Data Modeling

Foundational Curricula:
Cluster 8: Data
Module 15: Data Analytics, Modeling and Reporting
Unit 2: Data Modeling
FC-C8M15U2

Curriculum Developers: Angelique Blake, Rachelle Blake, Pauliina Hulkkonen, Sonja Huotari, Milla Jauhiainen, Johanna Tolonen, and Alpo Värri

44/60

This work is produced by the EU*US eHealth Work Project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 727552
EUUSEHEALTHWORK

Unit Objectives

- Define data modeling
- Define data representation
- Define primary data, secondary data, discrete data sources, data analysis, data modeling and data reporting
- Explain the different type of data models and systems including but not limited to distributed, centralized, relational, object oriented, warehouses/data marts
- Describe the principles of data representation, including data warehouses, clinical data repositories, etc.
- Identify primary and secondary uses of health and patient information and explain their role in data modeling

This work is produced by the EU*US eHealth Work Project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 727552
EUUSEHEALTHWORK

Data Modeling

- **Data modeling** means the process of creating models of the information by applying different techniques. Data model aims to define the structure and meaning of the data.
- Data models can be separated into different categories:
  - **Physical data models** organize the data into table and column structure that shows how the tables relate to each other
  - **Logical data models** includes all the attributes and relationship types required by the application it serves
  - **Conceptual data models** describe the semantics of the scope of the model

This work is produced by the EU*US eHealth Work Project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 727552
EUUSEHEALTHWORK

Terms and Definitions

- **Primary data** is collected by the one that uses it. **Secondary data** is collected by someone else than the one who uses it.
- **Discrete data sources** are stored to databases at the lowest level of granularity and the data is measurable and reportable. E.g. medication information (see the example figure)
- **Data analysis** is the process that aims to discover useful information, suggest conclusions and support decision making from data that is collected.
- **Data reporting** means collecting and submitting accurate data. Accurate data reporting makes it possible to conduct accurate data analysis.

Medication Name:	Prilosec
Dosage Qty:	1
Dosage Strength:	20
Dosage Units:	mg
Dosage Form:	Tablet
Frequency:	BID
Duration Number:	3
Duration Length:	Days

Non-discrete form: "*1 20mg tablet twice daily for three days*"

This work is produced by the EU*US eHealth Work Project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 727552 EUUSEHEALTHWORK

Data Models and Systems

Distributed database

- Distributed database consists of databases that are not physically connected but instead communicate via computer network

Centralized database

- Centralized database is a database that is located in a single location. All the data is located, stored and maintained in one place, for example in a central computer.

Relational data model

- Relational data model organizes data into tables that has rows and columns. Rows represent records and columns represent the attributes. Each row has a unique identifying key which is used to create relations between tables.

This work is produced by the EU*US eHealth Work Project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 727552
EUUSEHEALTHWORK

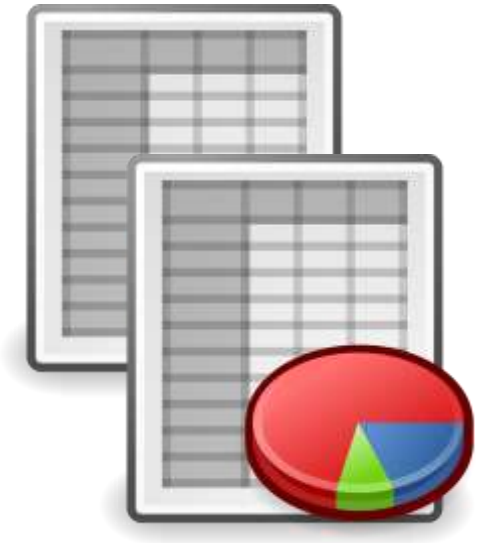
Data Models and Systems (cont'd)

Object oriented data model

- Object oriented data model is a data management system in which information is presented as objects. These objects are also used in object-oriented programming.

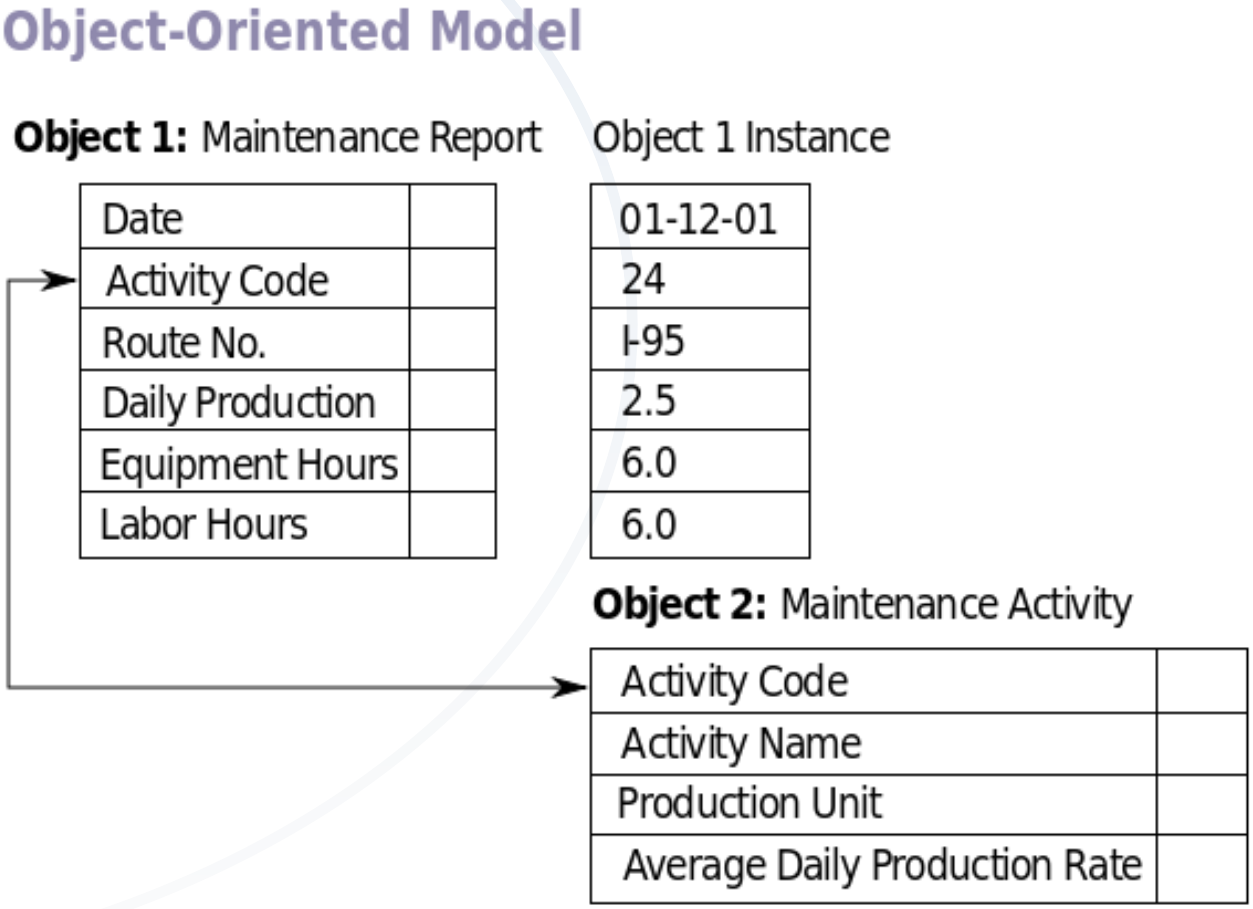
Data Warehouses

- Data Warehouses hold multiple subject areas and detailed information and they are used to get the data out to users. **Data marts** are a subset of warehouses, holding only one subject-area and more summarized data



This work is produced by the EU*US eHealth Work Project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 727552
EUUSEHEALTHWORK

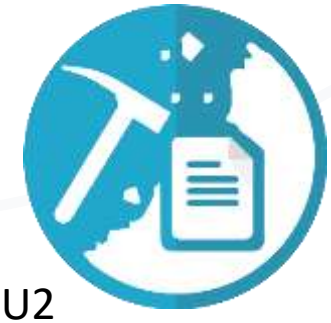
Example of an Object Oriented model



This work is produced by the EU*US eHealth Work Project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 727552
EUUSEHEALTHWORK

Data Representation Principles

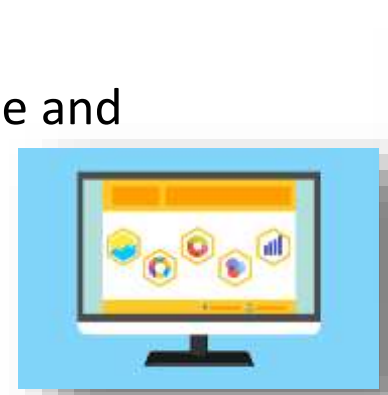
- **Data mining** is a potential method to use data to analyze and identify the best practices improving care and reducing costs. However, healthcare is slow to apply new methods and data mining includes methods that are not intuitively explained to the caregivers (e.g. decision trees in pattern recognition)
- **Clinical data repositories** (or also clinical data warehouses) are real time databases, that combine data from various sources and provide an unified view of a single patient
  - This could be used to provide a view of medications prescribed to the patient, thus reactive and allergenic medications are increasing



This work is produced by the EU*US eHealth Work Project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 727552 EUUSEHEALTHWORK

Primary and Secondary Uses of Health Information

- Primary use of health information is the use of information to give healthcare to the patient
- Secondary use of healthcare is when the data is already collected and stored, and it is used to health management planning, research or public health evaluation. For this, the data is mostly anonymized, thus only the exact information is used, not any personal information
  - Secondary uses are also disease surveillance and biomedicine/emerging medical technology



This work is produced by the EU*US eHealth Work Project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 727552 EUUSEHEALTHWORK

Unit Review Checklist

- ❑ Defined data modeling
- ❑ Defined data representation
- ❑ Defined primary data, secondary data, discrete data sources, data analysis, data modeling and data reporting (GL03)
- ❑ Explained the different type of data models and systems including but not limited to distributed, centralized, relational, object oriented, warehouses/data marts (GB07)
- ❑ Described the principles of data representation, including data warehouses, clinical data repositories, etc. (GB20)
- ❑ Identified primary and secondary uses of health and patient information and explained their role in data modeling (GB13)

This work is produced by the EU*US eHealth Work Project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 727552 EUUSEHEALTHWORK

Unit Review Exercise/Activity

1. Explain the difference between primary data and secondary data and their use in healthcare (give examples).
2. Give an example of healthcare information in non-discrete and discrete form.

This work is produced by the EU*US eHealth Work Project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 727552
EUUSEHEALTHWORK

Unit Exam

1. Data reporting aims to discover the right conclusions from collected data.
   a) True
   b) False
2. Data modeling means formatting information into pre-designed forms
   a) True
   b) False
3. Data warehouses include only one subject-area of information.
   a) True
   b) False

This work is produced by the EU*US eHealth Work Project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 727552
EUUSEHEALTHWORK

Unit Exam (cont'd)

4. Data models can be separated into different categories. The categories are:
   a) Logical, semantic and physical data model
   b) Physical, logical and conceptual data model
   c) Relational, semantic and physical data model
5. Data mining includes
   a) Identifying the best practices for reducing the costs
   b) Collecting identifiable databases of patients health care information
   c) Data warehouses that aim to hold the medication that a certain patient needs
   d) All of the above

This work is produced by the EU*US eHealth Work Project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 727552
EUUSEHEALTHWORK