DATA SHARING POLICY

As specified in the contest of free science and as described in the EU programme for research and innovation H2020, Horizon Europe and Programma Nazionale della Ricerca Sanitaria 2020-2022, data sharing is fundamental for the medical-scientific community. Accordingly, it encourages more connection and collaboration between researchers, which can result in important new findings within the field. Promoting the relevance of the huge amount of data obtained by clinical trials, a new instructive resource can be applied to reduce costs and time, to support the research, to optimize the care for the patient and to guide strong facts based future choices.

DOMAIN OF INTEREST/APPLICATION

This policy is applied to the preservation, sharing, security and treatment of personal data realized by all Operative Unites/joint organizations of Fondazione IRCCS Istituto Neurologico “Carlo Besta”.

OWNERSHIP OR CONTROL OVER DATA

The preservation of research data in a specific institutional respository is mandatory not only for the sharing of data themselves, but also because they can be useful to verify the integrity, the accuracy and the reproducibility of the research processes and results. As specified in D.Lgs. 196/2003 (“Codice in materia di protezione dei dati personali e sensibili”) and in the subsequent authorizations ("Autorizzazione Generale al trattamento dei dati genetici" 8/2014, "Autorizzazione generale al trattamento dei dati personali effettuato per scopi di ricerca scientifica" 9/2014), data and metadata will be stored following the in force regulations about personal data protection and the FAIR principles (Findability, Accessibility, Interoperability, and Reusability). In addition, every dataset will include information on what can be reused and what instead must not be shared.
Istituto Neurologico “Carlo Besta”, as Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) and owner of data treatment, in compliance with art. 9 par. 2, lett. j), of GDPR and with art. 110-bis comma 4 of D.lgs. 196/2003, informs that data could be used for:

1) Purpose of the research;
2) Lead of retrospective observational studies, supported by the Foundation, subjected to the achievement of the necessary authorization by the Ethics Committee in charge;

3) Subscription to the international registers, promoted and created by bodies and research institutions, consortia and societies, where data are stored and shared between researchers, in compliance with the principle of open science and open access to research data, data series (DNA sequences, protein structures, data from preclinical research and clinical experimentations, epidemiological data etc.). Fundamental is the fact that the patient data will be shared in anonymized or pseudonymized form and accessible only to authorized users.

Often the data that has to be shared contains personal or confidential information. To respect data protection and safeguard the rights and the fundamental freedom of those involved, it is necessary to verify that the owner of data usage has the informed consent from the patient. The informed consent, based on art. 4 of GDPR, is any free, specific, informed and unequivocal manifestation of will that he/she gives his/her approval, through an unequivocal declaration or positive action to use his/her personal data. Nevertheless, the sharing of data will happen with prior removal of sensitive data. This process of removal or concealment is called anonymization. These types of personal and confidential information comprehend:

- Personal data with names, addresses and ID numbers;
- Financial data about individuals or legal entities;
- Data that if aggregated can trace back the identity of an individual;
- Special categories of personal information such as "ethnic or racial origin, political, religious or philosophical opinions, genetic data, biometric data, sexual orientation etc" as stated in art.9 of GDPR.

The most straightforward way to proceed is the removal of these fields of data before sharing or publishing. Nevertheless, the simple removal of results can induce loss of data themselves and their sharing can lose its usefulness. Therefore, there is an additional method of anonymization, called pseudonymization, which is, as defined by art. 4 clause 5 of GDPR, "the use of personal data in a way that they cannot be related to a specific individual anymore without using additional information". This additional information will be stored separately and will undergo technical and organizational measures to guarantee that the personal data will not be attributed to an identified or identifiable individual.

TECHNIQUES/TECHNOLOGIES USED

Fondazione IRCCS Istituto Neurologico “Carlo Besta” as general guidelines, manages the research data sharing through these practices and instruments:

- REDCap: platform for the design of Case Report Form (CRF) for research. Especially, it is dedicated to the collection of data and clinical metadata and of translational research;
- XNAT: platform meant for the collection of data of neuroimaging such as images of magnetic resonance and CT-scan;
- Office 365 with institutional email and One Drive OTP (One-Time Password);
- Every file containing personal data or confidential will be protected by passwords;
- The creation of a user will occur by sharing the credentials by email and encrypted file. The file unblocking code must be sent with another method (ex. smartphone);
- Once the research project ends and once the database has been closed. The collected data will be stored in specific CD-ROM and DVD, that will be encrypted and protected with a password.

DATA STORAGE/USAGE

Shared data can contain personal or confidential information. If a research project requires to obtain patient data, it is the responsibility of the institutes involved to maintain ethical and legal standards, both during data usage and sharing. Therefore, in compliance with GDPR, Fondazione IRCCS Istituto Neurologico "Carlo Besta" will guarantee that these rules will be respected:

- Protection of patient identity;
- Notice of data sharing included in informed consent;
- Control of accesses of shared data.

In accordance with the points above, an informed consent form will be given to the patients, and will concern both protection and future usage of collected data. Especially:

- Modalities of data storage and confidentiality;
- Sharing of collected data;
- Sharing of extracted analyses;
- Storage of data and their usage in the long period;
- Removal of data once they are no longer necessary for research purposes.

Please refer to the data usage and procedure for their collection for further details related to management.

LICENSES

In terms of licenses, it is fundamental to consider two elements:

- Copyright
- License

The author of the work and the one that processes the data in digital format (publications, work sheets, informatic programs or diverse kinds of report) are those that possess copyright rights. When data or results derive from a collaboration, the different authors or institutions can share the copyright right. Every secondary user, instead, requires defined licences before using data protected by copyright.

Following the main themes regarding open science of the H2020 and Horizon Europe and of Programma Nazionale della Ricerca Sanitaria 2020-2022 programs and to allow the revision and check between peers, Fondazione IRCCS Istituto Neurologico "Carlo Besta" encourages the fulfillment of principles of open data. Open data are those data that can be freely used, re-used and shared by anyone, with the only limitation of citing the origin and sharing with the same type of license with whom they have been first released. Therefore, it is mandatory to confer to the data the last available version of the Creative Commons Attribution International Public License (CC BY) or Creative Commons Public Domain (CC 0) or one license with equivalent rights, following the principle "as open as possible, as the close as necessary".

DATA DISCOVERABILITY AND REUSE

Documenting data is vital when creating, organising, and managing data due to its importance not only for data discovery but also for (long-term) data preservation. In order to enable the shared data to be broadly discovered and reused the data has to be prepared. It would not be sufficient to collect and provide data to potential users. Instead, it is crucial to describe (document) the data to be shared in a discoverable and understandable way: how data was created, what it means, its structure and content (e.g. origin, purpose, creator, access conditions, and terms of use of the data collection). Documentation of data is done via the use of metadata: descriptors used to represent, trace, use and manage the objects deposited in an informatic system such as an institutional archive (tecnical, descriptive, semantic, structural metadata). Providing structured and well-defined searchable information helps the users find and classify the data underneath. Fondazione IRCCS Istituto Neurologico "Carlo Besta" will guarantee that metadata of the products registered in the institutional archive are always freely accessible and anonymized or pseudononymized.

DATA ACCESSIBILITY

The use of data is limited to the research participants, those present in the informed consent. To obtain access to these data, secondary users must accept and respect the conditions imposed by the applied licenses.
Nevertheless, the access to confidential data can be further ruled by:

- Specific authentications/procedures of authentication;
- Limited access to qualified users;
- Limited access only for data analyses, forbidding usage and download of data;
- Removal of confidential data during users' qualification.

Modalities of access and types of used data will be decided through a reciprocal deal, documented with a license, between the user and the data owner.

CHECKLIST

Does the storage subsystem provide redundancy against data loss?	
Does the storage subsystem provide redundancy for data availability?	
Is the confidential data encrypted at rest?	
Is network access of the storage subsystem protected by firewall?	
Is there a defined back-up procedure?	
Is authenticated access to the storage subsystem allowed to the smallest number of users?	
Are all default passwords, accounts and configurations removed or updated?	
Is workstation access authenticated?	
Are workstations full-disk encrypted?	
Are all user accounts in the databases accounted for and documented?	
Do storage maintenance procedures mandate supervised-only access to maintenance personnel?	
Is physical security of the storage subsystem ensured?	
Does the storage subsystem allow traceability of changes?	