

Subscribe to DeepL Pro to edit this document.  
Visit [www.DeepL.com/pro](https://www.deepl.com/pro?cta=edit-document) for more information.

Information, best practices and commitment for PRISIB users



Pau Pericàs

Balearic Islands Health Information Research Platform (PRISIB)

Balearic Islands Health Research Institute (Idisba)

Health Service of the Balearic Islands (IBSalut)

pau.pericas@ibsalut.es

28/02/2024

**Index**

[Introduction](#__RefHeading___Toc794_2961146142)

[The PRISIB](#__RefHeading___Toc1177_2961146142)

[Presentation](#__RefHeading___Toc1179_2961146142)

[Data sources](#__RefHeading___Toc1183_2961146142)

[Participation in projects and institutions](#__RefHeading___Toc1185_2961146142)

[Services](#__RefHeading___Toc1603_757473047)

[Secondary use of clinical data](#__RefHeading___Toc1199_2961146142)

[The data management plan](#__RefHeading___Toc810_2961146142)

[FAIR Principles](#__RefHeading___Toc806_2961146142)

[The metadata](#__RefHeading___Toc808_2961146142)

[Issues in the use of clinical data](#__RefHeading___Toc3040_2961146142)

[Data protection](#__RefHeading___Toc796_2961146142)

[Advantages of using clinical data](#__RefHeading___Toc3042_2961146142)

[Data quality 10](#__RefHeading___Toc3044_2961146142)

[Measures to improve the use of clinical data 11](#__RefHeading___Toc3046_2961146142)

[Types of studies 11](#__RefHeading___Toc1203_2961146142)

[Publication of data 12](#__RefHeading___Toc812_2961146142)

[Federated model 13](#__RefHeading___Toc1330_2961146142)

[Annexes 14](#__RefHeading___Toc1088_4231230472)

[Patient Autonomy Act 14](#__RefHeading___Toc1207_2961146142)

[GDPR 15](#__RefHeading___Toc798_2961146142)

[LOPDGDD 17](#__RefHeading___Toc1834_1563066921)

[RD Observational studies with medicines 20](#__RefHeading___Toc1209_2961146142)

[Science, Technology and Innovation Act 21](#__RefHeading___Toc1211_2961146142)

[Bibliography 23](#__RefHeading___Toc363_1859376467)

# Introduction

The use of health information has a number of profound ethical, legal and technical implications that this document aims to summarise. The information provided here is necessarily incomplete and may be rendered obsolete by changes in laws, regulations and customs pertaining to the use of clinical data for research, but the PRISIB Scientific Technical Committee will ensure that this document remains relevant and up to date.

# The PRISIB

## Presentation

PRISIB is the Health Information Research Platform of the Balearic Islands Health Research Institute (IdISBa).

This platform was created as a collaboration between the Servei de Salut de les Illes Balears (IBSalut) and the Research Institute with the aim of providing a single access route to obtain valid and reliable data from the electronic health record and other complementary sources to create knowledge, support research and facilitate innovation and evaluation in the healthcare field.

To guarantee the correct use of data, all research projects are assessed by the PRISIB Scientific-Technical Committee. This committee advises the platform and evaluates the applications received, ideally prior to or simultaneously with the evaluation by the Ethics Committee, so that the contributions that can be made in terms of data management are approved together with the project protocol.

## Data sources

Generally speaking, any source of data collected during the activity of the health service is eligible for use by PRISIB for research projects provided that this use has been approved by the relevant Research and Ethics Committees and Commissions.

IBSalut's healthcare activity makes use of multiple information repositories that are not always interconnected and whose governance can be complex to navigate. For this reason it is impossible to make a complete catalogue of all the information potentially accessible to PRISIB.

The following is a list of the data sources most commonly used by the platform:

* SIAP: Primary Care Information System. It registers all the activity of the health centres and their complementary units.
* Electronic Prescription: Records all prescriptions made using the health card and drug collections made at pharmacies.
* Laboratory: All laboratories in the health system are computerised and their data can be jointly exploited.
* CMBD: Minimum Basic Data Set. Repository with the coding of all hospital discharge reports.

Other sources of information have been accessed in the past on an ad hoc basis for research use:

* Corporate PACS: IBSalud's image archiving and communication system where radiological examinations performed in hospitals are stored and consulted in clinical practice.
* RADELEC: System for the collection and analysis of electrocardiograms performed in both primary care and hospital centres.
* IdISBa Biobank: This Biobank has several collections of samples collected for research purposes that PRISIB can complement with clinical information from donors.

## Participation in projects and institutions

Given the growing interest in the use of health data in the scientific community, the platform participates in the following projects aimed at exploring data for research that can serve to improve the functioning of the platform by assimilating it with that of other equivalent institutions and participating in multi-centre research projects whether they originate in other institutions or are driven by the Balearic Islands research community.

* [IMPaCT-Data](https://impact-data.bsc.es/): is the Data Science Programme of the Infrastructure for Precision Medicine associated with Science and Technology promoted by the Carlos III Institute, which aims to support the development of a common, interoperable and integrated system for the collection and analysis of clinical data.
* [EHDEN](https://ehden.eu/): is the European Health Data & Evidence Network, promoted by Horizon 2020 and EFPIA to create a federated network of health data research centres using the same data models and coding (OMOP CDM) with the intention of generating robust and reliable scientific evidence quickly and efficiently.
* [OHDSI](https://www.ohdsi.org/): the Observational Health Data Sciences and Informatics programme is an international collaboration for the development of large-scale health information analysis tools using open source and the federated data model.
* [EOSC-A](https://eosc.eu/): The European Open Science Cloud Association is the legal entity created by the European Commission to govern the European data portal and services for scientific research.

## Services

The services that can be requested from PRISIB are as follows:

* Extraction of data from the electronic medical record for research projects. The project requesting data from PRISIB must respond to a protocol duly drawn up by the research team developing it. This protocol must have been approved by the corresponding research committee and will be evaluated by the PRISIB Scientific-Technical Committee and the CEIm.

The evaluation by the PRISIB CST will judge aspects related to data management, such as the adequacy of the requested data to the stated objectives of the study or the means used to record, store and analyse the data. It is therefore recommended that a data management plan be drafted as an annex to the protocol (see section 'The data management plan').

This service includes the creation of the project's data model, the identification of data sources, the mapping of variables, the creation of extraction algorithms, the scheduling of periodic extractions, anonymised georeferencing of data, data quality analysis, descriptive statistics and anonymisation.

* Consultations for feasibility studies or exploratory studies. Free consultation service for feasibility studies that provides approximate, aggregated data, without quality control and whose objective is to have a first approximation to assess whether a more ambitious study can be developed. These are studies in which you want to know the N that exists in the databases of a certain problem or how certain information of interest is registered.
* Visualisation of data linked to research projects. These visualisations can range from the generation of vector maps or statistical graphs to the implementation of interactive reports in web format.
* Implementation of systems for research data collection, case randomisation and pseudonymisation.
* Custody of research data, which would be stored in the Platform's data repository at the Health Service's data processing centre.
* Advice on drafting applications, protocols, research projects, interoperability plans, data management plans and other documentation related to the use of data in health research.

# Secondary use of clinical data

## The data management plan

For most competitive calls, the submission of a Data Management Plan (DMP) together with the project protocol is being requested. The DMP is a document specifying all aspects of data processing and data protection during and after the implementation of a research project.

The PGD is a document that facilitates clarity on the legal aspects of data processing, collaboration with other researchers, evaluation of proposals and resolution of doubts concerning the use of data.

In summary, a PGD should specify:

* Where the data will be obtained from, what format it will be in, whether it will be static or change over time, how much data is expected to be collected.
* How these data will be documented and described, what criteria will be established to validate their quality and how they will be materialised.
* Where this data will be stored and how it will be protected against loss or theft.
* What confidentiality and privacy issues are foreseen and how will they be addressed?
* How other researchers might find, use and access these data
* How, why and for how long this data will be archived
* Who will be responsible for each of the aspects defined in the PGD?

There are different models of PGDs and several tools and guides to help in their drafting, among which we highlight the following:

* [ARGOS (Openaire)](https://argos.openaire.eu/splash/)
* [DMPonline (Data Curation Center)](https://dmponline.dcc.ac.uk/)
* [Cora (Consoci de Serveis Universitaris de Catalunya)](https://dmp.csuc.cat/)
* [10 Pasos for elaborate a plan at management plan from data management (CRUE)](https://www.rebiun.org/acceso-abierto/10-pasos-para-elaborar-un-plan-de-gestion-de-datos)

The drafting and publication of the PGD is seen as one of the cornerstones of the implementation of the FAIR principles.

## FAIR Principles

These principles are summarised in Mark D. Wilkinson et al., "The FAIR Guiding Principles for Scientific Data Management and Stewardship," Scientific Data (March 15, 2016) as follows

* To be Findable:

F1. Data and metadata are assigned a globally unique and persistent identifier (pe DOI).

F2. data is described with rich metadata (defined by R1 below).

F3. Metadata clearly and explicitly includes the identifier of the data it describes.

F4. The data is recorded or indexed in a searchable resource.

* To be Accessible:

A1. Data and metadata can be retrieved via their identifier using a standardised communication protocol.

A1.1 the protocol is open, free and universally implemented.

A1.2 the protocol allows for an authentication and authorisation procedure, where necessary.

A2. Metadata is accessible, even when the data is no longer available.

* To be Interoperable:

I1. Data and metadata use a formal, accessible, shared and widely applicable language for knowledge representation.

I2. The data and metadata use vocabularies that follow the FAIR principles.

I3. The data and metadata include qualified references to other data and metadata.

* To be Reusable:

R1. The metadata describes the data in detail with accurate and relevant attributes.

R1.1. Data and metadata are published with a clear and accessible data use licence.

R1.2. Data and metadata are associated with a detailed provenance.

R1.3. The data and metadata comply with the relevant EU standards for the domain.

## The metadata

To make the application of the FAIR principles practicable, a detailed description of the data generated in a study is needed. This description is what we call metadata, i.e. data about the data.

In health research, data often cannot be made public. Thus the publication of metadata is of even greater relevance, as it will explain what the dataset consists of and the persons and procedures to contact in order to request access to the data.

There are different models for documenting metadata specific to different fields of research, but there is no universal solution for all possible cases.

To facilitate communication between researchers and technicians, as well as the subsequent publication of at least the metadata, PRISIB uses its own data model to describe the datasets it generates. This model can be found in the institutional repository [Docusalut](http://hdl.handle.net/20.500.13003/18160) or in the open repository [Zenodo](http://10.5281/zenodo.7185046).

## Problems with the use of clinical data

The data that can be found in patients' medical records are collected for healthcare purposes, not scientific ones, and this means that their use for research must be done with the knowledge of the limitations that this implies and the necessary precautions.

* Observational data: As these are purely observational data, it is not possible to attribute causality to the hypotheses to be tested.
* Incomplete data: It is difficult, if not impossible, to know why a piece of information is missing from the medical record. It may not have been measured, or it may have been measured and not recorded, and there is usually no way of knowing what bias is introduced by this omission.
* Unreliable data: As the aim of the registry is not the actual collection of data, as in a scientific experiment, but patient care, errors may be made in data entry.
* Abundant data: Having a large amount of data can lead to the emergence of spurious associations to which one should avoid attributing relevant correlation.
* Biases: As the data is collected for healthcare purposes, it is biased from the outset, as information is only collected from those users who use the public health system, excluding those who have private healthcare or those who, for whatever reason, do not make use of healthcare resources (e.g. situations of social exclusion).

## Data protection

Secondary use of health data always involves balancing three ethical principles of utmost importance:

* Confidentiality

Healthcare personnel are obliged not to disclose confidential patient information without the patient's knowledge and authorisation.

* Autonomy

All people have unconditional intrinsic worth and should be empowered to make their own rational and moral decisions, and all should be able to exercise their capacity autonomously.

* Charity

It is the obligation of medical personnel to act in the best interests of their patients and is the basis of multiple ethical standards involving the protection and defence of rights, non-maleficence, assistance to persons in distress, elimination of health risks and assistance to persons with disabilities.

Making use of patient information without prior consent to the collection of information would violate the principle of confidentiality and autonomy. Not doing so would be in breach of the principle of beneficence by not making use of available resources for the improvement of patients' health. This ethical dilemma is present in most observational studies and its balance is a subject of constant discussion and movement in both the bioethical and legal spheres.

As a general rule, explicit written consent should be obtained from all patients to use their data. But according to RD 957/2020 of 3 November, which regulates observational studies with medicinal products for human use: "informed consent may be waived, provided that the IRB/IEC considers that the observational research is of important social value, that its conduct would not be feasible or viable without such a waiver, and that it entails minimal risks to the participants".

## Advantages of using clinical data

On the other hand, the use of medical record data has some advantages over the use of prospectively collected data in the context of a clinical trial.

* Multiple interventions are measured
* It is assessed on real patients representative of routine practice.
* It is carried out in the real environment where the measures studied are applied.
* Further data evolution is available
* The number of observed variables can be expanded
* Allows tracking of changes in results
* It is cheaper and simpler

In addition, they allow the analysis of relevant questions that would be very difficult or impossible to measure in a randomised clinical trial:

* Impact of the implementation of new technologies
* Study of rare diseases
* Detection of relevant variables in decision making
* Monitoring access to and use of health resources
* Identifying unmet health needs
* Detecting inappropriate use of health resources
* Impact of an intervention on care circuits

## Data quality

The quality of data depends on its purpose. A data may be of high quality for its primary purpose but if it is intended for a secondary use it may be flawed in one of the different dimensions in which data quality can be described.

"Data are of high quality if they are suitable for their intended uses in operations, decision-making and planning. Data are fit for use if they are free of defects and have the desired characteristics".

Data Quality: The Field Guide, Thomas C. Redman, Ph.D. Digital Press,2001

In the publication *Sidi,Fatimah, et al. "Data quality: A survey of data quality dimensions." 2012 describes the different* dimensions of data quality:

Accuracy: Does the data correctly represent the actual entity or event?

Consistency: Does the data contain no contradictions?

Availability: Can data be accessed now and over time?

Completeness: Does the data include all elements representing the entity or event?

Completeness: Does the data include all the elements it is intended to represent?

Compliance: Does the data follow accepted standards?

Processability: Is the data machine-readable?

Credibility: Are they based on reliable sources?

Relevance: Do they include an adequate amount of data?

Timeliness: Do they represent the actual situation and are they published sufficiently in advance?

## Measures to improve the use of clinical data

To avoid magnifying spurious associations and to truly measure what is intended in the scientific question, it is necessary to have a clear causal model. This implies that before deciding what information to extract from clinical records, it is necessary to be clear about all confounding factors and intermediate variables that may obscure and blur the relationships between the variables we want to investigate.

To represent these relationships, the use of different diagrams is recommended, such as [directed acyclic graphs](http://10.1016/j.chest.2020.03.011), which allow to clearly state the associations expected to be found in the data and those to be measured in order to answer the study question.

Another basic measure to alleviate the problems of secondary use of data is to establish criteria for validation and verification of data prior to data extraction in order to improve consistency and credibility without introducing new biases that could be incurred by deciding on the validity of data when they are already known.

Different statistical techniques can help to understand and mitigate biases, confounding factors and heterogeneities in the data that can introduce variances and deviations in our results.

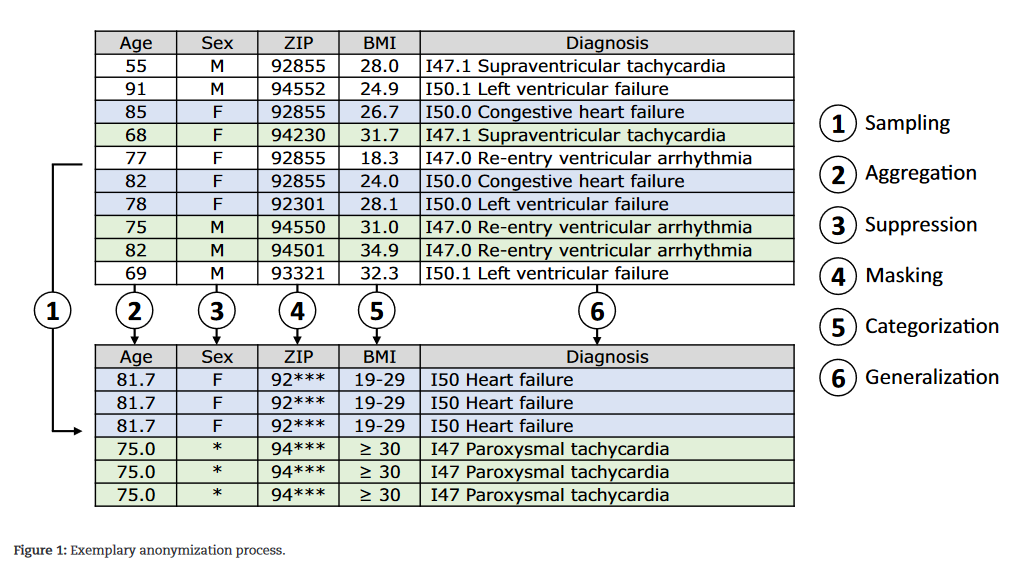
## Types of studies

Despite the above limitations, the secondary use of clinical data allows for different study models depending on whether an intervention is recorded, whether data are compared before and after the intervention, or whether results are compared with a control group:

* Purely observational: Collection of data at a specific point in time.
* Post-exposure: Collection of data from selected patients following an event.
* Pre-Post Exposure: Collection of data from selected patients before and after an event.
* Pragmatic clinical trial: Collection of data from selected patients before and after an event and from a comparable group without exposure to the event with exposure to the event being randomised.

## Publication of data

The publication of personal data is strongly regulated by different EU and national laws. In order to make this data public, it is necessary to guarantee that the data set does not allow the identification of any person and for this purpose different anonymisation techniques are used, which are exemplified in the following figure:

Figure 1A scalable software solution for anonymizing high-dimensional biomedical data, GigaScience, Volume 10, Issue 10, October 2021

Properly anonymised data are not considered personal data and can be published openly in accordance with European directives,

To facilitate the application of these techniques on large datasets, there are publicly available tools such as [ARX - Data Anonymization Tool](https://arx.deidentifier.org/) or [Amnesia (Openaire)](https://amnesia.openaire.eu/).

## Federated model

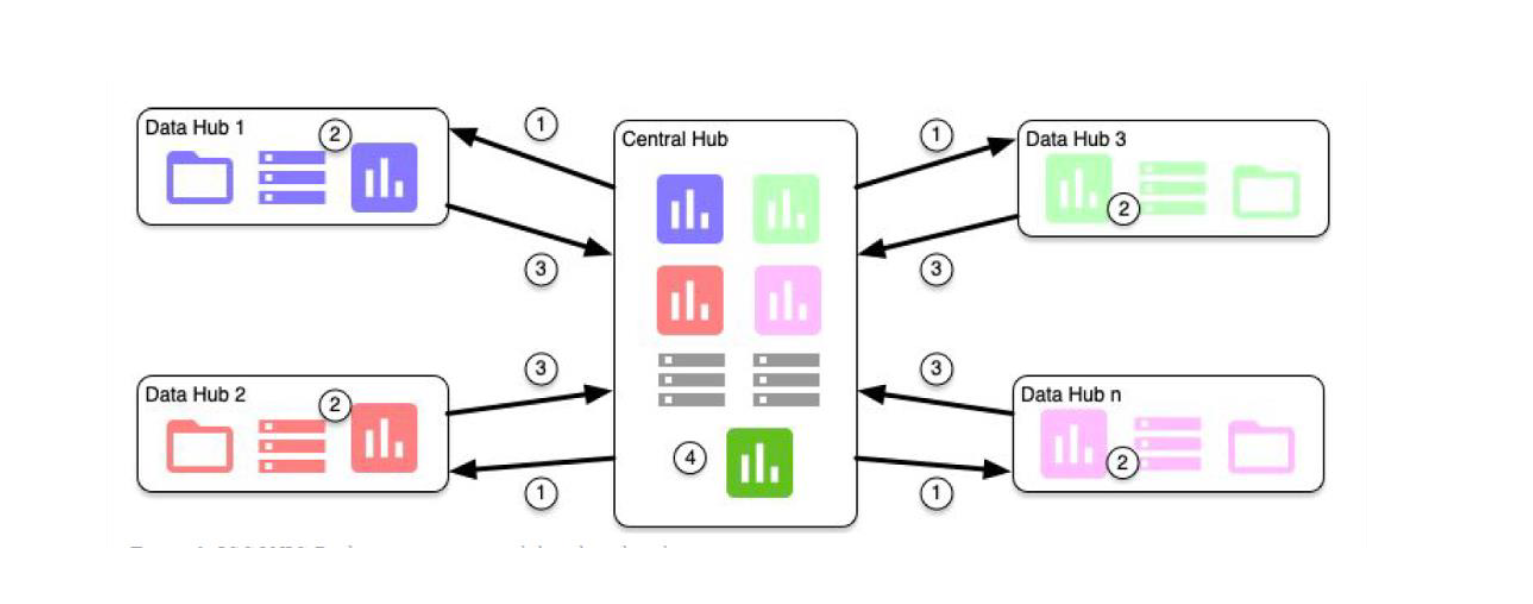
One way to generate stronger evidence and eliminate some biases is to combine data from different sources to conduct a multicentre study.

But the transfer of personal data for such studies, especially if they involve data of the entire population or of vulnerable groups, raises many ethical and legal concerns and problems.

To facilitate the conduct of these large studies without compromising data security and privacy, federated studies are possible. In these, a coordinating centre shares a data model and a clear definition of the data to be analysed and the algorithm with which the data are analysed with multiple participating nodes. Each participating node subsequently extracts and harmonises its data according to the data model and analyses it using the same algorithm using its own resources and returns to the coordinating centre only the aggregated result of this analysis. Finally, the coordinating centre analyses the results of the different nodes to produce the final results.

This model offers multiple advantages, mainly the fact that no personal data is transferred in any case, as it is analysed by each node within its own data space. In addition, it allows the computational and storage workload of the different datasets generated to be distributed.

If these studies are conducted between sites sharing a compatible infrastructure and a harmonised data model, conducting large multi-centre studies is reduced to running a single analysis package at each node.

Initiatives such as EHDEN, [Darwin-EU](https://www.darwin-eu.org/) (led by the European Medicines Agency) or IMPaCT, seek to implement this cooperation model that guarantees local data governance but allows for the creation of large population-based studies on an international scale.

# Annexes

## Patient Autonomy Act

[Law 41/2002 of 14 November 2002, the basic law regulating patient autonomy and the rights and obligations regarding clinical information and documentation](https://www.boe.es/eli/es/l/2002/11/14/41/con) establishes the following in point 3 of article 16:

"Access to clinical records for judicial, epidemiological, public health, research or teaching purposes is governed by the provisions of current legislation on the protection of personal data, and by Law 14/1986, of 25 April, General Health Act, and other applicable regulations in each case. Access to the clinical history for these purposes requires the preservation of the patient's personal identification data, separated from those of a clinical-healthcare nature, in such a way that, as a general rule, anonymity is ensured, unless the patient himself has given his consent not to separate them.

The cases of investigation provided for in section 2 of the seventeenth additional provision of the Organic Law on the Protection of Personal Data and the Guarantee of Digital Rights are excepted.

Likewise, exceptions are made in cases of investigation by the judicial authority in which the unification of identification data with clinical and healthcare data is considered essential, in which case the provisions of the judges and courts in the corresponding process shall apply. Access to data and documents from the clinical history is strictly limited to the specific purposes of each case.

When this is necessary for the prevention of a serious risk or danger to the health of the population, the health administrations referred to in Law 33/2011, of 4 October, General Public Health Act, may access patients' identification data for epidemiological reasons or for the protection of public health. Access shall in all cases be granted by a health professional subject to professional secrecy or by another person also subject to an equivalent obligation of secrecy, subject to prior justification by the Administration requesting access to the data".

## GDPR

The [General Data Protection Regulation or European Data Protection Regulation (GDPR/REPD)](https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32016R0679) of 2016 establishes the fundamental right to data protection of natural persons.

Article 9

Processing of special categories of personal data

1. The processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data intended to uniquely identify a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited.

2. Paragraph 1 shall not apply where one of the following circumstances applies: [...]

(i) processing is necessary for reasons of public interest in the field of public health, such as protection against serious cross-border threats to health, or to ensure high standards of quality and safety of healthcare and of medicinal products or medical devices, on the basis of Union or Member State law providing for appropriate and specific measures to protect the rights and freedoms of the data subject, in particular professional secrecy,

(j) the processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, in accordance with Article 89(1), on the basis of Union or Member State law, which must be proportionate to the aim pursued, respect in essence the right to data protection and provide for appropriate and specific measures to protect the interests and fundamental rights of the data subject.

It states that 'where the controller intends to further process personal data for a purpose other than that for which they were collected, the controller shall, prior to such further processing, provide the data subject with information about that other purpose and any relevant additional information [...]'. Although it exempts from the obligation to specifically inform about this further processing "to the extent that the data subject already has the information" about this processing.

This law further regulates the responsibilities of controllers and processors of personal data on the basis of "data protection by design and by default" and "taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of the processing, as well as the risks [...] the controller shall implement appropriate technical and organisational measures, such as pseudonymisation [... and ] data minimisation".

On the processor it says, inter alia, that it "shall process personal data only on the documented instructions of the controller", "shall ensure that persons authorised to process personal data have undertaken to respect confidentiality or are subject to a confidentiality obligation of a statutory nature" and "shall, at the choice of the controller, erase or return all personal data upon termination of the provision of processing services, and erase existing copies". "Where a processor uses another processor to carry out certain processing activities on behalf of the controller, the same data protection obligations shall be imposed on that other processor by contract or other legal act [...] as those stipulated in the contract or other legal act between the controller [...]. If that other processor fails to fulfil its data protection obligations, the original processor shall remain fully responsible'. Thus "The processor shall, without undue delay, notify the controller of any personal data security breaches of which it becomes aware".

Data protection impact assessment

Where a type of processing, in particular where it uses new technologies, is likely, by its nature, scope, context or purposes, to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the processing operations on the protection of personal data. A single assessment may address a number of similar processing operations involving similar high risks.

Article 40

Codes of conduct

1. The Member States, the supervisory authorities, the Committee and the Commission shall promote the development of codes of conduct aimed at contributing to the correct application of this Regulation, taking into account the specific characteristics of the different processing sectors and the specific needs of micro, small and medium-sized enterprises.

2. Associations and other bodies representing categories of controllers or processors may draw up codes of conduct or amend or extend such codes in order to specify the application of this Regulation, such as with regard to:

(a) fair and transparent treatment;

(b) the legitimate interests pursued by controllers in specific contexts;

(c) the collection of personal data;

(d) pseudonymisation of personal data;

(e) the information provided to the public and stakeholders;

(f) the exercise of the rights of data subjects;

(g) the information provided to and the protection of children, as well as the manner of obtaining the consent of the holders of parental responsibility or guardianship over the child;

(h) the measures and procedures referred to in Articles 24 and 25 and the measures to ensure the security of processing referred to in Article 32;

(i) the notification of personal data security breaches to supervisory authorities and the communication of such breaches to data subjects;

(j) the transfer of personal data to third countries or international organisations, or

(k) out-of-court and other dispute resolution procedures for resolving disputes between controllers and data subjects concerning processing, without prejudice to the rights of data subjects under Articles 77 and 79

## LOPDGDD

These rights are developed and regulated in Spain by [Organic Law 3/2018, of 5 December, on Personal Data Protection and guarantee of digital rights](https://www.boe.es/eli/es/lo/2018/12/05/3/con).

Of special interest is the seventeenth additional provision on the processing of health data, point 2 of which states:

Data processing in health research shall be governed by the following criteria:

a) The data subject or, where appropriate, his or her legal representative may give consent to the use of his or her data for health research purposes and, in particular, for biomedical research. Such purposes may cover categories related to general areas linked to a medical or research speciality.

b) Health authorities and public institutions with public health surveillance competences may carry out scientific studies without the consent of those concerned in situations of exceptional public health relevance and severity.

(c) the re-use of personal data for health and biomedical research purposes shall be considered lawful and compatible when, having obtained consent for a specific purpose, the data are used for purposes or areas of research related to the area in which the initial study was scientifically integrated.

In such cases, the data controllers shall publish the information required by Article 13 of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of individuals with regard to the processing of their personal data and on the free movement of such data, in an easily accessible place on the corporate website of the centre where the research or clinical study is conducted, and, where appropriate, on that of the sponsor, and shall notify the data subjects of the existence of this information by electronic means. When they do not have the means to access such information, they may request that it be sent in another format.

For the treatments provided for in this point, a prior favourable report from the research ethics committee shall be required.

(d) The use of pseudonymised personal data for health and, in particular, biomedical research purposes is considered lawful. The use of pseudonymised personal data for public health and biomedical research purposes shall require:

1. A technical and functional separation between the research team and those who carry out the pseudonymisation and keep the information that makes re-identification possible.

2. Pseudonymised data should only be accessible to the research team when:

(i) There is an express commitment to confidentiality and not to engage in any re-identification activity.

(ii) specific security measures are taken to prevent re-identification and access by unauthorised third parties.

The re-identification of data at source may take place when, in the course of an investigation using pseudonymised data, it is established that there is a real and concrete danger to the safety or health of a person or group of persons, or a serious threat to their rights, or is necessary to ensure adequate health care.

(e) Where personal data are processed for the purposes of health research, and in particular biomedical research, within the meaning of Article 89(2) of Regulation (EU) 2016/679, the rights of data subjects provided for in Articles 15, 16, 18 and 21 of Regulation (EU) 2016/679 may be waived where:

1. The aforementioned rights are exercised directly with researchers or research centres that use anonymised or pseudonymised data.

2. The exercise of these rights refers to the results of the research.

3. The purpose of the research is of essential public interest related to state security, defence, public safety or other important objectives of general public interest, provided that in the latter case the exception is expressly set out in a regulation with the status of a law.

(f) Where, in accordance with Article 89 of Regulation (EU) 2016/679, processing is carried out for the purposes of public health research and, in particular, biomedical research, the following shall be carried out:

1. Carry out an impact assessment identifying the risks arising from the processing in the cases provided for in Article 35 of Regulation (EU) 2016/679 or those established by the supervisory authority. This assessment shall specifically include the re-identification risks linked to the anonymisation or pseudonymisation of data.

2. Subject scientific research to quality standards and, where appropriate, to international guidelines on good clinical practice.

3. Adopt, where appropriate, measures aimed at ensuring that researchers do not access data that identifies the data subjects.

4. Designate a legal representative established in the European Union, in accordance with Article 74 of Regulation (EU) 536/2014, if the sponsor of a clinical trial is not established in the European Union. That legal representative may be the same as that provided for in Article 27(1) of Regulation (EU) 2016/679.

g) The use of pseudonymised personal data for public health and, in particular, biomedical research purposes must be subject to the prior report of the research ethics committee provided for in the sectoral regulations. In the absence of the existence of the aforementioned Committee, the entity responsible for the research shall require a prior report from the data protection officer or, failing this, from an expert with the prior knowledge set out in Article 37.5 of Regulation (EU) 2016/679.

## RD Observational studies with medicines

[Royal Decree 957/2020 of 3 November, which regulates observational studies with medicinal products for human use](https://www.boe.es/eli/es/rd/2020/11/03/957):

Article 4. Prerequisites for the initiation of observational studies with medicinal products.

In accordance with the provisions of Article 12 of Royal Decree 1090/2015 of 4 December 2015, as well as Chapter III of this Royal Decree, prior to their initiation, all observational studies with medicinal products shall require the favourable opinion of an IRB accredited in Spain. This opinion shall be unique, binding and recognised throughout the national territory. The procedure for obtaining the IRB's opinion shall follow the procedure regulated in Chapter III.

2. In the case of observational studies with prospective follow-up medicinal products, following a favourable opinion of the IRB/IEC, the competent health authorities may establish additional requirements for the initiation of such studies in the centres under their competence. The establishment of such requirements shall be justified on the basis of feasibility or relevance criteria, but not on the basis of aspects of the study already assessed by the relevant IRB/IEC. No additional requirements may be established for studies where the sponsor is a public authority or where it is demonstrated that the study is a non-commercial clinical investigation.

3. Studies involving the collection of information directly from the subject participant or from the health professional attending him/her in a health centre, service or establishment shall require the prior agreement of the person responsible for the same to the protocol and to the rest of the documentation that has obtained the favourable opinion of the IRB/IEC. The agreement shall be expressed by signing a contract with the sponsor. This contract will not be necessary in those cases in which the sponsor belongs to the centre, service or health establishment where the study is carried out, it being sufficient to obtain the express agreement of the person in charge of the same.

Article 5. Informed consent and protection of the personal data of participating subjects.

1. Observational studies with medicinal products that involve interviewing the subject participant shall require his/her informed consent. However, in accordance with the applicable provisions of current regulations and ethical principles for medical research involving human subjects, informed consent may be waived if the IRB/IEC considers that the observational research is of significant societal value, that its conduct would not be feasible or practicable without such a waiver, and that it entails minimal risk to the participants.

2. The request for informed consent shall take into account, where appropriate, the ethical rules and principles relating to the provisions for the collection, storage and possible future use of the biological samples of the subjects.

3. Promoters of studies using any source of information that includes the processing of personal data shall take into account the provisions of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC, and of Organic Law 3/2018 of 5 December on the Protection of Personal Data and the guarantee of digital rights, and in particular the following:

a) The developer shall have assessed and mitigated, through appropriate measures in each case, the impact that the conduct of the study may have on the protection of personal data.

b) The study sponsor and investigators should ensure the confidentiality of the data of the participating subjects.

c) Without prejudice to the provisions of paragraph 1, the consent of the subject participant shall be required unless another legitimate basis for the processing of his/her personal data applies from among those referred to in Articles 6.1 and 9.2 of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016. In addition, the sponsor and the researchers must apply the criteria governing the processing of data in health research in accordance with the seventeenth additional provision of Organic Law 3/2018, of 5 December.

(d) the conditions of access to personal data shall be detailed in the protocol, including the conditions of their international transmission outside the European Economic Area, if foreseen.

4. In the case of a study with anonymised data or data that have undergone pseudonymisation treatment, the protocol shall include the procedure followed to achieve such anonymisation or pseudonymisation.

5. Access to the data of the participating subjects shall in any case be subject to the conditions established by the data controller, so as to ensure compliance with the regulations on the protection of personal data.

## Science, Technology and Innovation Law

[Law 17/2022, of 5 September, which amends Law 14/2011, of 1 June, on Science, Technology and Innovation](https://www.boe.es/eli/es/l/2022/09/05/17) defines among its general objectives "To promote open science at the service of society and to promote initiatives aimed at facilitating free access to data, documents and results generated by research, developing open infrastructures and platforms, and encouraging open participation of civil society in scientific processes."

This objective is further elaborated in Article 37, which states:

"Research personnel in the public sector or whose research activity is mainly financed with public funds and who choose to disseminate their research results in scientific publications shall deposit a copy of the final version accepted for publication and the associated data in institutional or thematic open access repositories, simultaneously with the date of publication".

Both the [Universitat de les Illes Balears](https://dspace.uib.es/xmlui/) and the [IBSalut](https://docusalut.com/) have their own institutional repositories.

"5. [...In addition to open access, and always with the aim of making science more open, accessible, efficient, transparent and beneficial to society, the Ministries of Science and Innovation and of Universities, each in their respective fields of action, as well as the Autonomous Communities within the framework of their competences, will also promote other initiatives aimed at facilitating free access to and management of data generated by research (open data), in accordance with the international FAIR principles (easy to find, accessible, interoperable and reusable), to develop open infrastructures and platforms, to promote the publication of scientific results in open access, and the open participation of civil society in scientific processes, as developed in article 38."

# Bibliography

1. <https://impact-data.bsc.es/>
2. <https://eosc.eu/>
3. <https://ehden.eu/>
4. <https://www.ohdsi.org/>
5. Suttorp, M. M et al. Graphical presentation of confounding in directed acyclic graphs. 2015 Nephrology Dialysis Transplantation
6. Etminan M. Using Causal Diagrams to Improve the Design and Interpretation of Medical Research. Chest. 2020 DOI: [10.1016/j.chest.2020.03.011](https://doi.org/10.1016/j.chest.2020.03.011)
7. <https://www.darwin-eu.org/>
8. <https://www.phiri.eu/>
9. <https://www.boe.es/eli/es/l/2002/11/14/41/con>
10. <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32016R0679>
11. <https://www.boe.es/eli/es/lo/2018/12/05/3/con>
12. <https://www.boe.es/eli/es/rd/2020/11/03/957>
13. <https://www.boe.es/eli/es/l/2022/09/05/17>
14. <https://www.go-fair.org/fair-principles/>
15. 10.1038/sdata.2016.18
16. <https://dspace.uib.es/xmlui/>
17. https://docusalut.com/
18. https://datamanagement.hms.harvard.edu/plan-design/data-management-plans
19. El Emam, K., & Dankar, F. K. (2008). Protecting privacy using k-anonymity. *Journal of the American Medical Informatics Association : JAMIA*, *15*(5), 627-637. https://doi.org/10.1197/jamia.M2716
20. Thierry Meurers, Raffael Bild, Kieu-Mi Do, Fabian Prasser, A scalable software solution for anonymizing high-dimensional biomedical data, GigaScience, Volume 10, Issue 10, October 2021, [giab068](https://doi.org/10.1093/gigascience/giab068), https://doi.org/10.1093/gigascience/giab068
21. https://www.rebiun.org/acceso-abierto/10-pasos-para-elaborar-un-plan-de-gestion-de-datos
22. https://www.rebiun.org/acceso-abierto/los-datos-de-investigacion-importan-gestiona-comparte-publica-y-reutiliza