Data Management for Health Data

The main motor, in Italy, with universal healthcare, is **cost management** .

There is a flow of information in an hospital, mostly vertical, while the movement of the patient is horizontal.

- emergency room
- acceptation in "reparto"
- "ricovero"
- exams
- therapy
- "dimissione" (let go the patient)

For every step, DRG (*diagnostic related group*) is evaluated. It comprehends the calculation of relative costs of the patients, to calculate reimbursement(?) and standard tarifss, with all accessory costs.

DRGs are an example of **medical dictionary**, standard codes for operations that have descriptions and details. Sometimes different application of the same methodology can have the same code, and that could be a mess.

What are medical data?

A **medical data** is a single observation on a patient. They are needed for medical decision making.
Necessary to define diagnosis, need of additional informations, treatments (actions).

Every medical activity **produces** data, **analyzes** data, **uses** data. Every medical data is collected by an operator, its value depends from the condition of its collection.

The **clinical folder** is born to track those data, because the medic is never alone.

Data types

Some are **numeric fields**. Could be **time series** (ECG, etc). Could be **images** (2D, 3D) or **stacks of images** (CT) and **movies** (eco) or just long **free text**.

Our main problem in the medical field is the heterogenicity of data. The purposes of data management are:

- recording (for clinical history and legal point of views)
- communication (between different reparti, coordinating different professionals)
- more levels: risk assessment (preventions, diagnosi precoce, unattended behaviours)
- research: clinical, epidemiological..

Research is prospective or retrospective.

from data to knowledge

From data, knowledge must be distilled. It derives from the analysis of the data, of its context, in the light of formal studies, assumptions, heuristics, and models. May require more observations. And so the aim is not just data storage, but the transformation into knowledge.

When registering a data, context must be included. **Uncertainty** must be taken into consideration.

There are many unreliability sources: data reliability:

- reported from patient?
- "translated" by a clinic (e.g. anamnesis)
- measured by machines that has an error (systematic, random, artifacts..)

Data modifications:

- patient conditions
- recording conditions
- healthy/diseased thresholds that can change over time

Intra-subjective variability. Multiple actors can be involved. All these factors can influence the data. A missing data is different from a wrong data, that can be worse.

Medical data property

Differently from, (i.e.), banking data, the patient is the owner of the data, but he's never the **user**. Medical operators are the user. **Medical data needs to be shared** between different people and systems. Online unsecure communications are a nightmare (mother that sends a picture to the medic by Whatsapp), and privacy is a huge issue. Some things must be ensured:

- Integrity
- Privacy
- Responsibility
- Authenticity
- Security and backup
- Continuity (disponibility)

Good data derives from good data collection procedures. The digitalization of clinical data has many advantages: there is more control on data (accessibility, availability, update capabilities). Redundancies are eliminated and data flow is more efficient. Data can be reused (for research), mined, and used in DDSs (*decision support systems*).

Knowledge bases

Beyond databases, that are structured collections of uninterpreted data, knowledge bases aggregates facts, heuristics, models, that can be use for problem solving and data analysis. The added value is given by interpretative rules.

informative systems

Informations are generated, stored, distributed and used. Only storage and distribution are pertinent issue for medical informatics. All medical processes can **query** a common DBMS (*database management system*) for CRUD (create, read, update, delete) operations.

Medical informatics in Italy

2012, start of AgID (Agenzia per l'Italia Digitale), that created multiple common infrastructures (SPID, storage, cloud, connectivity, etc). The basics is the **Fasciscolo Sanitario Elettronico** (FSE) that maximises interoperability.

Databases

(just get the slides from the Database course)