

# DATABASE METHODOLOGY

## Database Fundamentals

### Introduction

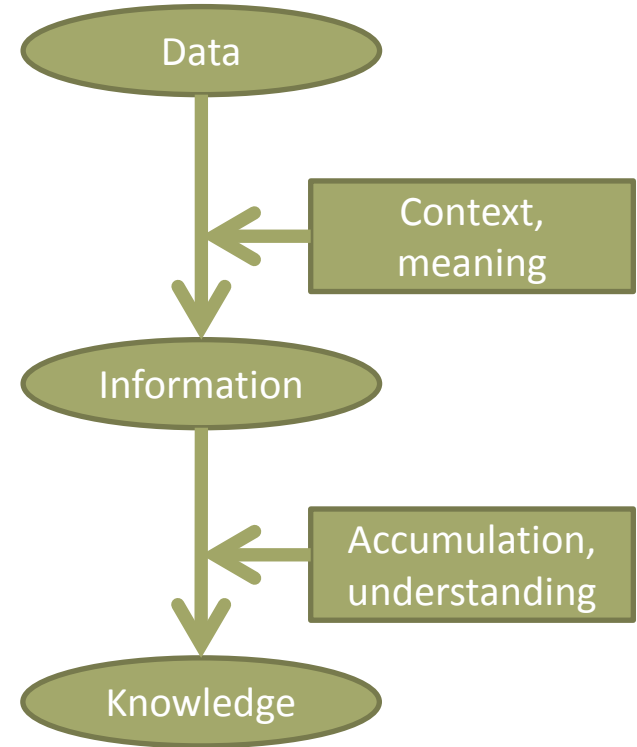
# Databases - Basics

- In this module you will learn some introductory basics about databases
  - What is data? Information? Knowledge?
  - Traditional file-based systems
  - Databases defined
  - Database Management Systems (DBMSs)

# What Is Data? Information? Knowledge?

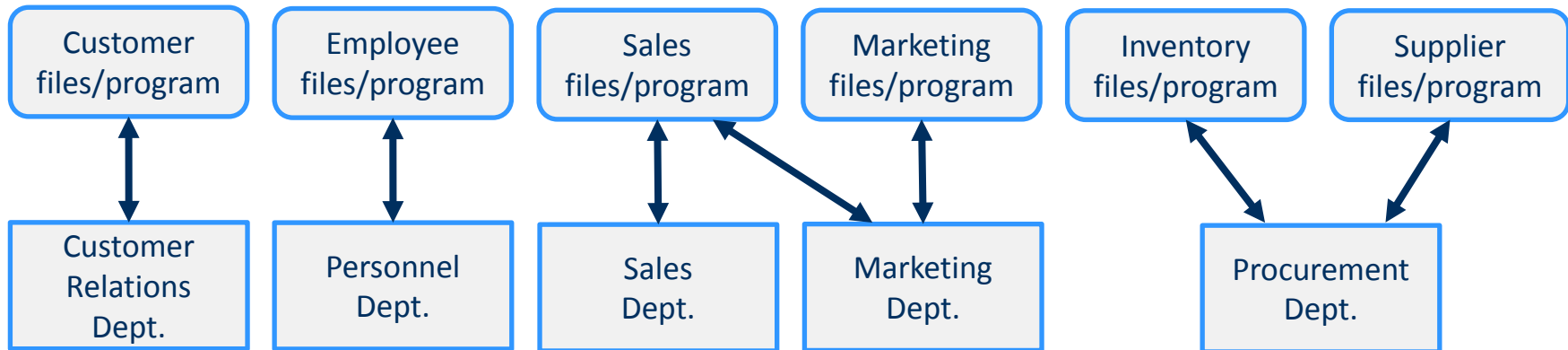
## Often confused and used as synonyms

- **Data** are raw facts without context (without meaning)
  - Strings, numbers, dates, ...
    - 'John', 77, 'kg', 1995-04-18
  - In general not so useful to us
- **Information** is data put into context (given meaning)
  - The person 'John' weighs 77 'kg' and was born on 1995-04-18
  - Information *makes sense* to us
- **Knowledge** is accumulated information, about a certain domain of interest, that we are aware of/understands
  - Can e.g. be used to derive new knowledge



# Traditional File-Based Systems

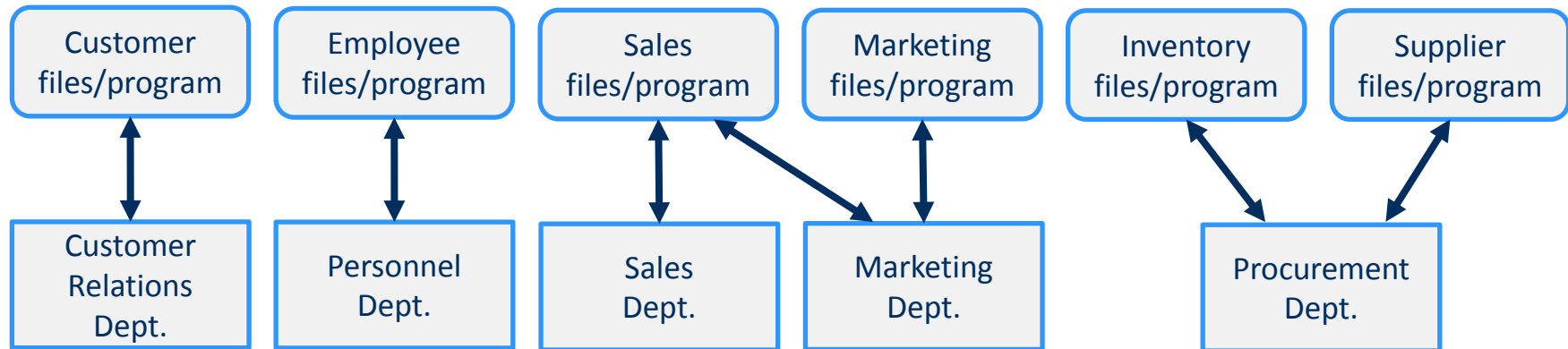
- The first data storage approach: **File-based systems**
  - “A collection of application programs that perform services for the end-users, such as the production of reports. Each program defines and manages its own data” *Connolly/Begg, “Database Systems”*
  - Used from the early days of the computer era



One possible structure of a file-based system

# Traditional File-Based Systems, Cont.

- The first data storage approach: **File-based systems**
  - Mimics manual filing systems used to keep track of documents about projects, products, etc
  - Key feature: Each project, product, etc, has its own separate filing system (program)



One possible structure of a file-based system

# Problems With File-Based Systems

- **Data is separated and isolated**
  - Aggregating data from several files is hard

- **Incompatible file formats**
  - File structures are specific to program languages

- **Data is often duplicated**
  - Waste of space, updates in several places

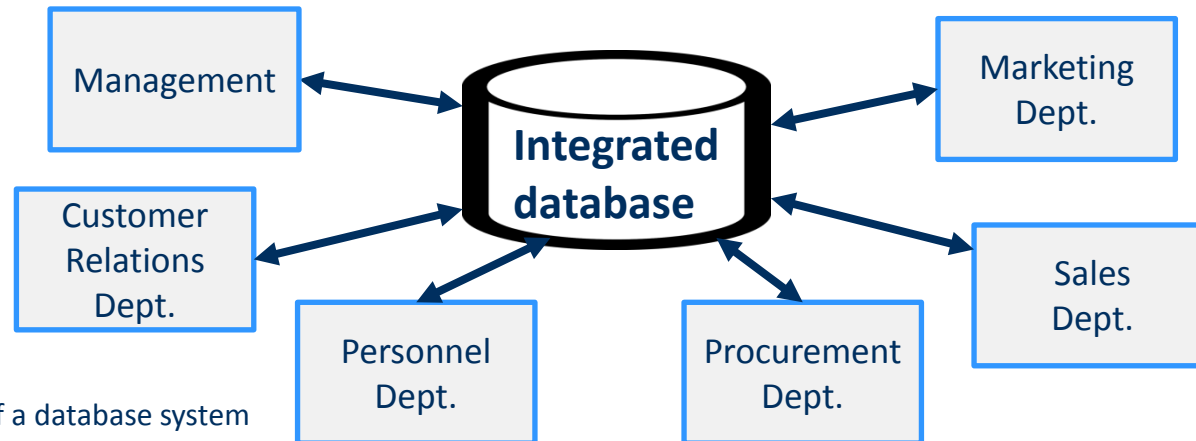
- **Only pre-programmed queries about the data**
  - “Ad hoc” querying prohibited by the complexity

- **Program-data structure dependence**
  - Changes to one often lead to changes to the other

- **Little or no built-in security or integrity**
  - Problems with recovery, data control, and sharing

# Solution, Databases Defined

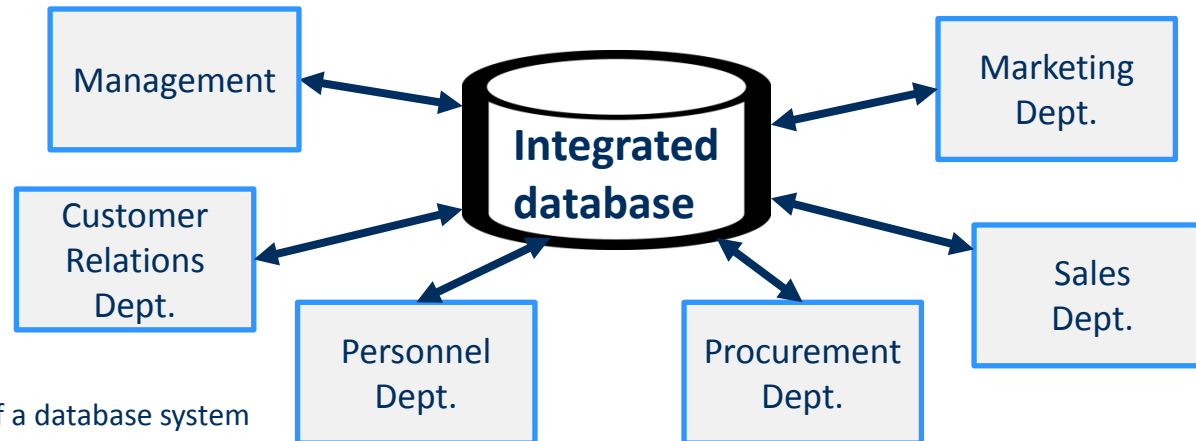
- The next-gen data storage approach: **The Database**
  - **“A shared collection of logically related data and its description, designed to meet the information needs of an organization”** *Connolly/Begg, “Database Systems”*
  - Data + metadata, a description of the data
    - Description in the *system catalog (data dictionary)*



One possible structure of a database system

# Databases Defined, Cont.

- The next data storage approach: **The Database**
  - Data items integrated to reduce duplication, and format issues
  - Shared by all users
  - Program-data structure **in**dependence
  - Simplified “ad hoc” querying of the data
  - Increased possibility to build-in security and integrity



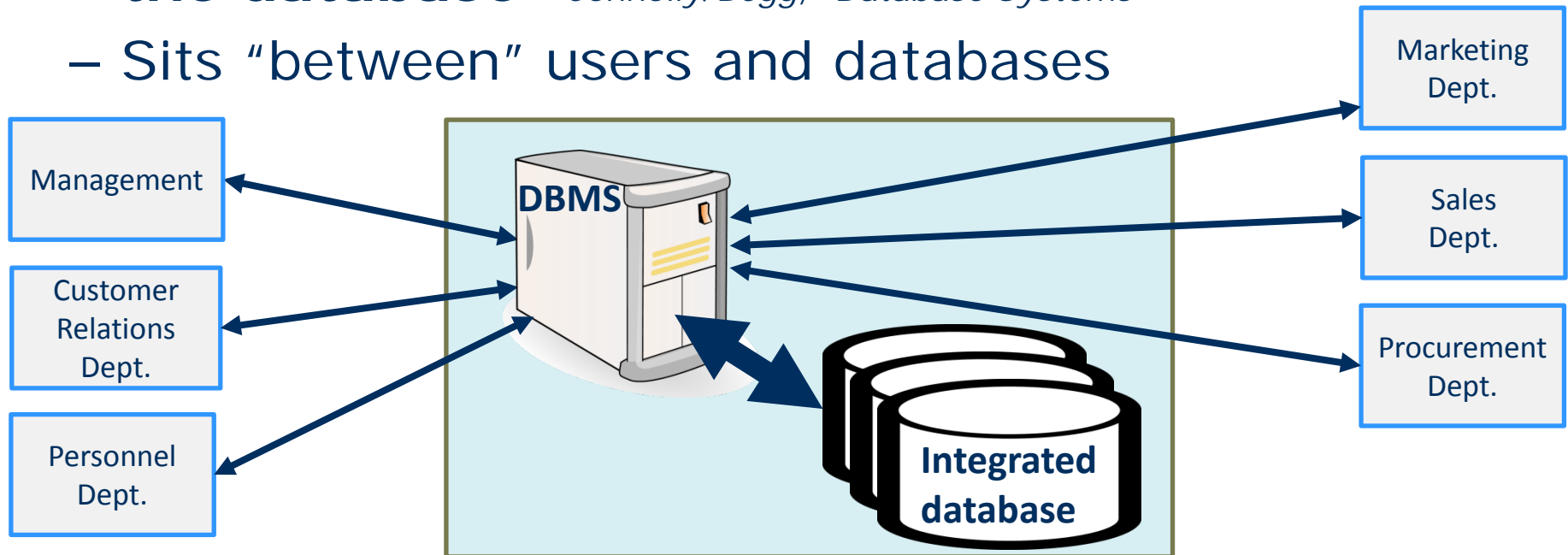
One possible structure of a database system



# Managing Database Systems

- **Database Management System (DBMS)**

- “A software system that enables users to define, create, maintain, and control access to the database” *Connolly/Begg, “Database Systems”*
- Sits “between” users and databases



# DBMS: Advantages And Disadvantages

Advantages		Disadvantages
Less data redundancy (saves storage)	Economy of scale (cost savings)	Complex piece of software
Data consistency (fewer places to keep updated)	Balance requirement conflicts	Large piece of software
Connect data to get more information ("1 + 1 = 3")	Better accessibility and responsiveness ("ad hoc" querying)	Cost of DBMS, including maintenance
Data sharing is easier	Increased productivity (less time spent on programming)	Additional hardware costs (servers, storage)
Better data integrity (business constraints enforced more easily)	Easier maintenance through program-data structure independence	Conversion cost (from old "legacy" system)
Improved security (access control)	Increased concurrency (thousands of users)	Performance (dedicated software is often faster)
Enforcing standards of various kinds	Easier backup and recovery (automation)	Failures could have very high impact

# Databases - Basics

- So, now you know the basics about
  - Data, information, and knowledge
  - Traditional file-based systems
  - The definition of databases
  - Database Management Systems (DBMSs)

# Medverkande

Anders Thelemyr – Lärare

Lars In de Betou – Mediepedagog

Inspelat 2015-08-27

Institutionen för data- och systemvetenskap, DSV



Stockholms  
universitet