iMeMex: Escapes from the Personal Information Jungle

Jens-Peter Dittrich   Marcos A Vaz Salles

Donald Kossmann   Lukas Blunschi

ETH Zurich, Switzerland

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

iMeMex.org
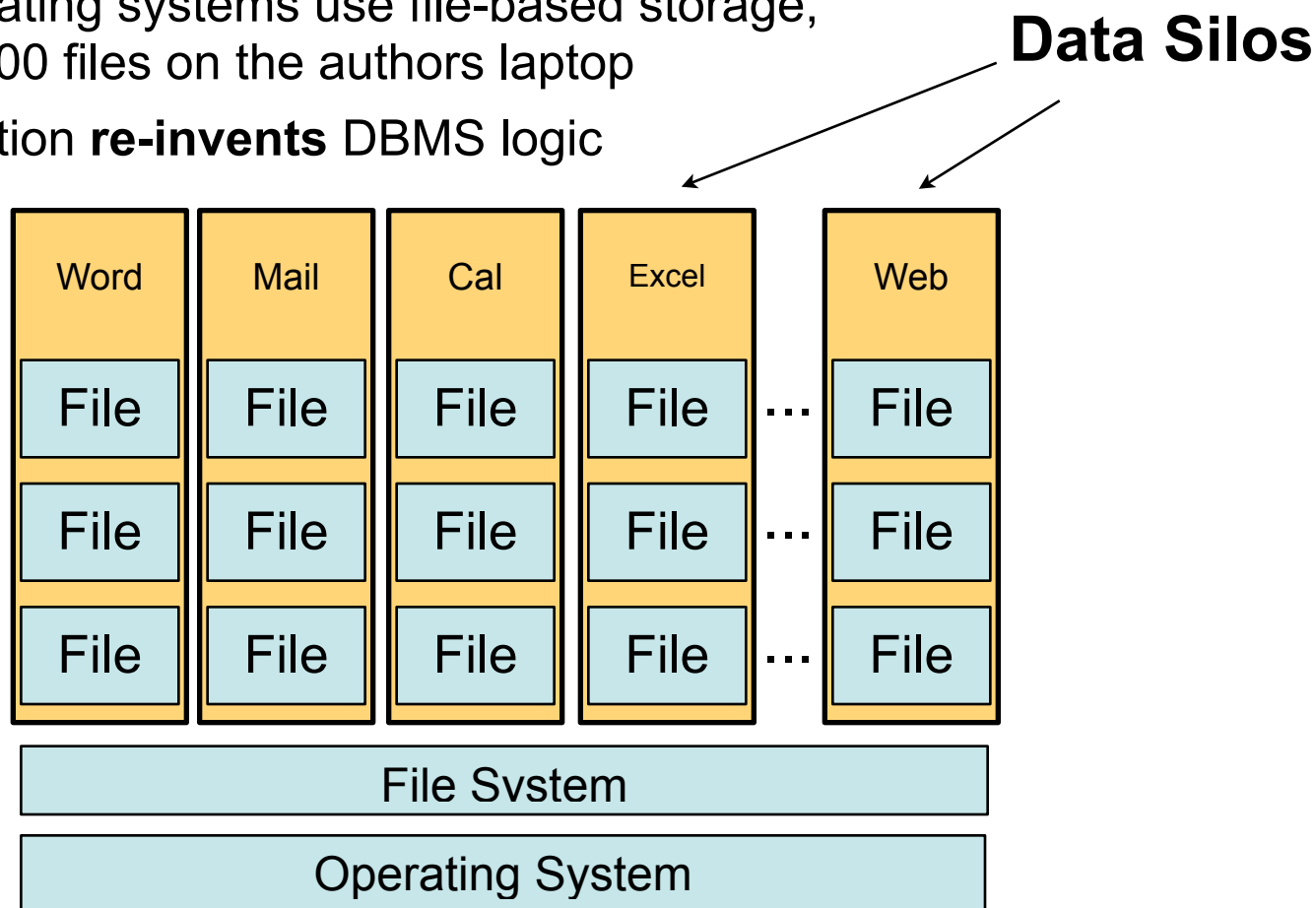Personal Information Management System

What is Personal Information?

- emails
- contacts
- pdfs
- excels, word, presentations, txt
- code (java, C++, ...)
- latex-sources
- pictures
- music
- application preferences
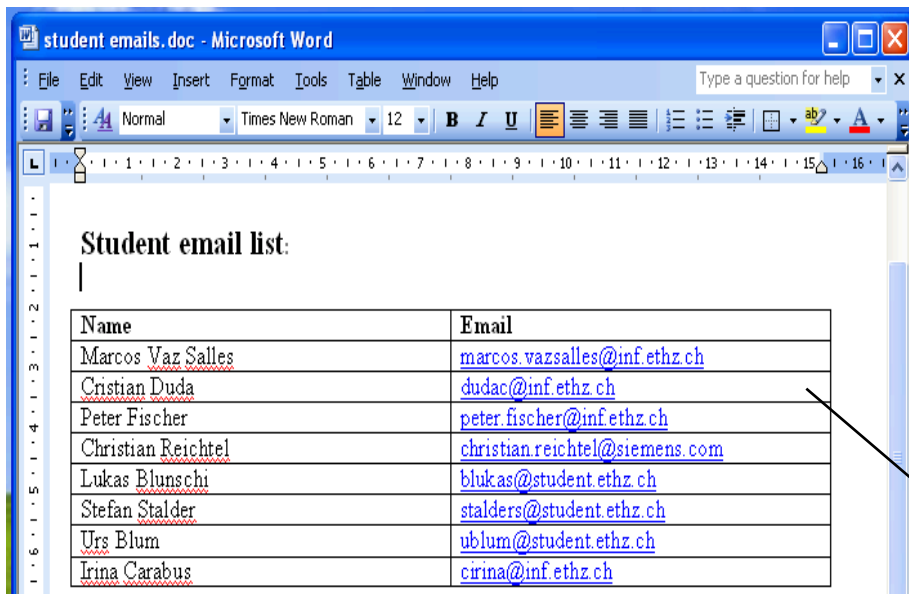- OS configuration information
- passwords
- etc.
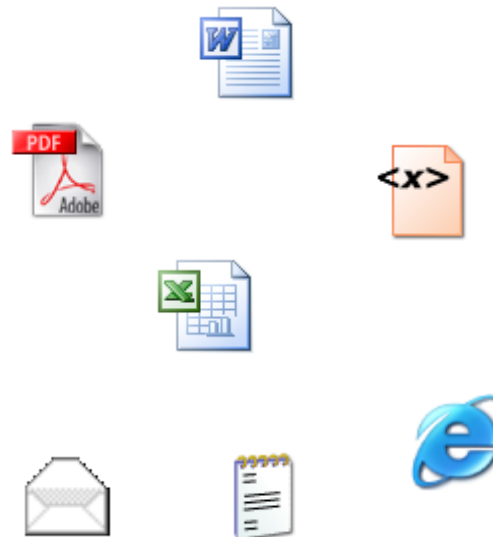
1st Problem: Data Silos

- Most data lies outside DBMS
- Current operating systems use file-based storage, e.g., > 700,000 files on the authors laptop
- each application **re-invents** DBMS logic

Data Silos

Word	Mail	Cal	Excel		Web
File	File	File	File	…	File
File	File	File	File	…	File
File	File	File	File	…	File

File System

Operating System

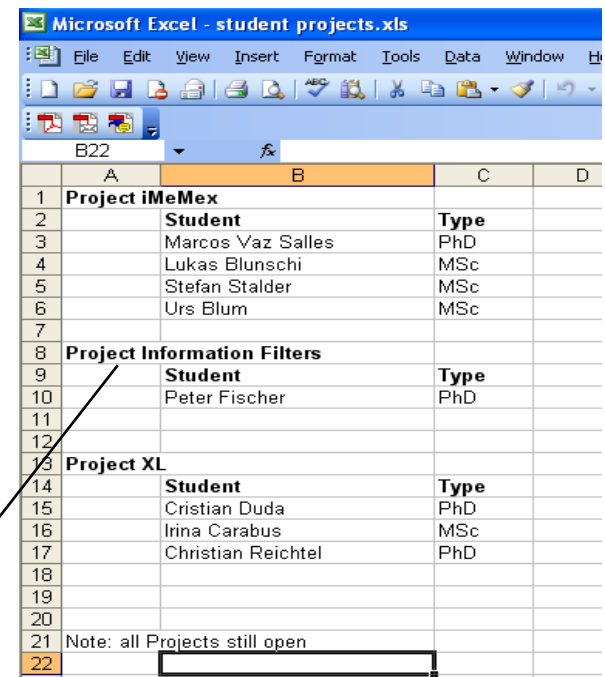# 2nd Problem: Lack of Query Processing Capabilities

- How to do queries that go beyond simple keyword search?
- How to search for stuff exploiting structure/schema information?
- How to join data from different files?

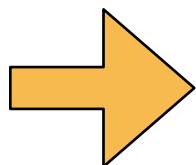Generate a mailing list for the projects…

Consequence: a Desktop is an Information Jungle
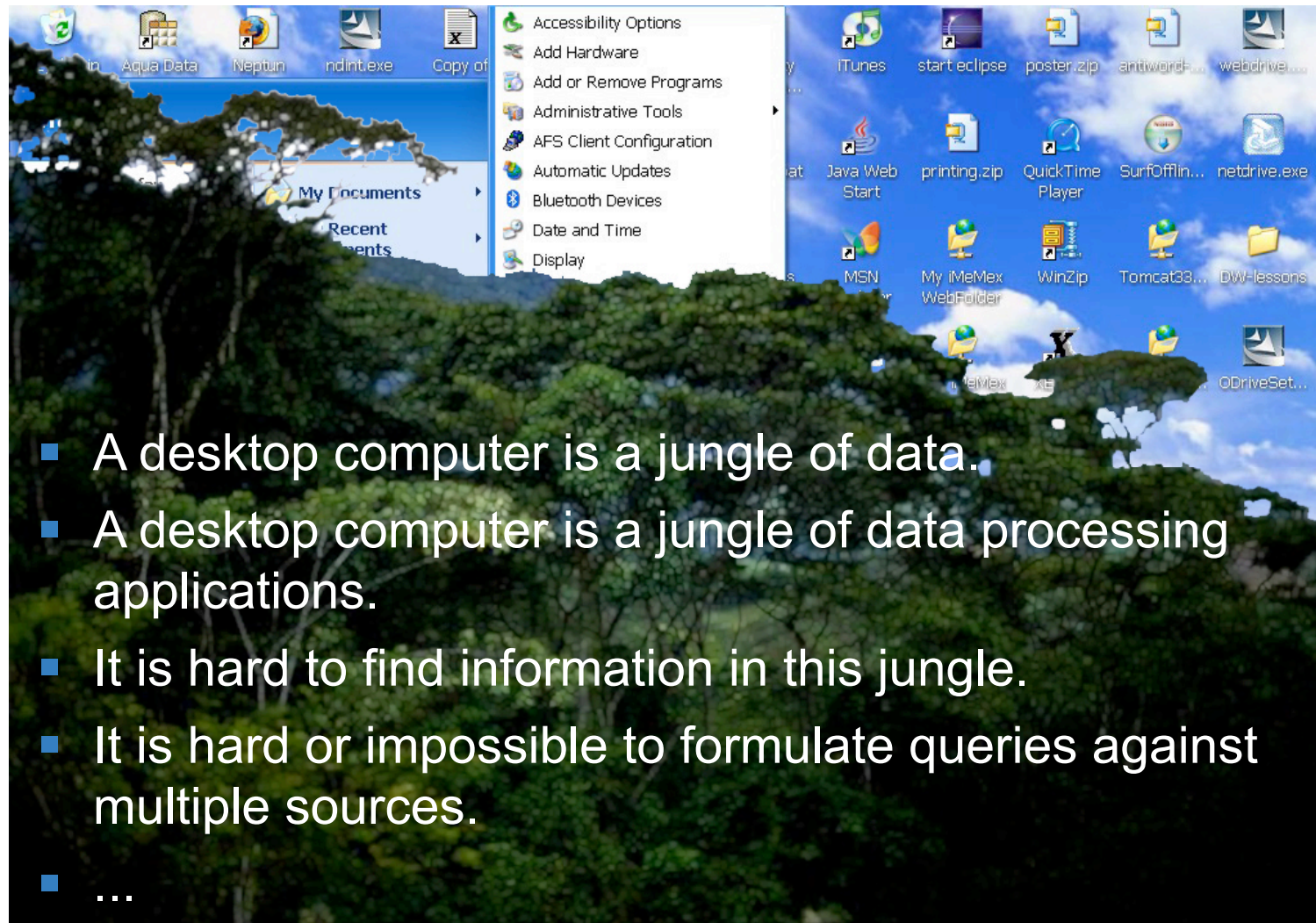
- A desktop computer is a jungle of data.
- A desktop computer is a jungle of data processing applications.
- It is hard to find information in this jungle.
- It is hard or impossible to formulate queries against multiple sources.
- ...

Our Solution: <u>integrated</u> memex

How does that work?

See you at the Demo Grounds!!!

Why is there a potato?