Code of Conduct for user access to the RD-Connect Genome-Phenome Analysis Platform (GPAP) for health-related information[1]

Effective as of 25 May 2018

Revised and approved by the RD-Connect Executive Management Committee (EMC) as of 22 May 2018 with reference to the General Data Protection Regulation (GDPR), to be applied as of 25 May 2018

Communication to Mats G. Hansson, Centre for Research Ethics & Bioethics at Uppsala University: mats.hansson@crb.uu.se

1. Preamble

RD-Connect was established in November 2012 through a grant from the European Commission under the seventh framework programme (FP7). It provides infrastructure, tools and resources to facilitate and accelerate rare disease research by maximizing the availability, analysis and (re)use of rare disease data and biosamples. It is sustained on an ongoing basis by European and national funding mechanisms and close connection with pan-European biomedical research infrastructures, in particular ELIXIR and BBMRI-ERIC.

This Code of Conduct has been developed to regulate the terms on which users gain access to the RD-Connect Genome-Phenome Analysis Platform. Other RD-Connect tools and resources share the same goal of enabling rare disease research and data and sample sharing for the benefit of patients. They may have related but distinct Codes of Conduct to serve their particular purpose.

2. Definition of terms

RD-Connect Genome-Phenome Analysis Platform (GPAP)

The RD-Connect Genome-Phenome Analysis Platform (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 on individuals with rare diseases. The GPAP interface enables clinicians and researchers to analyse and interpret the full genomic datasets they submit for both diagnosis and gene discovery on an individual patient basis and to link these with phenotypic data and biosample availability for the same individual.

[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, in which form it was effective from 9 November 2015 until the coming into effect of this revision. It is adapted from a draft version by EHR4CR, 27 August 2014. It was revised in Barcelona 27 January 2015, at the Annual Meeting of RD-Connect, acknowledging comments from: Bartha Knoppers. Olaf Riess, Manuel Posada, Peter-Bram 't Hoen, Mirella Filocamo, Marina Mora, Sabina Gainotti, Pauline McCormack and Mathieu Boudes on previous drafts. It was approved at the Executive Management Committee of RD-Connect on 11 June 2015. It was approved by Parc de Salut MAR - Clinical Research Ethics Committee in Barcelona on 29 October 2015. It was revised by Mats G. Hansson and Jane Reichel as of 3 November 2017 with reference to the General Data Protection Regulation (GDPR) and revised by Sergi Beltran, Rachel Thompson and Mats G. Hansson as of 18 May 2018 with reference to the specific mechanisms by which the RD-GPAP is implementing the GDPR requirements. It was approved by the RD-Connect Executive Management Committee on 22 May 2018.

It should always be referred to according to the effective date. It may need revisions due to changes in the regulatory frameworks.

Personal data

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) , GDPR, Article 4)

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 donor.

Human biological samples

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

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")

Processing of personal data

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)

Data controller

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 processor

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

Secondary use of data

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

Identification, coding and anonymised data

As defined in the GDPR and adapted here to fit both personal data and identifiable biospecimens the following definitions apply:

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 RD-Connect GPAP.

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 RD-Connect GPAP contains pseudonymised data.

Anonymised data

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.

Anonymous data

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'.

Data identifier

A code that enables identification of a dataset within a database. The RD-Connect 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 RD-Connect GPAP and the personal data that they may hold on their patients in a secure and GDPR-compliant manner in their local environment. The RD-Connect GPAP never receives this information.

Aggregated data

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

Incidental finding

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[2].

Genetic data

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.

[2] As defined in Wolf SM, Lawrenz FP, Nelson CA et al: Managing incidental findings in human subjects research: analysis and recommendations. J Law Med Ethics 2008; 36: 219–248, 211).

Broad consent

Consent for an unspecified range of future research projects subject to a few content and/or process restrictions, e.g. ongoing governance.

Withdrawal of consent

A research subject has the right to withdraw consent at any time without justification. In case research subjects withdraw their consent for use of their data, the data will be maintained up to the point of withdrawal. The data controller is responsible for deleting all data to be withdrawn, unless an applicable law would prevent it.

2. Aim

The aims of this Code of Conduct are: to enable the process of clinical development in rare diseases through research; to ensure collaborative research towards innovation, quality, safety and value in health care; and to protect the privacy of patients and research subjects.

3. Legal and ethical framework

All proceedings in accordance with this Code of Conduct should be compliant with national and European law. It is the national law of the data controller/person responsible for a registry or collection of personal data that applies to the processing of data, irrespective of where the data is used. Processing must be compliant with the provisions of the informed consent and/or decision of an Ethical Review Board, or else compliant with national law.

The aim of promoting clinical development through research while protecting privacy of patients and research subjects is motivated by The Charter of Fundamental Rights of the European Union (2010/C 83/02). This charter emphasizes the right of each individual to integrity within the fields of medicine and biology, implying a free and informed consent according to the procedures laid down by law (Article 3). Article 8 of the Charter grants the individual the right to the protection of personal data implying that the processing of such data requires consent of the person concerned or other legally-recognized means. These articles conform to the European Convention for the Protection of Human Rights and Fundamental Freedoms, the Social Charters adopted by the European Union and by the Council of Europe.

These rights may be motivated by a fundamental respect of each individual's autonomy and right to have control of matters related to oneself, e.g. the processing of personal data. They may imply a right to know about genetic and other medical information about oneself but also, as has been frequently discussed in the ethical and legal literature, the right not to know such information. In addition to these autonomy rights the Charter of Fundamental Rights of the European Union also lays down rights of each individual to social security benefits and social services in cases of illness (Article 34) and the rights of access to preventive health care and the right to benefit from medical treatment under the conditions established by national laws and practices (Article 35). As described, the charter of the European Union recognizes both the privacy right leading to requirements of respecting autonomy, obtaining consent etc., and the right to health care and social services in cases of illness as fundamental individual rights, notwithstanding that there may also be societal and public health related interests concerned.

In order to balance these fundamental rights there are several information and consent procedures available to be used in different contexts. Broad consent for the use of pseudonymised or anonymised and aggregated data is the preferred solution in prospective sampling and collection of data. Renewed consent for secondary purposes using previously collected samples and data where broad consent has not been used is recommended as long as this is practically feasible and does not jeopardize the right of patients and research subjects to benefit from medical research. In such cases the basis for processing of personal data should be replaced with a decision by an Ethical Review Board.

Sharing data and biological samples (bio-resources) is essential for accelerating biomedical research projects that provides benefits to current and future patients. Research conducted through biobanks and registries is more effective if access to sufficient and quality data and biological samples is granted. The use of data and samples is maximized through data sharing. Ensuring secure data and sample sharing ethically and legally protects bio-resources, as well as the wishes of the donors to contribute to research.

The following five principles for the stewardship of bio-specimens and data repositories constitute the common premise for sharing and access and apply to all users of the RD-Connect GPAP[3].

I. Respect for privacy and autonomy: stewardship implies protection of participants' privacy. Privacy protection measures should be in place and informed consent must provide provisions for future research purposes described in general terms using data and bio-specimens.

II. Reciprocity: stewardship also implies giving back. Feedback of general results should be channelled to institutions and patients.

III. 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.

IV. 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.

V. 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.

Article 1 – General provisions

The principles of protection do not apply to data that have been rendered anonymous and are not identifiable. To determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person. This Code aims at addressing the processing of data to be used or re-used in collaborative research projects.

This Code is inter alia compliant with:

- Art. 8 of the European Convention on Human Rights
- Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data 1981 (convention 108)
- Charter of Fundamental Rights of the European Union (2010/C 83/02)
- GDPR of the European Union (2016/679)

Article 2 – Collection, use and transfer of data

RULE 1: The entity providing personal data to the collaborative project shall verify that:

- The initial collection of these data has been compliant with the requirements of the original purpose.
- The collection and the provision of the data to the collaborative research projects meets all legal requirements to which the entity is subject.
- Further storage and processing of the data after completion of a collaborative research project is in compliance with applicable law.

[3] See: Mascalzoni D, Dove E, Rubinstein Y, Dawkins H, Kole A, Mc McCormack P, Woods S, Riess O, Schaefer F, Lochmüller H, Knoppers B, Hansson M, International Charter of Principles for Sharing Bio-specimens and Data, European Journal of Human Genetics, Online: DOI EJHG.2014.197 and see also the Global Alliance for Genomics and Health, Framework for Responsible Sharing of Genomic and Health-Related Data (www.genomicsandhealth.org)

RULE 2: The entity which provides personal data to a collaborative project 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.).

RULE 3: The secondary use of data in collaborative research projects shall take place using anonymised data. If, however, the purposes of the research cannot be achieved using anonymised data, pseudonymised data may be used pursuant to this Code. The project shall decide, based on the scientific purpose, legal constraints, and the benefits or risks of harm to research subjects, whether anonymised or pseudonymised data shall be used.

RULE 4: In compliance with this Code, pseudonymised and anonymised data can be shared within the scope of the project independently of geographic boundaries.

RULE 5: Pseudonymised data is considered personal data (in accordance with Article 4 GDPR)

Article 3 – Protection of anonymised data

RULE 6: Aggregated data is considered anonymised data.

RULE 7: Anonymised data should receive adequate protection against future risks of re-identification.

Article 4 – Information, consent and withdrawal

1. Basic information and consent for prospective research projects

RULE 8: Information and consent procedures shall be approved by the relevant ethics board.

RULE 9: Data collectors collecting personal data for a prospective collaborative research project shall inform the study participants about the project in an appropriate manner, including:

- the identity of the data controller
- the voluntariness of the collection of data/samples
- the broad/general purposes of the processing
- the nature of the processed data, including its type (identifiable, pseudonymised, anonymised)
- the existence of the right of access to, and the right to rectify the data concerning them
- if the research project reasonably anticipates the sharing of data across research groups and national borders
- if complementary information will be added through linkage to different data registries, medical records, etc.
- if the project involves genetic and other molecular analyses
- if the project involves collaboration with both academic and commercial partners
- foreseen secondary use of study participants' data (anonymised or pseudonymised) in future research projects
- if information about project results and incidental findings will be provided and the process for this if applicable, and
- that consent may be withdrawn and how this is done

2. Information and consent procedures for previously collected samples and data

RULE 10: For the use of previously collected data and bio-specimens where consent is absent or not fit for purpose, an assessment has to be made by an ethics board governed by applicable national law. Re-consent, notification, opt-out and implied consent or absence thereof are all available and legitimate depending on national legislation and should be selected after an inventory of available appropriate safeguards, a proper

balancing of autonomy rights, benefits, risks of harm and right to medical care as described in the legal and ethical framework of this Code of Conduct.

3. Information on project results and incidental findings

RULE 11: Research project results or outcomes should be made available to study participants in a manner allowing non-specialists to understand the results.

RULE 12: If data are still identifiable, any incidental finding should be communicated to the initial collector of the data in accordance with what has been described in the information and consent form.

4. Consent withdrawal

RULE 13: Study participants have the right to withdraw their consent at any time without justification.

RULE 14: In case study participants withdraw their consent for use of their data, all data but not already derived data or results shall be erased from all research project databases (if still identifiable and unless applicable law requires further storage or use).

Article 5 – Sharing and access to data

RULE 15: Data Transfer Agreements, subject to the authorisation from the competent supervisory authority in accordance with Article 46 GDPR, should be used for parties outside the EU, unless there are specific legal provisions between the EU and the third country (e.g., Privacy Shield for U.S., or countries accepted by the European Commission in accordance with an Adequacy Decision under Article 45 GDPR or equivalent such as Commission Decision 2001/497/EC, C(2004)5721.

RULE 16: Data shall always be collected, stored and exchanged in a secure manner, through secure channels. For the RD-Connect GPAP, data will be stored in a cluster with a restricted access policy, limited internet access and daily backups. Databases use distributed filesystems, limiting the risk of physical attacks. All communications are encrypted. Security of the RD-Connect GPAP was audited in October 2017 with no major risks being identified. Platform requests and user actions are securely logged for audit purposes.

RULE 17: Access to and use of data should always be based on the scientific validity, quality and potential success of the request for use. By default, submitted data will be available to the other verified and authenticated users of the platform immediately. Those who contribute data are entitled to a maximum of six months from submission for exclusive use of the data, easily selectable during the online submission process. If there is an important reason why the embargo period should be extended beyond this period, requests can be made using the appropriate online mechanism at the time of submission.

RULE 18: The RD-Connect GPAP will provide transparent procedures for initial authorisation of access to the platform, periodic renewal and for the running of the services of the platform. Full information about all access procedures is always available through a link from the main RD-Connect platform webpage at https://platform.rd-connect.eu. The procedure for assessing applications for access (user registration) includes validation of user identity with a copy of passport/national ID, validation of institutional email address, validation of research aims of the user and signature of the Adherence Agreement for this Code of Conduct. Validation of user identity and signature of the Adherence Agreement will only be necessary for Principal Investigators, Group Leaders or similar positions, who can then include other users in their group under their responsibility. However, all users will be asked to abide by this Code of Conduct and will be periodically required to explicitly renew their authorisation for access in order that the GPAP can be certain that all users have an ongoing valid reason for access. 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 revoke their participation at any time upon request. Information about actions

performed by former users and relationship to the datasets submitted by them will be retained for audit and data provenance purposes.

RULE 19: 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 GPAP, the username of the data submitter will become visible to the other authorised users of the system viewing that dataset. Data submitters are entitled to view which other users have queried their data within the RD-Connect GPAP. When a user tags a genetic variant in the system (e.g. as likely pathogenic or pathogenic), the content of the tag and the name of the user will be visible to other users viewing that variant in the system.

RULE 20: For transparency reasons, the RD-Connect GPAP 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 21: 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 22: Quality of data shall be ensured by the provider.

RULE 23: In the case that the RD-Connect GPAP will be discontinued, users will be notified and transparent mechanisms to transfer the data to long-term funded European infrastructures will be sought.