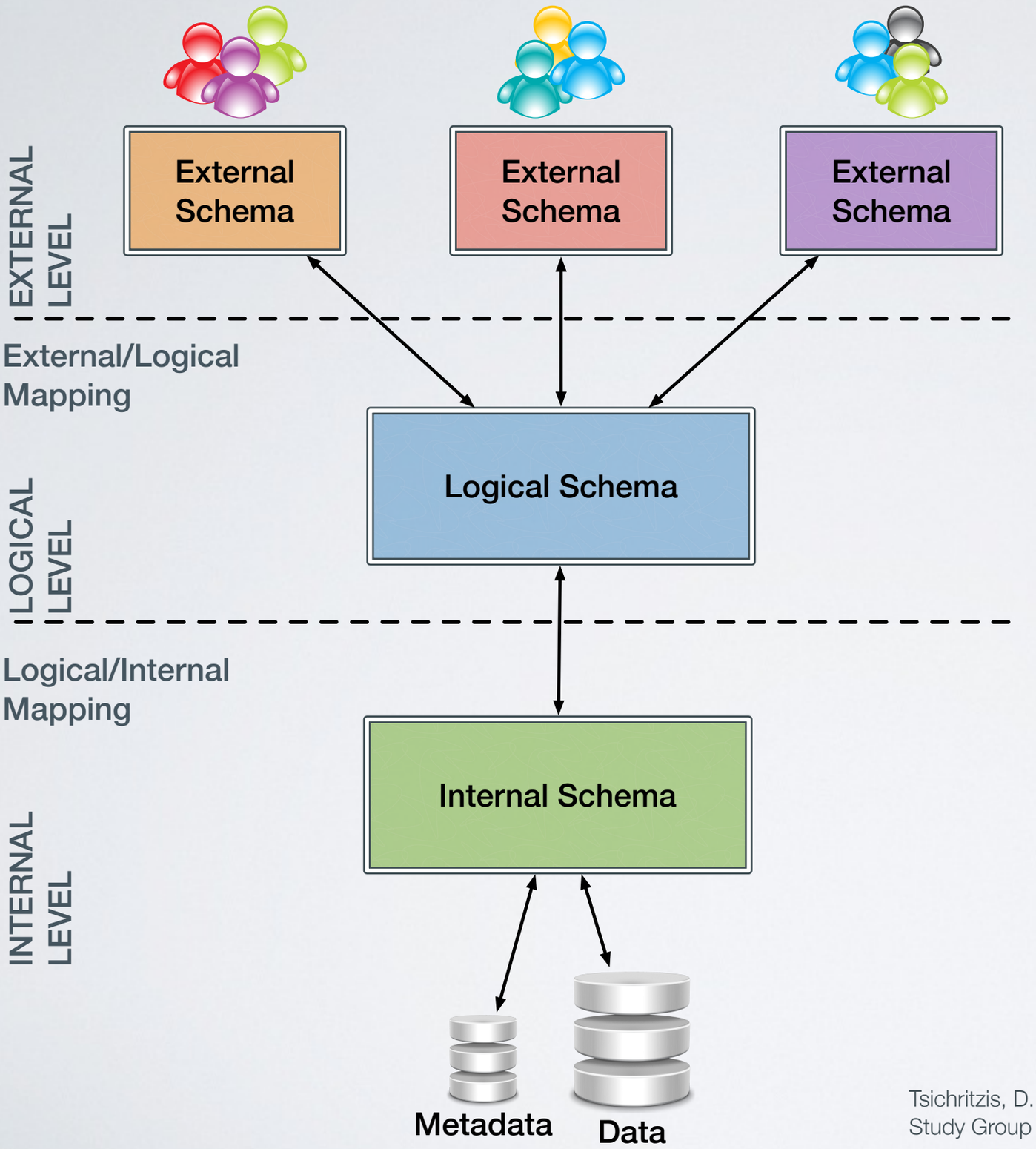
- Duplication and high data redundancy
  - different representations of the data in different archives (files)
  - inconsistencies possible
  - waste of resources
- Different life-cycles for the same data
  - inconsistencies possible
- ~~Different physical features~~
  - ~~local and not global optimisation~~
- Different integrity constraints
  - inconsistencies possible

From Archives to Databases



- ~~Duplication and high data redundancy~~
  - ~~different representations of the data in different archives (files)~~
  - ~~inconsistencies possible~~
  - ~~waste of resources~~
- ~~Different life-cycles for the same data~~
  - ~~inconsistencies possible~~
- ~~Different physical features~~
  - ~~local and not global optimisation~~
- ~~Different integrity constraints~~
  - ~~inconsistencies possible~~

Standard ANSI/SPARC Architecture



- **External schema**: one for each application, where only the data relevant to that application are described
- **Logical schema**: integrated representation of the data, independent from the physical representation
- **Internal schema**: physical representation into data structures and storage units

Tsichritzis, D. and Klug, A. The ANSI/X3/SPARC DBMS Framework Report of the Study Group on Database Management Systems. *Information Systems*, 3(3):173–191, 1978.

Data Independence

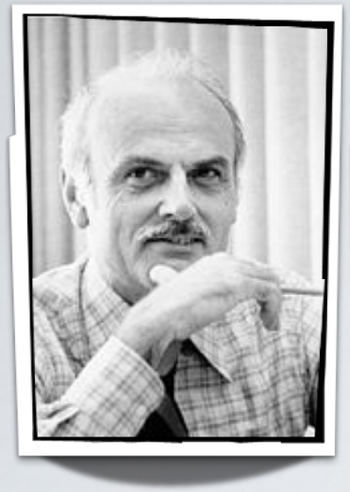
Thanks to the two-level architecture, **access happens only through the external level** (which may coincide with the logical level)

- Two types of data independence
  - **physical independence**: the external and logical levels are independent from the physical level, which may changes without affecting them
  - **logical independence**: the external level is independent from the logical one
    - additions or modifications to the external views do not require changes at logical level

DBMS Actors

- Designers and developers of DBMS
- Database administrators (DBA)
- Designer of databases
- Designer and developers of applications
- Users
  - parametric end users who perform pre-defined activities
  - casual end users who perform general and not pre-defined activities

A Bit of History



1970
Codd Paper on the Relational Model

1976
Chen Paper on the ER Model

1977
IBM System R

1979
Oracle

1983
IBM DB2

1987
ISO SQL/86

1989
ISO SQL/89

1992
ISO SQL/92

1998
Open Source Initiative

1999
ISO SQL:1999

2003
ISO SQL:2003

2006
ISO SQL:2006

2008
ISO SQL:2008

2011
ISO SQL:2011

2016
ISO SQL:2016

2021
ISO SQL:2021?

1998
MySQL (Open Source)

1977 - 1985
Ingres (RM) at Berkeley

1986 - 1994
Postgres (ORM) at Berkeley

1994 - 1995
Postgres95 (SQL) at Berkeley

1996 - 2021
PostgreSQL Global Development Team

1970

2023

Wrap-up on Databases (1/2)

- Databases are **self-descriptive**
  - **metadata catalog** and **data**
- **Separation** between programs and data
  - data abstraction
- **Multiple views** on data
  - further data abstraction
- Data **sharing** and **transaction management**
  - **concurrency control, isolation** and **atomicity**

Wrap-up on Databases (2/2)

- Redundancy control
  - it avoids **errors** due to data **duplication** and **saves space**
- Access control
  - **authentication** and **authorization**
- Durability
  - resilience to **hardware/software malfunctioning**
- Efficiency
  - complex **query** processing by **optimizing space** and **time** resource consumption

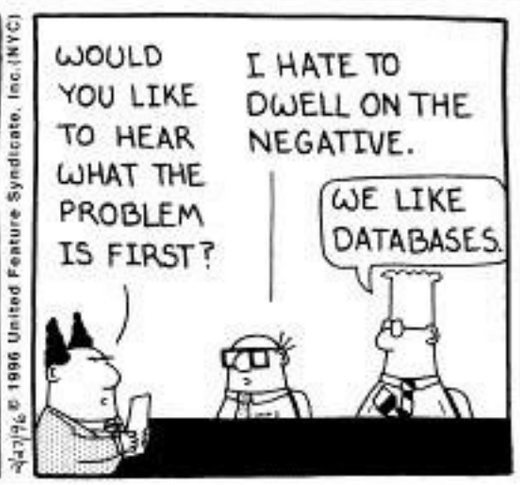
Further Readings

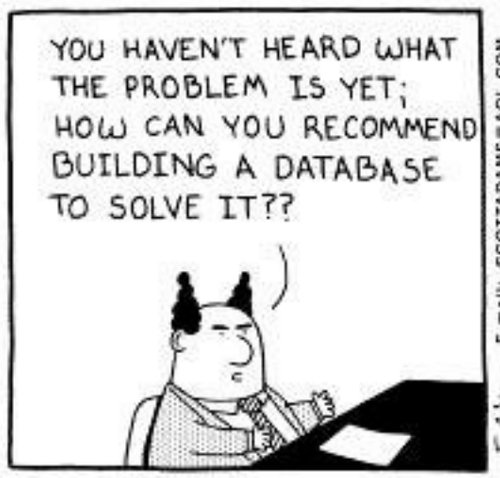
- Ackoff, R. L. (1989). From Data to Wisdom. *Journal of Applied Systems Analysis*, 16:3–9.
- Batini, C., De Petra, G., Lenzerini, M., and Santucci, G. (2002). *La progettazione concettuale dei dati*. Franco Angeli, Milano.
- Garcia-Molina, H., Ullman, J. D., and Widom, J. (2009). *Database Systems. The Complete Book. 2nd Edition*, Pearson Prentice Hall, Upper Saddle River (NJ), USA.
- Rowley, J. (2007). The Wisdom Hierarchy: Representations of the DIKW Hierarchy. *Journal of Information Science*, 33(2):163–180.
- Tsichritzis, D. and Klug, A. (1978). The ANSI/X3/SPARC DBMS Framework Report of the Study Group on Database Management Systems. *Information Systems*, 3(3):173–191.
- Zeleny, M. (1987). Management Support Systems: Towards Integrated Knowledge Management. *Human Systems Management*, 7(1):59–70.

Questions?


I JUST GOT OUR CONSULTANT'S REPORT. HE'S IDENTIFIED OUR BIGGEST PROBLEM.
S. Adams E-mail: SCOTTADAMS@AOL.COM


I RECOMMEND THAT WE BUILD A TRACKING DATABASE.
WE CAN PUT IT ON THE NETWORK.
S. Adams E-mail: SCOTTADAMS@AOL.COM


WOULD YOU LIKE TO HEAR WHAT THE PROBLEM IS FIRST?
I HATE TO DWELL ON THE NEGATIVE.
WE LIKE DATABASES.
© 1996 United Feature Syndicate, Inc.(NYC)


YOU HAVEN'T HEARD WHAT THE PROBLEM IS YET; HOW CAN YOU RECOMMEND BUILDING A DATABASE TO SOLVE IT??
S. Adams E-mail: SCOTTADAMS@AOL.COM


WE ALWAYS BUILD A DATABASE.
AND WE'LL NEED COFFEE MUGS FOR THE PROJECT TEAM.
© 1996 United Feature Syndicate, Inc.(NYC)


THE PROBLEM IS THAT WE HAVE POOR PROCESSES.
THAT COULD BE THE SLOGAN ON OUR MUGS!