



Co-financed by the Connecting Europe
Facility of the European Union



Relational databases

DDH - Module A - lecture 1

Enea Parimbelli - enea.parimbelli@unipv.it

Lecture summary

- Managing data
- Relational databases
- The relational data model
 - Relationships
 - Facts
 - Redundancy and duplication
- ~~ Normalization
- Designing a DB: the ER diagram

Data management

- Most computer systems need to handle data in a persistent manner
 - File-based approach
 - Data management software-based approach

The file-based approach

- No distinction between data and applications
- Security delegated to the operating system
- Problem: large amounts of data
- Problem: sharing and concurrent access

The structured approach

- Creation of a database
- Read/write accesses to data
- Data Sharing
 - Among different users
 - Among several applications
- Consistency of shared data
- Data Protection
- Reliability of data in case of failure

Relational databases

- The term Database denotes a logically grouped set of data (usually concerning the same topic, or several related topics), structured in such a way that the data can be used for different applications.
- In addition to the actual data, the database must also contain information about their representations and the relationships between them. There may also be data structures (indexes) that speed up frequent operations, typically at the expense of less frequent operations.

Database management system

- The Database Management System (DBMS) is that system that allows data to be "managed," that is, stored, retrieved, checked for correctness, etc. Among other things, they are manage:
 - Backup
 - Security and access, which authorizes only certain user profiles to perform certain operations on certain types of data.
 - Programs that are executed, either automatically or at the request of authorized users, to perform processing on data. A typical automatism is to run a program every time a data item of a certain type is changed
 - Links to external data, i.e., references to local or remote files that are not part of the database.

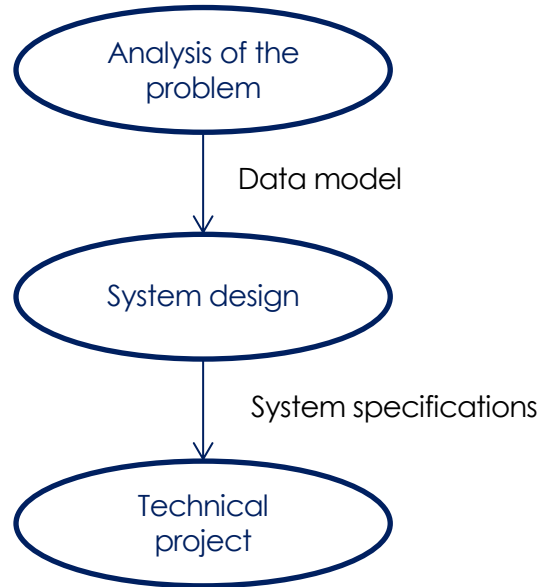
RDBMS vendors



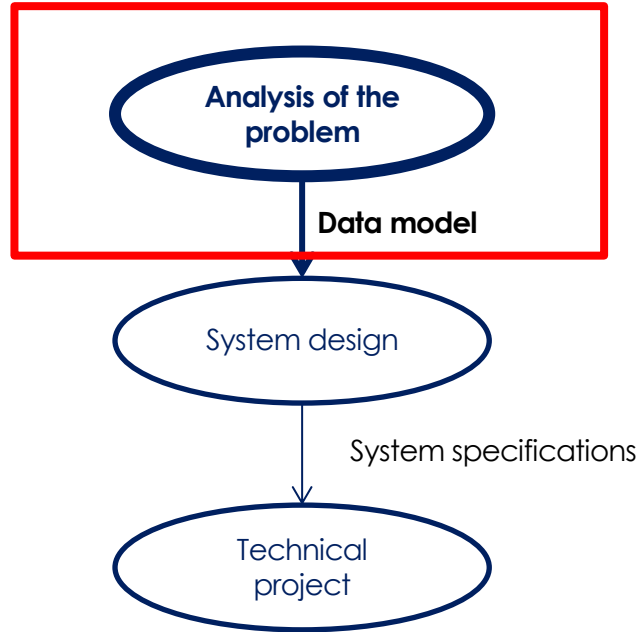
FileMaker.



Database design



Database design



The relational model

- Theory '70s - Implemented by early '80s
- Use tabular structures (tables) as a natural way to store data
- Provides procedures for choosing such facilities in the best possible way
- A relational model can easily be translated into a logical database definition
- Relationships provide a good means of communication between user and designer.

The reports

- A relationship is a table in which the columns represent variables and the rows the values assumed by those variables (the cases).
- In practice, a table usually represents a set of logically related variables.

Relationships - example

PATIENT

Folder No.	Tax code	Department	Days of hospitalization in the department
101/A	GTASDU23S12D182F	PEDIATRICS	32
1200/B	HGUIUH56B569873H	SURGERY	15
...

DEPARTMENT

Department	Head of unit	# beds	# doctors
OBSTETRICS	RED GIUSEPPE	20	11
PEDIATRICS	LUIGINA VERDI	40	18
...

Relationships - terminology

- Relationship = Table
- Tuple = table row
- Domain = set of values that can be associated with each attribute of the relationship (numeric, text, code...)

Relationships - terminology

Relation/table name

Attributes

PATIENT

Folder No.	Tax code	Department	Days of hospitalization in the department
101/A	GTASDU23S12D182F	PEDIATRICS	32
1200/B	HGUIUH56B569873H	SURGERY	15
...

Domains

Tupla

Relationships - terminology

Relationship/table name

Attributes

PATIENT

Folder No.	Tax code	Department	Days of hospitalization in the department
101/A	GTASDU23S12D182F	PEDIATRICS	32
1200/B	HGUIUH56B569873H	SURGERY	15
...

Domains

Tuple

Relationships - rules

- Relationships are logical, nonphysical representations of data. They must be unambiguous and explanatory of the data themselves:
 - In a table, there are no duplicate rows

Relationships - rules

PATIENT

Folder No.	Tax code	Department	Days of hospitalization in the department
101/A	GTASDU23S12D182F	PEDIATRICS	32
1200/B	HGUIUH56B569873H	SURGERY	15
1200/B	HGUIUH56B569873H	SURGERY	15
...

Relationships - rules

- Relationships are logical, nonphysical representations of data. They must be unambiguous and explanatory of the data themselves:
 - In a table, there are no duplicate rows
 - The order of rows and columns is not important

Relationships - rules

PATIENT

Folder No.	Tax code	Department	Days of hospitalization in the department
1200/B	HGUIUH56B569873H	SURGERY	15
101/A	GTASDU23S12D182F	PEDIATRICS	32
...

PATIENT

Folder No.	Tax code	Department	Days of hospitalization in the department
101/A	GTASDU23S12D182F	PEDIATRICS	32
1200/B	HGUIUH56B569873H	SURGERY	15
...

Relationships - rules

- Relationships are logical, nonphysical representations of data. They must be unambiguous and explanatory of the data themselves:
 - There are no duplicate rows in a table
 - The order of rows and columns is not important
 - Each column has a different name from all the others.

Relationships - rules

PATIENT

Folder No.	Department	Department	Days of hospitalization in the department
101/A	GTASDU23S12D182F	PEDIATRICS	32
1200/B	HGUIUH56B569873H	SURGERY	15
...

Relationships - rules

- Relationships are logical, nonphysical representations of data. They must be unambiguous and explanatory of the data themselves:
 - In a table, there are no duplicate rows
 - The order of rows and columns is not important
 - Each column has a different name from all the others.
 - Finally, it is advisable to use labels, for attributes, that are useful in identifying the meaning of the attribute itself

Relationships - rules

PATIENT

A	B	C	D
101/A	GTASDU23S12D182F	PEDIATRICS	32
1200/B	HGUIUH56B569873H	SURGERY	15
...

The universal relationship

- A single report with all attributes
- REDUNDANCY
 - Waste of memory
 - Poor CONSISTENCY
- Consistency is automatically maintained if each row is independent of the others

1	Paziente	D.N.	Sesso	Richiesta	Data prelievo	Corr.	
71	[REDACTED]	12/07/1948	Maschio	090176083	18/09/2015	LABO	0,56
72	[REDACTED]	23/07/1964	Femmina	020091955	11/02/2015	LABO	0,02
73	[REDACTED]	26/07/1999	Femmina	110057389	12/11/2014	LABO	0,01
74	[REDACTED]	29/10/1954	Femmina	080021117	07/08/2014	LABO	0,02
75	[REDACTED]	23/01/1979	Femmina	010215384	04/01/2016	LABO	0
76	[REDACTED]	22/11/1987	Femmina	010217729	11/01/2016	LABO	0
77	[REDACTED]	02/08/1968	Femmina	100191624	28/10/2015	LABO	0
78	[REDACTED]	01/03/2002	Maschio	070011154	08/07/2014	LABO	0,04
79	[REDACTED]	08/04/1977	Femmina	010075960	02/01/2015	LABO	0,01
80	[REDACTED]	21/06/1971	Femmina	010081009	15/01/2015	LABO	0,02
81	[REDACTED]	10/07/1998	Femmina	070150159	02/07/2015	LABO	0,03
82	[REDACTED]	18/09/1967	Femmina	080023081	13/08/2014	LABO	0,02
83	[REDACTED]	03/10/1969	Maschio	120214906	31/12/2015	LABO	0,01
84	[REDACTED]	07/04/1979	Maschio	070159608	30/07/2015	LABO	0,01
85	[REDACTED]	12/09/1926	Femmina	040113123	01/04/2015	LABO	0,01
86	[REDACTED]	11/06/1988	Femmina	040118713	15/04/2015	LABO	0,02
87	[REDACTED]	23/07/1969	Femmina	060144102	17/06/2015	LABO	0,01
88	[REDACTED]	29/05/1967	Maschio	120213927	28/12/2015	LABO	0
89	[REDACTED]	09/02/1994	Maschio	010222830	21/01/2016	LABO	0,02
90	[REDACTED]	10/07/1958	Femmina	110056608	10/11/2014	LABO	0
91	[REDACTED]	13/04/1965	Femmina	120208384	10/12/2015	LABO	0,02
92	[REDACTED]	26/07/1967	Maschio	060138362	03/06/2015	LABO	>100
93	[REDACTED]	16/08/1971	Femmina	050128900	11/05/2015	LABO	1,07
94	M	24/07/2008		070074		RHO	0,01

(In)consistency - Example

Blood sampling

Last Name and First Name	Sex	ID	Date of birth	Sampling date	Type of sampling	Source	Measure 1
Mario Rossi	M	1001	20/12/1982	20/9/2015	A	outpat	12
Luca Bianchi	0	1002	March 8, 76	20/9/2015	B	inpatient	11*
Silvia Verdi	1	1003	1955 Nov 23	23/9/2015	A	inpatient	3
Mario Rossi	M	1004	20/11/1982	28/9/2015	A	outpat	54
Silvia Verdi	1	1003	1955 Nov 23	4/10/2015	B	outpat	>20

The facts

- A fact exists when the value of one attribute determines at least the value of another attribute.
 - Ex: Patient ID determines Name, Date of birth, ...
- There can be facts with multiple values
 - Ex: Patient ID determines more than one Phone, Visit, Diagnosis, ...

Derived facts

- Facts can be divided into basic facts and derived facts.
- Ex:
 - patient P120 was born on 28-10-1992 (basic fact)
 - patient P120 was born on a Wednesday (derived fact)
- Derived facts can be derived from the basic facts
- A good database should not contain derived facts

Redundancy

- A database may contain redundancy for two reasons:
 1. derived facts are stored
 2. the same fact is stored multiple times

Redundancy (1)

derived facts are stored

Patient	Department
D	OBSTETRICS
A	PEDIATRICS
...	...

head	Phone
Angels	503674
Reds	502567
...	...

head	Department
Reds	OBSTETRICS
Angels	PEDIATRICS
...	...

Patient	head physician tel.
D	502567
A	503674
...	...

Redundancy (1)

Patient	Department
D	OBSTETRICS
A	PEDIATRICS
...	...

Primary	Phone
Angels	503674
Reds	502567
...	...

Primary	Department
Reds	OBSTETRICS
Angels	PEDIATRICS
...	...

Patient	Primary caregiver's tel.
D	502567
A	503674
...	...

- the last table is redundant, since the telephone of the head physician of the ward where a certain patient is admitted can be **derived** from the other three tables.

Redundancy (1)

Patient	Department	Primary	Phone
D	OBSTETRIC	Angels	503674
A	PEDIATRIC	Reds	502567

Primary	Department	nt	Primary caregiver's tel.
Reds	OBSTETRICS		502567
Angels	PEDIATRICS		503674
...	...		
	



- Try to think, in the presence of the fourth table, how to update the database in the following cases:
 - a patient changes departments
 - one department changes head medical officer
 - a head of department changes phone number

Redundancy (2)

the same fact is stored multiple times

Patient	Department	Head	# beds
RSSGVN56G56...	SURGERY	Rossi Antonio	30
BNCPRO45H67...	OBSTETRICS	Luigi Verdi	225
FDASRE54N78...	PEDIATRICS	Marini Mario	34
QRSEWS76B34...	SURGERY	Rossi Antonio	30
...
UOASPE53B79	PEDIATRICS	Marini Mario	34

Redundancy (2)

Patient	Department	Primary	# beds
RSSGVN56G56...	SURGERY	Rossi Antonio	30
BNCPRO45H67...	OBSTETRICS	Luigi Verdi	225
FDASRE54N78...	PEDIATRICS	Marini Mario	34
QRSEWS76B34...	SURGERY	Rossi Antonio	30
...
UOASPE53B79	PEDIATRICS	Marini Mario	34

- For each patient admitted to the same ward, the name of the head physician and the number of beds are repeated
- Redundancy makes it difficult to update the database (e.g., change of dept. head)

Redundancy (2)

Patient	Department	Primary	# beds
RSSGVN56G56...	SURGERY	Rossi Antonio	30
BNCPRO45H67...	OBSTETRICS	Luigi Verdi	225
FDASRE54N78...	PEDIATRICS	Marini Mario	34
QRSEWS76B34...	SURGERY	Rossi Antonio	30
...
UOASPE53B79	PEDIATRICS	Marini Mario	34

- In case a ward is without inpatients, a line such as the following would be needed

Patient	Department	Primary	# beds
	OCULISTICS	Giorgini Giorgio	6

Redundancy (2)

Patient	Department	Primary	# beds
RSSGVN56G56...	SURGERY	Rossi Antonio	30
BNCPRO45H67...	OBSTETRICS	Luigi Verdi	225
FDASRE54N78...	PEDIATRICS	Marini Mario	34
QRSEWS76B34...	SURGERY	Rossi Antonio	30
...
UOASPE53B79	PEDIATRICS	Marini Mario	34

- In the case where a ward is without inpatients, a line such as the following would

Patient	Department	Primary
		Giorgina

Incomplete tuple!!!

Difficult update!!!

Getting rid of redundancies

- Relationships that contain derived facts are removed
- Relationships that store the same fact more than once are decomposed, as we will see next

Duplication

- Duplication does NOT mean redundancy
- Duplication is sometimes necessary

RECIPIENTS

Patient	Department	Date of birth	Date of admission
RSSGVN56G56...	SURGERY	12-12-1975	12-11-1996
BNCPRO45H67...	OBSTETRICS	13-11-1970	12-08-1997
FDASRE54N78...	PEDIATRICS	20-02-1990	20-08-1997
RSSGVN56G56...	SURGERY	12-12-1975	30-05-1997
...
UOASPE53B79	PEDIATRICS	23-05-1993	23-05-1993

Duplication

- Duplication does NOT mean redundancy
- Duplication is sometimes necessary

RECIPIENTS

Patient	Department	Date of birth	Date of admission
RSSGVN56G56...	SURGERY	12-12-1975	12-11-1996
BNCPRO45H67...	OBSTETRICS	13-11-1970	12-08-1997
FDASRE54N78...	PEDIATRICS	20-02-1990	20-08-1997
RSSGVN56G56...	SURGERY	12-12-1975	30-05-1997
...
UOASPE53B79	PEDIATRICS	23-05-1993	23-05-1993



Duplication



Redundancy

Duplication

What do we do here?

RECIPIENTS

Patient	Department	Date of birth	Date of admission
RSSGVN56G56...	SURGERY	12-12-1975	12-11-1996
BNCPRO45H67...	OBSTETRICS	13-11-1970	12-08-1997
FDASRE54N78...	PEDIATRICS	20-02-1990	20-08-1997
GBTMTT83T20...	SURGERY	12-12-1975	31-11-1999
...
UOASPE53B79	PEDIATRICS	23-05-1993	23-05-1993

Normalization

- Goal: eliminate redundancy
- 2 concepts:
 - Dependence between two objects
 - A value of the former corresponds to one or more values of the latter
 - Ex: COD_PAT -> DATE_BIRTH.
COD_PAT, DATE_ADM -> DEPARTMENT
COD_PAT -> -> VACCINATIONS (Multivalue)
 - Relationship keys
 - The attribute, or set of attributes, whose value uniquely identifies a row in the table.

The relationship keys

- The attribute, or set of attributes, whose value uniquely identifies a row in the table.
- Properties:
 - a given set of attribute values belonging to the key (this set is called the key value) identifies only one row of the relationship; in other words, no two rows can exist with the same key value
 - no subset of the key attributes still forms a key; in other words, the key is the minimum set of attributes needed to uniquely identify a row in the report
 - key attributes cannot take null values

Normalization - example

VISITS

Date of visit	Temp	Blood pressure	Name	Last name	Date of birth	Sex
15-10-2015	37.6	120/80	Mario	Reds	28-04-1965	M
18-10-2015	35.9	90/60	Laura	White	13-12-1984	F
22-10-2015	36.9		Mario	Reds	28-04-1965	M
24-10-2015		100/70	Laura	White	13-12-1984	F

Normalization - example

	Date of visit	Name	Last name	Temp	Blood pressure
VISITS	15-10-2015	Mario	Reds	37.6	120/80
	18-10-2015	Laura	White	35.9	90/60
	22-10-2015	Mario	Reds	36.9	
	24-10-2015	Laura	White		100/70

	Name	Last name	Date of birth	Sex
PATIENTS	Mario	Reds	28-04-1965	M
	Laura	White	13-12-1984	F

K

VISITS

Date of visit	Name	Last name	Temp	Blood pressure
15-10-2015	Mario	Reds	37.6	120/80
18-10-2015	Laura	White	35.9	90/60
22-10-2015	Mario	Reds	36.9	
24-10-2015	Laura	White		100/70

K

PATIENTS

Name	Last name	Date of birth	Sex
Mario	Reds	28-04-1965	M
Laura	White	13-12-1984	F

Normalization - example

VISITS

Date of visit	Patient Id	Temp	Blood pressure
15-10-2015	PZ001	37.6	120/80
18-10-2015	PZ002	35.9	90/60
22-10-2015	PZ001	36.9	
24-10-2015	PZ002		100/70

PATIENTS

Patient Id	Name	Last name	Date of birth	Sex
PZ001	Mario	Reds	28-04-1965	M
PZ002	Laura	White	13-12-1984	F

K

VISITS

Date of visit	Patient Id	Temp	Blood pressure
15-10-2015	PZ001	37.6	120/80
18-10-2015	PZ002	35.9	90/60
22-10-2015	PZ001	36.9	
24-10-2015	PZ002		100/70

K

PATIENTS

Patient Id	Name	Last name	Date of birth	Sex
PZ001	Mario	Reds	28-04-1965	M
PZ002	Laura	White	13-12-1984	F

Normalization - example

MEASUREMENTS

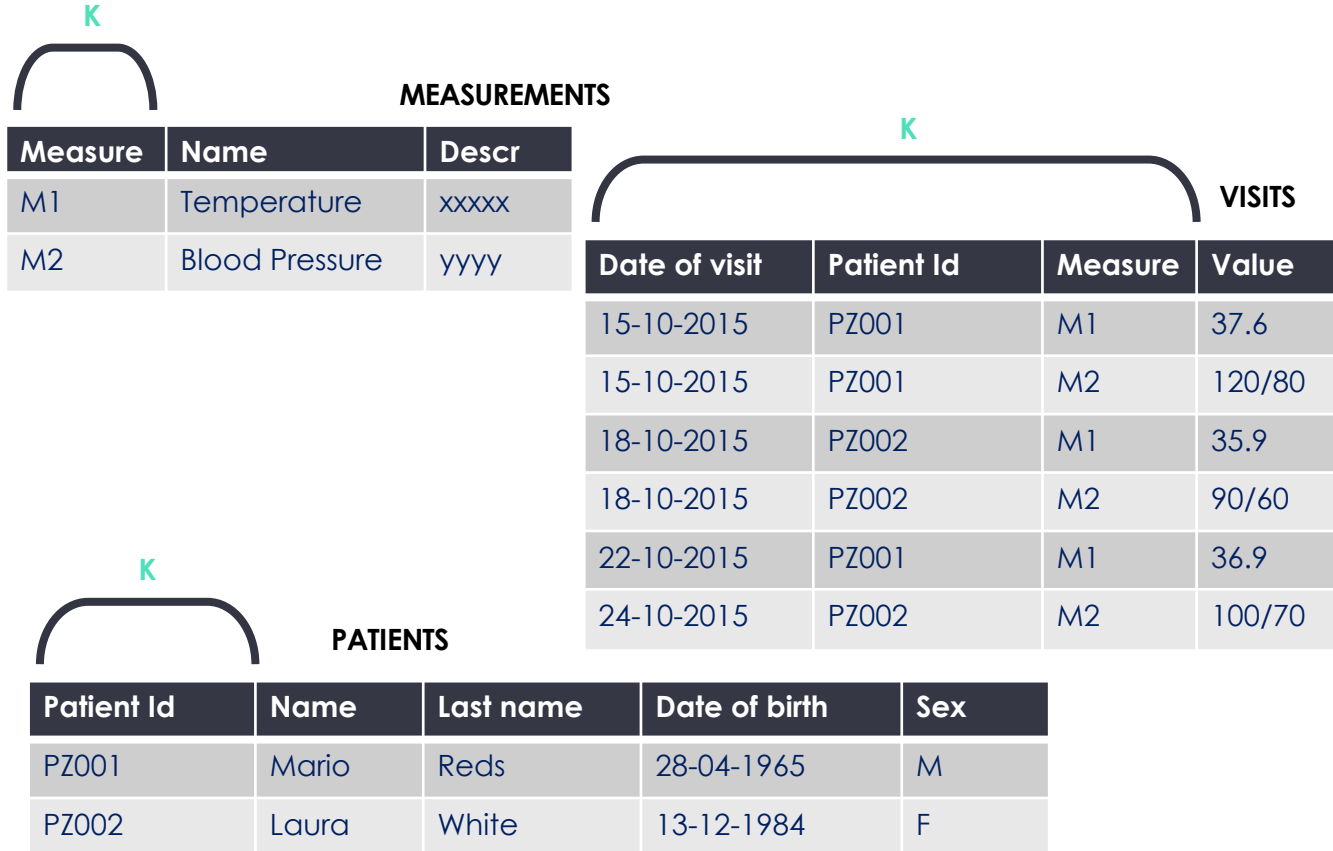
Measure	Name	Descr
M1	Temperature	xxxxx
M2	Blood Pressure	yyyy

VISITS

Date of visit	Patient Id	Measure	Value
15-10-2015	PZ001	M1	37.6
15-10-2015	PZ001	M2	120/80
18-10-2015	PZ002	M1	35.9
18-10-2015	PZ002	M2	90/60
22-10-2015	PZ001	M1	36.9
24-10-2015	PZ002	M2	100/70

PATIENTS

Patient Id	Name	Last name	Date of birth	Sex
PZ001	Mario	Reds	28-04-1965	M
PZ002	Laura	White	13-12-1984	F



Example (homework)

Assignment – look
@course web

We want to create a database for patients admitted to the surgical department who undergo two different types of surgery (Intervention1 and Intervention2).

The purpose of this database, in addition to maintaining an up-to-date and easily searchable archive, is also to perform statistical analyses to compare the outcomes of interventions. Intervention1 is riskier in terms of intraoperative mortality, while Intervention2 carries no risk of death, but is suspected to be less effective.

The success of the surgery is measured during a follow-up visit, in which the patient is monitored by ultrasound and certain symptoms (PROs), such as pain, are collected.

Example cont.d (homework)

The database should contain the following information:

- patient demographic data
- Data on the surgery, such as the type, the surgeon who performed it, the length of hospitalization, and any complications, including death
- data on follow-up visits, specifically the result of any ultrasound examination in terms of the number of lesions found, size, and location. In addition, the presence and type of symptoms, if any, reported by the patient should be recorded.
- one or more medications can also be prescribed during the visit, characterized by the name of the drug, its ATC code and recommended dosage.

Example 2 (looking ahead)

- Observational study of patients given pre-PCI femoral vessel echo with ultrasound endpoint (plaque yes/no, quantification and site)
- Clinical endpoint: complications at access site
- Design a database in which to collect the study data
 - What information do we want to store?
 - How many and what tables could be included?

Patient Id	Name	Last name	Date of birth	Sex
PZ001	Mario	Reds	28-04-1965	M
PZ002	Laura	White	13-12-1984	F

Patient Id	Exam Date	Operator	Plaque Presence	Thickness	Site	Indication to the procedure
PZ001						
PZ002						

Patient Id	Date PCI	Type of access
PZ001		
PZ002		

Patient Id	Date Complication	Type Complication
PZ001		
PZ002		



Master Data Form

Patient Id
Name
Last name
Date of Birth
Sex

Form EcoFemoralVessels

Exam Date
Operator
Plaque Presence
Thickness
Headquarters
Indication to the Procedure


Form PCI

Date PCI
Type Access

Complications Form

Date Complication
Type Complication





UniPV
University of Pavia - Department of Electrical, Computer and Biomedical Engineering

Logged in as **lucia.sacchi** | [Log out](#)

My Projects

- [Project Home](#)
- [Project Setup](#)

Project status: Development

Data Collection [Edit instruments](#)

- [Record Status Dashboard](#)
- View data collection status of all records
- [Add / Edit Records](#)
- Create new records or edit/view existing ones

Data Collection Instruments:

- Anagrafica
- EcoVasiFemorali
- PCI
- Complicanza

Applications

- [Calendar](#)
- [Data Exports, Reports, and Stats](#)

Studio Complicanze

[Record Status Dashboard \(all records\)](#)

Displayed below is a table listing all existing records/responses and their status for every data collection instrument (and if longitudinal, for every event). You may click any of the colored buttons in the table to open a new tab/window in your browser to view that record on that particular data collection instrument. Please note that if your form-level user privileges are restricted for certain data collection instruments, you will only be able to view those instruments, and if you belong to a Data Access Group, you will only be able to view records that belong to your group.

Displaying record through of 3 records

Displaying: [Instrument status only](#) | [Lock status only](#) | [All status types](#)

Record ID	Anagrafica	EcoVasiFemorali	PCI	Complicanza
1				
2				
3				

The E-R Diagram

- It is a mean of communication between the database designer and the domain expert
- Serves to ensure that all aspects of the domain, data and interactions, are understood and represented
- Entities and relationships among entities considered "important" in the domain under consideration are then represented in a formalized graphical representation

The E-R Diagram



The E-R Diagram

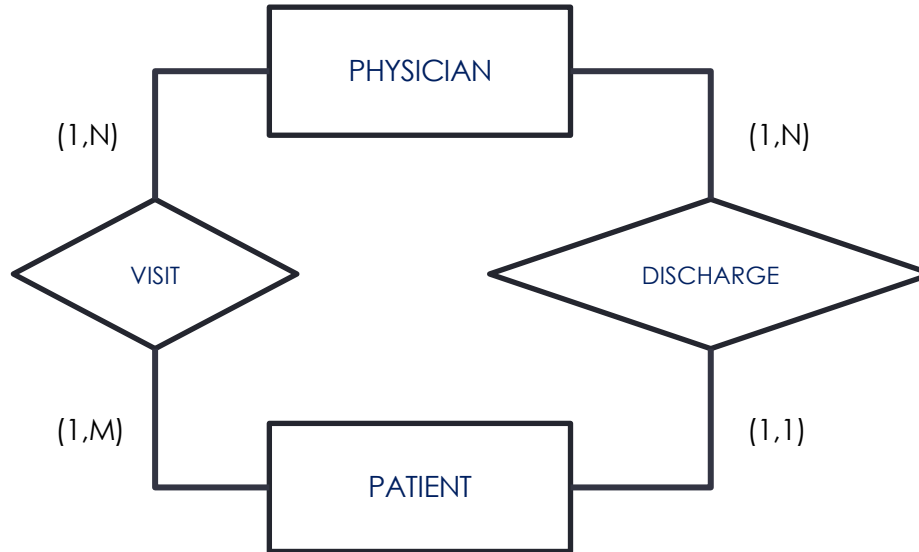
- Entity sets are represented with rectangles labeled by nouns
- Sets of relationships are represented with rhombuses labeled by verbs or prepositions indicating actions
- Some authors prefer to use nouns even for relations so as not to induce a “verb” -> BEWARE THIS IS NOT A USE-CASE or ACTIVITY diagram!!!!



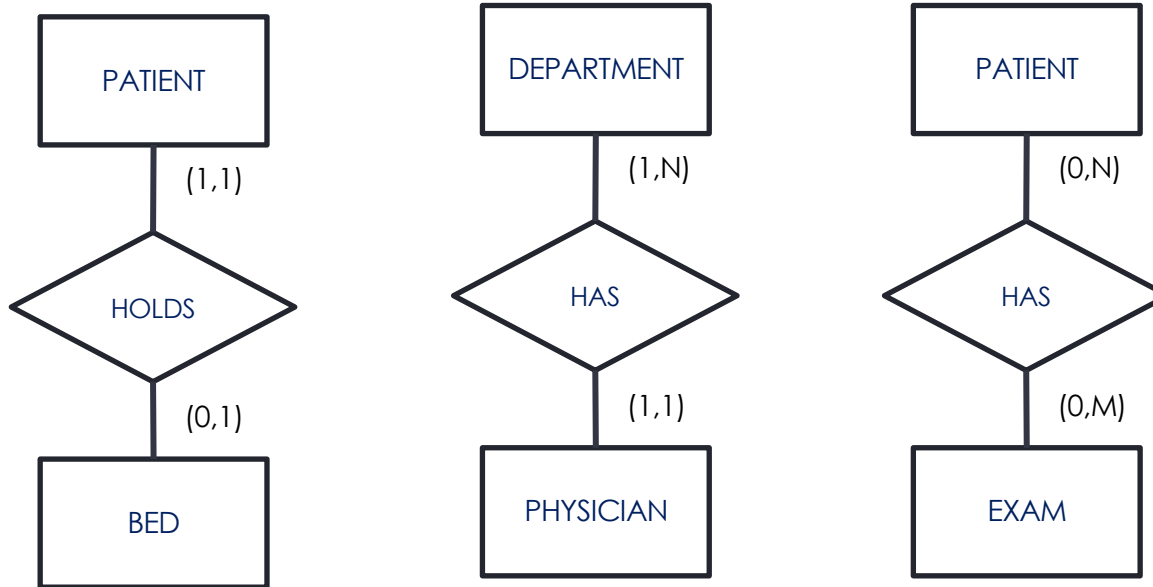
Cardinality

- Minimum Cardinality is the minimum number of times an instance of an entity can be involved in a fact of a relationship;
 - if the minimum cardinality is 0, it means that there can be an instance of the entity not involved in any fact of the relationship;
 - a value of 1 (in general N) means that there is no instance of the entity that is not involved in at least 1 (or N) facts of the relationship
- Maximum Cardinality is the maximum number of times an instance of an entity can be involved in a fact of a relationship

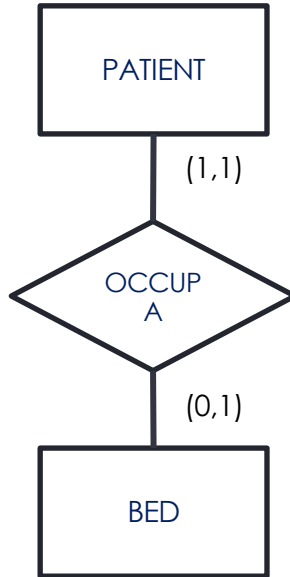
Cardinality



Cardinality - Examples



Cardinality - Examples



If there is a relationship between two sets of entities, it is not necessarily the case that there must exist for every entity an instance of relationship with another entity. In this case, participation in the relationship is said to be **OPTIONAL**.

The 'optionality' is detected directly from the minimum cardinality (0,...).

For example, while a patient for sure occupies a bed (in case inpatients only are being represented), a bed does not necessarily have to be occupied by a patient at all times (in fact, there are generally vacant beds in a ward).

The Identifiers

- Entities and relationships are characterized by attributes. They are indicated next to rectangles and rhombuses.
- Attributes include IDENTIFIERS (an attribute or set of attributes), which uniquely identify the entity or relationship. -> memento the “key”
- The attributes of a relationship certainly include the identifiers of the entities being related, plus other attributes that describe the properties of the interaction (relationship) itself between the entities.

The Identifiers

Identifiers are usually underlined.



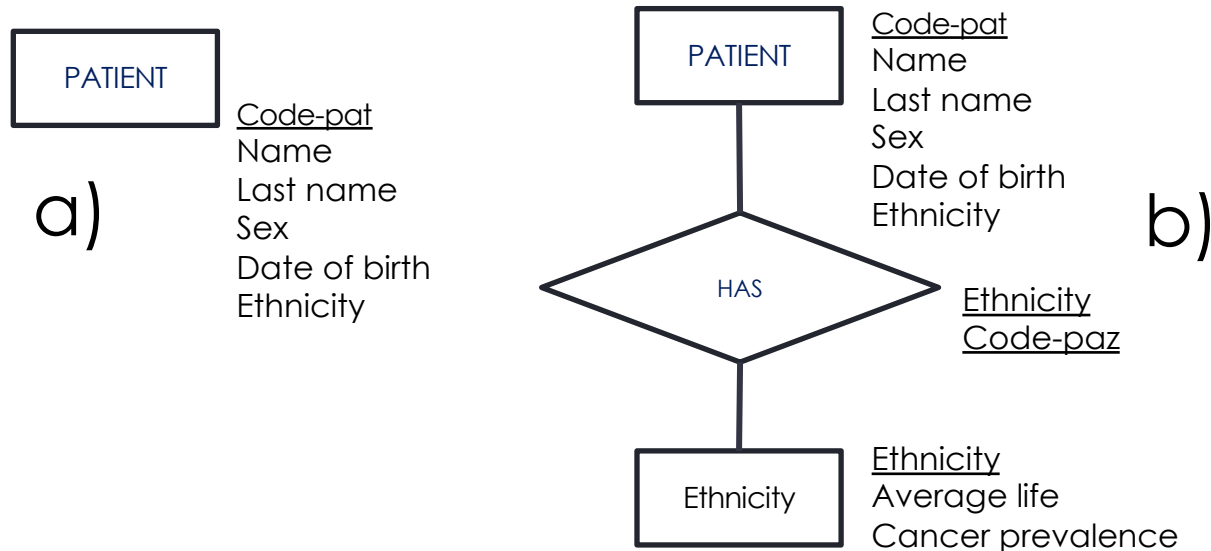
Code-pat
Name
Last name
Sex
Date of birth
Ethnicity

Code-pat
No. of bed
Inpatient days

No. of bed
Physician
Responsible
Room

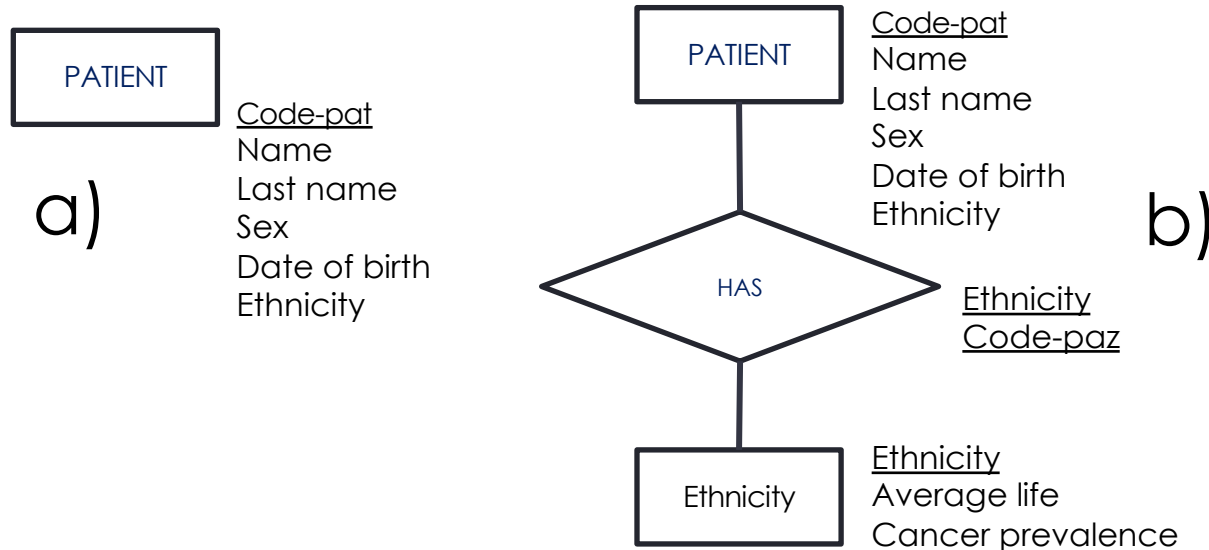
The Identifiers

- Typically, the relationship identifier is composed of the identifiers of the related entities.
- The choice between modeling a certain concept as an entity or as an attribute of an entity is not always trivial.



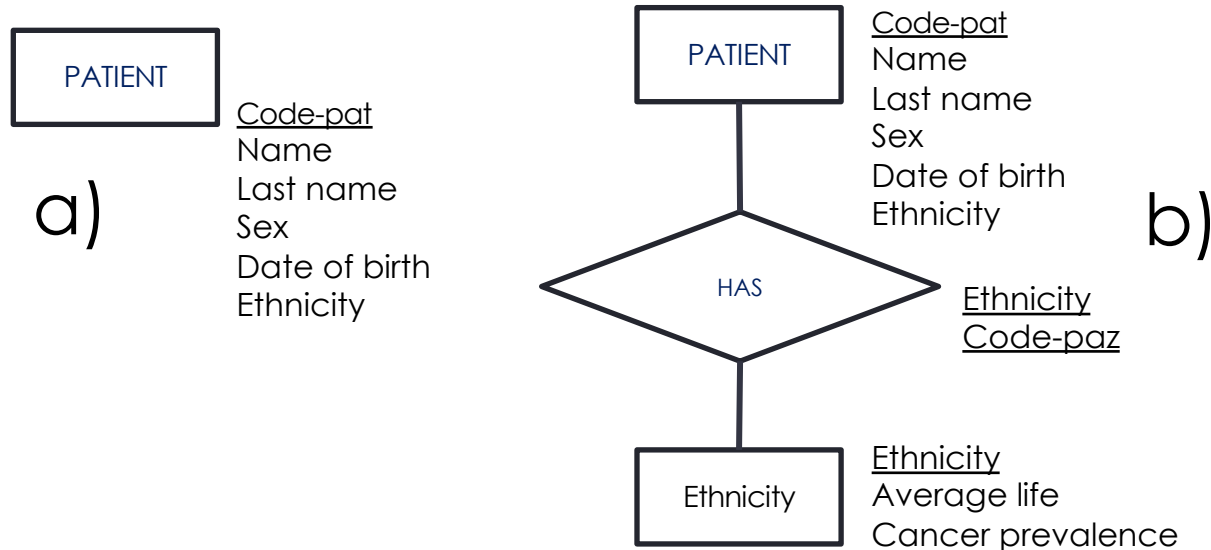
The Identifiers

- In case (a), ethnicity is simply an attribute of "patient," at the same level as first name, last name, etc.
- In case (b), on the other hand, ethnicity was represented as an entity, since it is considered important not only as a characteristic of a patient, but in that it itself possesses attributes that need to be stored.



The Identifiers

In this second case (b), each ethnicity is associated with the corresponding average life span and the number of individuals belonging to it. Thus, if information on ethnicity and its impact on cancer prevalence is important regardless of whether or not there is an individual of that ethnicity in the database, model (b) should be adopted, otherwise model (a) is sufficient.



Assignment

Assignment – look
@course web

Construct an E-R diagram to represent the following context within a hospital with several departments:

- A department has several laboratories, each identified by a number.
- Each lab has several machines and each machine 3 operators, who take turns over the course of 24 hours.
- One operator can work on more than one machine.
- Each operator must be on call at one phone number at all times.
- Different machines perform different examinations.
- You want to be able to report the number of examinations performed each day.