Data-ownership

RomRom Core

Introduction Minimal Dataset Rule and Ownership Paradigm

Data are stored to be (re)used; otherwise, there's no reason to store data. This basic rule follows the 'minimal dataset' paradigm: only data with a clear pre-defined goal for (re)use are stored.

This Minimal-Dataset Rule is essential for personal data: privacy is best served by only collecting and storing relevant data. The RomRom Core algorithms are aimed explicitly at health data; privacy and ownership are essential to account for at every coding level. This document elaborates on the basic outlines concerning the Minimal-Dataset Rule and Ownership Paradigm. This document will be revised regularly: please confirm you're using the most recent version.

People own their data

It sounds superfluous, but it's essential to realize that personal data are owned by the person it relates to. Everyone else can only have permission to use personal data; consent must be given willingly by the owner. The European Union's (EU) General Data Protection Regulation (GDPR) is based on this paradigm.

This document contains the 'rule book' for the RomRom Core algorithms that produce personal data statistics; see the Rules of data-ownership RomRom Core at the end of this document. There are three anonymous levels defined in this document:
1. personal data (non-anonymous)
2. clustered internal data (semi-anonymous)
3. anonymous data that can be reported publicly

Consequences for the RomRom Core algorithms

The developers have to be aware of the anonymous level of the statistical data they produce based on their algorithms. When reporting, the anonymous level has to be part of the report. Users of the RomRom Core reports must be made aware of the anonymous level of the outcomes they get. By implementing access rules in the organization's reporting tools, unauthorized access by people with an insufficient confidence level is prevented.

Description of the three anonymous levels

The three levels seem simple, but differentiation of the rules makes the levels more complex. Reports can easily contain personal information, sometimes only discernable by group members or leadership. A person owns personal information that relates to her or him; this person/ owner can use this information at will. Sharing an overview, which contains information tracible to a person, is only possible with explicit permission from or by the person the information relates to.

Table 1: anonymous levels for reporting, as a specification of the EU GDPR rules

Anonymous level	Numbers	Personal level	Description
Personal data (non-anonymous)	1 - 14	Overviews with information that can be traced to individuals by group members or their leadership/ management.	These reports are considered confidential, access only by persons with a non-disclosure agreement or the tracible individuals themselves.
Clustered personal data (semi-anonymous)	15 - 29	Organized groups, i.e., 'KLM pilots' or 'Dutch ENT-surgeons'.	These reports contain sensitive information only for use by the group or within an organization.
Anonymous data	30 or more	No specific identifiers, other than the reported subject.	These reports can be shared in the public domain, standard authorship rules.

Rules data-ownership RomRom Core

Name rule	Description	Author
Minimal Dataset	Only data that serve a pre-defined goal are stored.	Jan Rombout
Ownership assignment	People own their data; they have to give consent for (re)use willingly, which has to be adequately documented. Anonymous data are owned by the person or organization that generated or collected the data following authorship legislation.	Jan Rombout
Ownership group data	A representative of a specific group handles the data that is related to that group of persons.	Jan Rombout
Anonymous levels	30 or more: Data that concern 30 or more persons and individuals or groups cannot be distinguished, are considered anonymous, and can be reported publicly. 15 - 30: Data that concern a group of 15 – 30 persons or specific recognizable groups are semi-anonymous and can be used only internally by the organization or by a representative of that group. 1 - 14: Data related to 1-15 persons or overviews in which individuals can be recognized are non-anonymous and can be handled only by people with special permission and a non-disclosure agreement. Individuals can share their data at will.	Jan Rombout

Digital Twins reporting

When looking for digital twins the following sequence guaranties anonymous reporting.

1. The episode of care EoC of an actual patient is chosen.
2. The main issue that must be in common (selection criterium) is selected.
3. All patients from a database with this EoC and selection criterium are selected
4. The patients are sorted according to their similarity with the actual patient comparing their whole history. The more Snomed CT codes in common, the higher the ranking.
5. Of the first 1000 or so the content of the common Snomed CT codes is compared and a 10% selection (100 other patients) is used for reporting.
6. Of the top 100 digital twins a report is generated about only the EoC, with at least 15 patients per category.

This way the data of all patients can be re-used to find digital twins

Date of revision, revised issues, author name, and allocated hours

Date	Issue	Name	Hours
23-08-2021	Digital Twins reporting added	Jan Rombout	4
27-03-2021	Start	Jan Rombout	4