



KDI ● **Knowledge and Data Integration**

A World of Data

Phase: 1. Introduction &
Representation Diversity

W1.L12.M1.T211

Contents

1 The Web Data

2 Corporate Data

3 Diversity in the Web and Corporate Data

Contents

1 The Web Data

2 Corporate Data

3 Diversity in the Web and Corporate Data

Unstructured Data

Data without any metadata or structure.

- Text
- Figure (without metadata)
- Audio (without metadata)
- Video (without metadata)



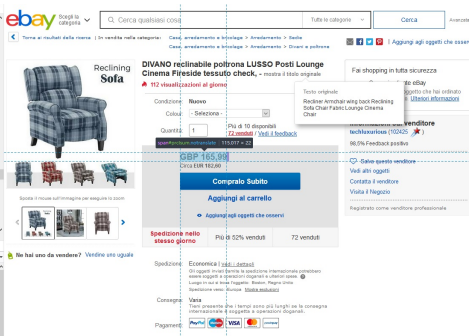
Unstructured data example: figures

Data with some metadata (such as tags but without formal semantics).

- ```

<script>
 $(function() {
 // 1. 为 .btn 添加 click 事件
 $(".btn").click(function() {
 // 2. 获取 .btn 的 text 属性
 var text = $(this).text();
 // 3. 判断 text 属性是否为 "显示"
 if (text === "显示") {
 // 4. 将 .btn 的 text 属性修改为 "隐藏"
 $(this).text("隐藏");
 } else {
 // 5. 将 .btn 的 text 属性修改为 "显示"
 $(this).text("显示");
 }
 });
 });

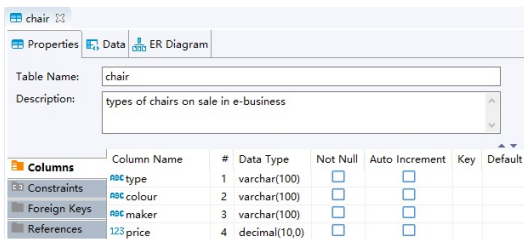
```



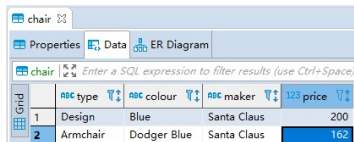
## Semi-structured data example: html file

# Structured Data

Data in formal structures (such as in related database tables).



|              | Column Name | # | Data Type     | Not Null                 | Auto Increment           | Key | Default |
|--------------|-------------|---|---------------|--------------------------|--------------------------|-----|---------|
| Columns      | ABC type    | 1 | varchar(100)  | <input type="checkbox"/> | <input type="checkbox"/> |     |         |
| Constraints  | ABC colour  | 2 | varchar(100)  | <input type="checkbox"/> | <input type="checkbox"/> |     |         |
| Foreign Keys | ABC maker   | 3 | varchar(100)  | <input type="checkbox"/> | <input type="checkbox"/> |     |         |
| References   | 123 price   | 4 | decimal(10,0) | <input type="checkbox"/> | <input type="checkbox"/> |     |         |



|   | ABC type | ABC colour  | ABC maker   | 123 price |
|---|----------|-------------|-------------|-----------|
| 1 | Design   | Blue        | Santa Claus | 200       |
| 2 | Armchair | Dodger Blue | Santa Claus | 162       |

Structured data example: schemas and records in database tables

# Contents

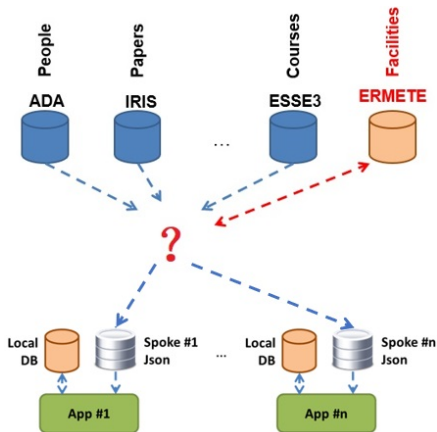
**1** The Web Data

**2** Corporate Data

**3** Diversity in the Web and Corporate Data

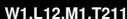
# Case Study

Digital University (DU) is an application running at the University of Trento which provides **extensive query answering services** across existing resources, such as ADA for people, IRIS for papers, ESSE3 for courses, ERMETE for faculties, etc.





Examples of the extensive queries across multiple legacy systems:  
*university profile generation, research activity explore, student  
 performance assessment, etc.*



# Contents

**1** The Web Data

**2** Corporate Data

**3** Diversity in the Web and Corporate Data

# Diversity Everywhere

The data on the web or within the corporate organization are represented in unstructured, semi-structured and/or structured fashions.

For the **same** entity in the real world, it can be represented ...

- in **different** languages.
- with **different** tags.
- with **different** predefined attributes in databases.

For the **different** entities in the real world, they can be represented ...

- with the **same** language symbols.
- with the **same** tags and tag values.
- with the **same** record values.

In conclusion, in the world of data, diversity is **everywhere**.



**W1.L12.M1.T211**



## **A World of Data**

Phase: 1. Introduction &  
Representation Diversity