Code of Conduct
for the use and provision of the
RD-Connect Genome-Phenome Analysis Platform
(RDC-GPAP)[1]

Revised and approved by the RD-Connect Data Access Committee as of 28 July 2023, to be applied as of 8 August 2023.

1. Context and purpose

The present **Code of Conduct for the use and provision of the RD-Connect for Genome-Phenome Analysis Platform**[2] (hereinafter RDC-GPAP) is set out in two parts: the **Code of Conduct for RDC-GPAP Users** the terms under which researchers and clinicians may use the RDC-GPAP while the **Code of Conduct for RDC-GPAP Service Provision** set out the principles to which the RDC-GPAP providers will adhere. Key terms used in the Code of Conduct are defined in Annex 1.

1.1. RDC-GPAP is operated and supported by Consorcio para la Explotación del Centro Nacional de Análisis Genómico (Consorcio CNAG; http://www.cnag.eu/), a research centre for genomics funded by the Spanish and Catalan Governments. Consorcio CNAG collaborates with the European Bioinformatics Institute (EBI) (https://www.ebi.ac.uk) for the purposes of providing services to store raw data in the European Genome-Phenome Archive (EGA) (https://ega-archive.org).

1.2. RDC-GPAP services are offered on a contractual basis to clinicians and researchers in both public and private research entities. All clinicians and researchers using the RDC-GPAP interface and services are hereinafter referred to as RDC-GPAP users. In addition, a clinician or researcher registering a research group in RDC-GPAP will be referred to as RDC-GPAP Group Leader.

1.3. Consorcio CNAG is advised in the operation of RDC-GPAP by a Data Access Committee (DAC). The role of the DAC is to ensure that access to the RDC-GPAP and use of the data it stores is always based on the scientific validity, quality and potential success of the request for use (specific details of DAC responsibilities are provided in the RDC-GPAP Data Access Committee Terms of Reference).

1.4. RDC-GPAP institutional users include, but are not limited to, the accredited centres of the European Reference Network for Rare Diseases.

1.5. RDC-GPAP individual users are accredited clinicians and researchers from accredited bodies who submit and access pseudonymised data of patients and relatives. RDC-GPAP users can include clinicians or researchers from commercial or pre-commercial organizations working in a project approved by DAC's RDC-GPAP.

1.6. Data submitted to RDC-GPAP will be available to the other verified and authenticated users of the platform after submission. An RDC-GPAP user may however choose to limit such access for a maximum of six months for exclusive use of the data for the submitter. Extension of this embargo period can be requested to help@rd.connect.eu at the time of submission and will be evaluated by the DAC. However, this extension is discouraged as the RDC-GPAP promotes IRDiRC's goal of "receiving an accurate diagnosis, care, and available therapy within one year of coming to medical attention".

1.7. RDC-GPAP is operated in close connection with pan-European biomedical research infrastructures like ELIXIR (https://elixir-europe.org/about-us). Some of the authentication and accreditation tools for user registration

1 This Code of Conduct was originally known as the RD-Connect Code of Practice for integrated user access to the RD-Connect platform for health-related information and human biological samples. The current version was approved by the Data Access Committee (DAC) on 28. July 2023.
It should always be referred to according to the effective date.

2 The RDC-GPAP is an online, controlled-access suite of software tools and underlying secure database that enables the standardised collection, integration, storage, real-time analysis and reuse of linked genomic and phenotypic data and metadata of individuals with rare diseases. RDC-GPAP provides a platform for processing and storing personal data of rare disease patients and their family members. The RDC-GPAP enables clinicians and researchers to analyse and interpret the full genomic datasets for research on both diagnosis and gene discovery of an individual patient basis. The tool has been developed to help clinicians and researchers find confirmatory cases from among the datasets submitted by other researchers. The RDC-GPAP is a mechanism for data sharing, allowing clinicians and researchers to access information about similar patients submitted by other users and enabling other users to query the datasets that have been submitted. It allows for matchmaking, finding second families, finding patient cohorts for validation studies, and allowing basic science researchers to find human analogues for their animal models. RDC-GPAP comes from the project "RD-CONNECT: An integrated platform connecting registries, biobanks and clinical bioinformatics for RD research", funded by the EC Seventh Framework Programme (FP7) from Nov 2012 to Oct 2018 under grant n° 305444.

created by EU Infrastructures might be used for RDC-GPAP user authentication.

1.8. The present Code of Conduct is to be included as an annex to the Adherence Agreement which forms part of the Contract between researchers or clinicians and Consorcio CNAG to use the RDC-GPAP interface and services.

1.9. Signature of the Adherence Agreement binds the individual RDC-GPAP Group Leader to the terms of this Code of Conduct and also creates a binding duty to ensure that any RDC-GPAP user working under their control have read and understood this Code of Conduct.

1.10. It should be noted that this Code of Conduct addresses only the use of data. It does not address the conduct of trials or research as a whole. Accordingly consent for inclusion in a study or trial should not be conflated with consent for data collection or use.

2. Core Principles

The following five principles for the stewardship of bio-specimens and data repositories constitute the foundational principles for the use and provision of the RDC-GPAP interface.

2.1. **Respect for privacy and autonomy:** stewardship implies protection of participants' privacy. Privacy protection measures should be in place and data subjects must be informed in general terms the potential use of data and bio-specimens for future research purposes and their rights to consent to such further use.

2.2. **Reciprocity:** stewardship also implies giving back. Feedback of general, non-patient specific, results should be channelled to institutions and patients.

2.3. **Freedom of scientific enquiry**: stewardship should encourage openness of scientific enquiry and maximize data and bio-specimen use and sharing so as to exploit their full potential to promote health.

2.4. **Attribution:** the intellectual investment of investigators involved in the creation of data registries and bio-repositories is often substantial and should be acknowledged by mutual agreement.

2.5. **Respect for intellectual property:** the sharing of data and bio-specimens needs to protect proprietary information and address the requirements of institutions and third-party funders.

3. Applicable law and legal provisions

3.1. Each experimental genotype dataset and phenotypic profile submitted to the RDC-GPAP interface is assigned an identifier to maintain links between the two and to enable joint genome-phenome analysis. These identifiers are alphanumeric codes and are not based on or related to any personal information. **The data processed in RDC-GPAP are accordingly classified as pseudonymised and are therefore governed by** 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 (hereinafter referred to as "GDPR"), and repealing Directive 95/46/EC.

3.2. Both Consorcio CNAG, as the provider of RDC-GPAP, and the RDC-GPAP users have a duty to comply with GDPR. Both are acting as **data controllers**, since the researcher obtains the genomic data and enhances the use of the RDC-GPAP according to its research needs; while Consorcio CNAG maintains the data in the RDC-GPAP for use by other researchers and may also conduct its own research with the RDC-GPAP.

3.3. Consorcio CNAG, being established in Spain, must operate RDC-GPAP in accordance with **Organic Law 3/2018 on the Protection of Personal Data and the Guarantee of Digital Rights**, being the adaptation of the Spanish legal framework to the GDPR and the development of its provisions.

3.4. Consorcio CNAG must ensure that all data processed and held in the RDC-GPAP system are processed in accordance with one of the legal bases provided for in Article 6 of GDPR: consent, contract, legal obligation, vital interests, public interests, legitimate interest; and one of the additional bases required to process sensitive data provided in Article 9 of GDPR, which could be the explicit consent of the data subject, public interest in public health, or necessary for research purposes.

3.5. Consorcio CNAG obtains data in the context of its contract with researchers on the basis of the public interest. Consorcio CNAG is accordingly using **public interest** (GDPR Article 6(1)(e)) as its primary legal basis for obtaining the data. As the data to be processed are sensitive data, Consorcio CNAG will use processing for **research purposes** (GDPR Article 9(2)(j)) as the legal basis for processing sensitive data.

3.6. The use of 'research purposes' as the legal basis for processing sensitive data means that RDC-GPAP

processing must also be compliant with the guidance and requirements of GDPR Article 89 and Organic law 3/2018 Additional Provision 17, in particular the requirement that the corresponding Ethical Review Board includes a person with data protection knowledge and responsibility.

3.7. The RDC-GPAP user, acting as a controller in its own right, must also have a legal basis for collecting and processing sensitive data. The appropriate primary legal basis will depend on the RDC-GPAP user's legal nature (public authority, health care provider, research institute) and its relationship with data subject (direct care, clinical trial, research project).

3.8. The legal bases used by the RDC-GPAP user shall be confirmed in the contract between Consorcio CNAG and the researcher, these may be:

3.8.1. The **consent** of the data subject in accordance with GDPR Article 6(1)(a) coupled with **explicit consent** for processing sensitive data in accordance with Article 9(2)(a); or

3.8.2. The **public interest or legitimate research** purposes in accordance with GDPR Article 6(1)(e) or(f) (depending on the legal nature of the RDC-GPAP user) coupled with the legal basis for processing sensitive data for the purposes of **medical diagnosis** (Article 9(2) (h)) or public interest in **public health** (Article 9(2) (i)) or **scientific research** (Article 9(2) (j)), depending on the legal status of the RDC-GPAP user.

3.9. Access to RDC-GPAP by organizations performing Commercial/Pre-commercial activities will be restricted to fulfilling the objectives of a specific research project previously approved by the DAC. Some results from this research project will be made publicly available to the research community upon agreement with the DAC.

3.10. Data in RDC-GPAP may be accessed by non-EU countries for research purposes under very specific circumstances. Groups from non-EU countries will have signed a Data Access Agreement and a Data Transfer Agreement with RDC-GPAP. In addition, such access shall be based on one of the following criteria being met:

3.10.1. The country in question has been granted an **adequacy decision** in accordance with Article 45(3) GDPR.

3.10.2. **Appropriate safeguards** for such data transfer have been adopted in accordance with Article 46 GDPR. These include: a legally binding instrument between Consorcio CNAG and the receiving organisation if this is a public body; binding clause adopted by the European Commission or Spanish Data Protection Authority in accordance with Article 46 GDPR; an approved Code of Conduct adopted in accordance with GDPR Article 40.

3.10.3. The **explicit consent** of the data subject in accordance with GDPR Article 49 (1)(a) has been obtained and can be documented.

3.10.4. The transfer is necessary for important reasons of public interest – in accordance with Article (49 (1) (d)). This exception is available only if the public interest reason can be identified in national or EU level law. For the purposes of RDC-GPAP such a reason could be established with the Cross-Border Care Directive (2011/24/EU) which creates the ERNs.

Code of Conduct for RDC-GPAP Users

RULE R1: Duty to ensure that data submitted to RDC-GPAP have been collected lawfully

The RDC-GPAP user submitting data to RDC-GPAP for the purposes of patient diagnosis, individual research or for a collaborative research project shall:

R1.1 Confirm that the data submitted to RDC-GPAP have been collected in compliance with the legal and ethical requirements for the collection and processing of such data in their country of registration, and specify the legal basis on which the data were collected.

R1.2 Where the legal basis is the **consent** of the data subject in accordance with GDPR Article 6(1)(a) coupled with **explicit consent** to process sensitive data (Article 9(2)(a)), the RDC-GPAP user shall confirm that appropriate information has been provided to the data subject; and that the consent is sufficient to cover the relevant research purpose of using RDC-GPAP, including the right of the data subject to withdraw consent and withdraw data from RDC-GPAP.

R1.3 Where the legal basis for data processing is **public interest or legitimate interest** in research in accordance

with GDPR Article 6(1)(e) or (f), and the basis of processing sensitive data is **medical diagnosis, public interest in public health or scientific research** in accordance with Article 9(2) (h), (i) or (j) respectively the RDC-GPAP user shall confirm appropriate information has been provided to the data subject.

R1.4 Where the legal basis for processing sensitive data is scientific research in accordance with Article 9(2)(j), the RDC-GPAP user shall confirm that safeguards in accordance with GDPR Article 89(1) have been adopted, meaning that the data have been pseudonymised or anonymised.

R1.5 Where data are submitted in the context of a collaborative project, the RDC-GPAP user shall document any restriction of use or obligation applicable to these data (e.g., the limited scope of purpose imposed by the consent form, the obligation to report incidental findings, etc.).

R1.6 Whichever legal basis is used, RDC-GPAP users shall confirm that information about the use of their data for research has been made available to data subjects, in accordance with GDPR Article 13 (1) and (2). Where submission to RDC-GPAP was not foreseen at the time of data collection, such information shall be provided to the data subject before submission of data to RDC-GPAP in accordance with GDPR Article 13(3).

RULE R2: Duty to protect the privacy of data subjects

R2.1 The RDC-GPAP user shall submit data in pseudonymised form.

R2.2 The RDC-GPAP user is permitted to attach additional local identifiers to profiles or experimental datasets in order to track their data and facilitate their usage of the system. The RDC-GPAP user must ensure that identifiers do not contain any personally identifiable information and that do not allow identification of the individual. Any such labelling is undertaken under the RDC-GPAP user's responsibility as a data controller and does not create legal duties for Consorcio CNAG. RDC-GPAP administrators and the RDC-GPAP interface never receive this information.

R2.3 There will be no attempt to try to identify or contact data or donor subjects.

R2.4 The RDC-GPAP user is not allowed to:

- R2.4.1 Copy or download RDC-GPAP data (with the exception of copying or downloading your own submitted data, or if data is part of a project that specifically allows for it).
- R2.4.2 Share or redistribute RDC-GPAP data (with the exception of sharing or redistributing your own submitted data, or if data is part of a project that specifically allows for it).
- R2.4.3 Sell or in any way commercialize with RDC-GPAP data.

R2.5 Access codes and user logins are specific to the identified user and are strictly non-transferable.

R2.6 RDC-GPAP Group leaders must inform RDC-GPAP in case they change their research Institution. It will be the responsibility of the Group leader to reach an agreement with the involved Institutions regarding stewardship of the submitted data and to communicate it to RDC-GPAP.

R2.7 RDC-GPAP Group leaders must inform RDC-GPAP if any of their RDC-GPAP users leave their research group or Institution.

R2.8 RDC-GPAP users must inform RDC-GPAP of any change in their Institution affiliation or research group pertinence.

R2.9 RDC-GPAP Group leaders that are interested in requesting to unsubscribe from the RDC-GPAP, must contact Consorcio CNAG. The unsubscription of the RDC-GPAP Group Leaders shall imply that the subscription of the RDC-GPAP users of the group shall also be cancelled.

Rule R3: Duty to act responsibly with research results

R3.1 RDC-GPAP Group Leaders are responsible to verify the validity of the results obtained with the support of the RDC-GPAP.

R3.2 The RDC-GPAP user shall ensure that research project results or outcomes can be made available to study participants in a manner allowing non-specialists to understand the results.

R3.3 Where data remain identifiable, the RDC-GPAP user shall ensure that any incidental finding can be communicated to the initial collector of the data in accordance with their relationship with the data subject.

Rule R4: Duty to respect intellectual property

R3.1 Proper attribution and intellectual property should be accorded as appropriate. Sharing of data shall follow criteria for the acknowledgement of intellectual contributions and originality through rules of authorship and intellectual property rights.

R3.2 All RDC-GPAP users are encouraged to follow the publication guidelines described in the "RDC-GPAP Publication Policy" document, available in the RD-Connect GPAP website.

Code of Conduct for RDC-GPAP Service Provision

RULE R5: Duty to ensure that data submitted to RDC-GPAP are processed lawfully

Consorcio CNAG as the provider of the RDC-GPAP services shall:

R5.1 Record in its official documents its legal basis for processing health related data using the RDC-GPAP tools and interface as **'public research'** as provided for in GDPR Article 6(1)(e) along with '**scientific research**' as provided for in GDPR Article 9(2)(j)) as the legal basis for processing sensitive data.

R5.2 Adopt suitable measures to ensure that RDC-GPAP processing is compliant with the guidance and requirements of GDPR Article 89 (1) and Additional Provision 17 of Spanish Organic Law 3/2018.

R5.3 Renew as required the ethical approval to provided RDC-GPAP services received on 27th October 2015 from the Parc de Salut MAR – Clinical Research Ethics Committee (ref. no. 2015/6456/I).

R5.4 Where Consorcio CNAG holds the personal data of RDC-GPAP users, such as their ID information and academic accreditation, as well as data on their use of RDC-GPAP, such data shall be reported as held for the purpose of performance of a contract in accordance with GDPR Article 6(1)(b).

Rule 6: Duty to maintain high level of data security in data processing

R6.1 Consorcio CNAG will store RDC-GPAP data in a cluster in distributed file systems with a restricted access policy, limited internet access and daily backups.

R6.2 Consorcio CNAG will store the de-identified raw data in the EGA, a secure, controlled-access repository which it provides collaboratively with EBI.

R6.3 Consorcio CNAG will ensure that access to and use of RDC-GPAP data is based on the scientific validity, quality and potential success of the request for use.

R6.4 Consorcio CNAG gives no warranty and accepts no responsibility or liability for the accuracy, completeness or validity of the data it contains.

R6.5 The procedure for assessing Group Leader applications for access (user registration) will include validation of institutional email address, validation of research aims, signature of the Adherence Agreement based upon the present Code of Conduct, and approval by the Data Access Committee. The Adherence Agreement should be signed both by the Group Leader and the Legal Representative of the institution of the Group Leader, and the user registration will include the validation of the signatures (either the eSignatures or handwritten signatures with copies of the ID documents). Group Leaders can include other users in their group and under their responsibility by providing a full name and a valid institutional address to RDC-GPAP.

R6.6 RDC-GPAP will perform RDC-GPAP user access controls regularly. Access will be automatically revoked if the renewal request is not acted upon or if the original reason for access is no longer valid (e.g. when a user moves to a new institution). Any user will be able to end their participation at any time upon request.

R6.7 Any detected misconduct or malpractice by an RDC-GPAP user will be communicated to the DAC. The DAC will then have the authority to revoke RDC-GPAP user access to RDC-GPAP. If necessary, legal actions might be pursued in case of misconduct or malpractice.

R6.8 Information about action performed by former users and relationship to the datasets submitted by them will be retained for audit and data provenance purposes. For traceability, all submitted datasets are linked to the username of the data submitter within the system. To facilitate connections between users who may have cases of interest to each other, when a dataset becomes accessible to other users within the RDC-GPAP, the username of the data submitter will become visible to the other authorised users of the system viewing that dataset.

R6.9 For transparency reasons Consorcio CNAG may publish a list of names of the institutions to which its authorised users belong. The name of the individual users or groups will not be published without explicit

consent.

Rule 7: Duty to respect intellectual property

R7.1 Proper attribution and intellectual property should be accorded as appropriate. Sharing of data shall follow criteria for the acknowledgement of intellectual contributions and originality through rules of authorship and intellectual property rights.

Rule 8: Duty to manage the closure of RDC-GPAP

R8.1 In the case that the RDC-GPAP will be discontinued, users will be notified and transparent mechanisms to transfer the data to long-term funded European infrastructures will be sought.

Annex I

Key terms in Alphabetical Order

Some terms are specific to RDC-GPAP and some are terms found in the GDPR

Adequacy Decision

Cross-Border Data Transfers to a recipient in a third country may take place, without a need to obtain any further authorisation, if the Commission has decided that such third country ensures an adequate level of data protection (an "Adequate Jurisdiction"). The basis for this principle is that such jurisdictions provide sufficient protection for the rights and freedoms of data subjects without the need for further safeguards.

Aggregated data

Means the Data of several individuals that have been combined to show general trends or values without allowing identification of individuals.

Anonymous data

Means there are no links to the individual donor, the data and bio-specimens were never associated with identifiers, and the risk of identification of individuals is very low. There may be general descriptions such as ‘man, aged 50–55 years, cholesterol level 240 mg per 100 ml’.

Anonymised data

Means data and samples that have been identified earlier or pseudonymised, but the identification, or the code and the code key have been destroyed, and thus there is no longer any link to the individual.

Data controller

Means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law. (As defined in GDPR, Article 4).

Data identifier

Means a code that enables identification of a dataset within a database. The RDC-GPAP assigns an identifier to each patient phenotypic profile and to each experimental dataset on submission and maintains links between the two to enable joint genome-phenome analysis. These identifiers are alphanumeric codes provided by the database and are not based on or related to any personal information. Data submitters are allowed to provide additional local identifiers for each profile or experimental dataset submitted in order to track their data and facilitate their usage of the system. It is the responsibility of the submitters to provide identifiers that do not contain any personally identifiable information and that do not allow identification of the individual. Data submitters are responsible for maintaining any pseudonymisation links between the identifiers used in the RDC-GPAP and the personal data that they may hold on their patients in a secure and GDPR-compliant manner in their local environment. The RDC-GPAP never receives this information.

Data processor

Means the natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller.

Genetic data

Means all personal data relating to the genetic characteristics of an individual, inherited or acquired, as a result of an analysis of a biological sample from the individual in question, in particular by chromosomal, DNA

or RNA analysis or analysis of any other element enabling equivalent information to be obtained.

Human biological samples

Means the constituent parts of the human body, or human biological material, including organs and parts of organs, cells and tissues, and body fluids.

Identified data

Data labelled or linked to the individual in a way that makes them directly identifiable (name and surname or social security numbers). There is no identified data within the RDC-GPAP.

Identifiable person

One who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, genetic, mental, economic, cultural or social identity. (As defined in GDPR, Article 4, see "Personal data")

Incidental finding

Means a finding concerning an individual research participant that has a potential health or reproductive importance and is discovered in the course of conducting research or diagnostic procedures but is beyond the aims of the study.

Personal data

Means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. (As defined in 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 (General Data Protection Regulation, hereafter referred to as GDPR).

Personal data may be available as recorded and registered data or in the format of human biological samples under the provision that there is a code through which the sample can be connected to the identity of an individual Identifiable person.

Pseudonymised data

Means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person (GDPR Article 4.5). [Guidelines for pseudonymisation are expected to be part of the Code of Conduct on Processing of Personal Data for Purposes of Scientific Research in the Area of Health]. The RDC-GPAP contains pseudonymised data.

Processing of personal data

Means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction. (As defined in GDPR, Article 4.2.)

Secondary use of data

Means the processing of already existing health-related personal information for a purpose different from the purpose for which it was originally collected.