DMP du projet "Optimizing response to Li treatment through personalized evaluation of individuals with bipolar I disorder: the R-LiNK initiative"

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "Horizon 2020 FAIR DMP (anglais) - Personnalisé" fourni par CEA Commissariat à l'énergie atomique et aux énergies alternatives.

Renseignements sur le plan

Titre du plan	DMP du projet "Optimizing response to Li treatment through personalized evaluation of individuals with bipolar I disorder: the R-LiNK initiative"
Langue	fra
Date de création	2018-02-15
Date de dernière modification	2020-01-20
Identifiant	R-LiNK

Renseignements sur le projet

Titre du projet	Optimizing response to Li treatment through personalized evaluation of individuals with bipolar I disorder: the R-LiNK initiative
Résumé	Scientific description - Bipolar disorder (BD) is a prevalent mental disorder and a leading cause of suicide. Lithium (Li) is the key mood stabilizer for prevention of BD relapse and suicide. Whilst many cases become asymptomatic with Li treatment, the majority show sub-optimal response. Identifying biomarkers for predicting Li response would enable personalization of treatment, define criteria for stratification of BD cases and further refine the clinical response phenotype. The objectives of this proposal are to (i) improve outcomes of bipolar I disorder (BDI) cases prescribed Li through the application of stratified approaches; (ii) optimize the early prediction of Li response using a set of multi-modal biomarkers (“blood omics”, Magnetic Resonance Imaging and Li7-Magnetic Resonance Imaging derived-markers); (iii) develop a multidisciplinary multinational network of experts to undertake this and future projects on personalized diagnostics and therapeutics; and (iv) implement new, powerful technologies to characterize brain Li distribution and the blood molecular signature of Li in responders and non-responders. This cutting edge approach will identify the eligibility criteria for treatment with Li in BD in terms of response, safety and tolerability. The assessment of each putative biomarker (singly and combined) will be guided by preliminary findings already obtained by R-LiNK; our expertise will allow exploratory analyses and innovative modeling of multi-modal data. Likely impacts include improved outcomes and quality of life for BDI cases; development of a screening tool for clinicians; and an evaluation of the cost-effectiveness of this stratified approach. The network will develop new avenues of research on Li mechanisms of action and disease mechanisms; our industrial partnerships will enable development of medical devices to improve treatment adherence and patient’s autonomy, diagnostic kits, and tools based on the molecular signature in treatment responders. The data sharing strategy will follow the "as open as possible as close as necessary" principle. The Executive committee will regulate the data sharing following (i) the scientific objectives of the R-LiNK consortium's members; (ii) the regulatory constraints (GPRD, etc.) and (iii) the European recommendations about openscience.
Sources de financement	• European Union's Horizon 2020 research and innovation programme : 754907

Produits de recherche :

1. Generation of patient sheets and labels for biolgical samples (Logiciel)

Contributeurs

Nom	Affiliation	Rôles
Dimitri Papadopoulos Orfanos - https://orcid.org/0000-0002-1242-8990		• Responsable du plan
Frank Bellivier - https://orcid.org/0000-0002-3660-6640		• Coordinateur du projet • Personne contact pour les données

DMP du projet "Optimizing response to Li treatment through personalized evaluation of individuals with bipolar I disorder: the R-LiNK initiative"

1. Data summary

Provide a summary of the data addressing the following issues:

- **State the purpose of the data collection/generation**
- **Explain the relation to the objectives of the project**
- **Specify the types and formats of data generated/collected**
- **Specify if existing data is being re-used (if any)**
- **Specify the origin of the data**
- **State the expected size of the data (if known)**
- **Outline the data utility: to whom will it be useful**

Two clinical studies will be conducted by the R-LiNK project.

The first study is a prospective, multicentre clinical cohort study of individuals with bipolar 1 disorder (BDI) initiating lithium and examining putative early biomarkers predicting long term response/non-response to lithium (Li) treatment. We recruit a naturalistic cohort of **300 patients** initiating Li in all phases of the disorder. We collect data for extensive baseline assessment and intensive follow-up over two years, in **15 recruitment centres** across Europe. Baseline data are used to identify predictors of long-term response/non-response to Li treatment. All components of this study allow us to establish if incorporating any measures of potential biomarkers into day-to-day clinical practice is both clinically viable and economically beneficial. Collected data are :

- **Clinical data** are recorded by recruiting centres directly into a centralised eCRF (electronic case report form) system. Patients are invited to self-report daily symptoms and medication adherence between assessments in a dedicated mobile health system. Data are exported into files in tabular format such as CSV (comma-separated values) or TSV (tabulation-separated values).
- **Biological samples** are collected before and 12 weeks after Lithium initiation. Blood samples are prepared for total blood mRNA, circulating microRNA, metabolomics and proteomic studies. For a given modality, all samples are processed by the same platform. Data are then published in a file format appropriate for each specific modality, independant of any specific programming language.
- **Neuroimaging data** are collected before and 12 weeks after Lithium initiation. The MRI protocol includes T1-weighted, FLAIR and diffusion imaging (DTI) images of the brain,
  single voxel proton magnetic resonance spectroscopy (H-MRS) and is implemented in all recruiting centres of the study. In some centres only, an additional Li7-MRI protocol is implemented to measure brain Li concentration and distribution. Imaging data are collected in DICOM format. Raw data are quality-controlled then preprocessed. Raw and proprecessed data are pusblished in NIfTI (Neuroimaging Informatics Technology Initiative) format.
- Sleep-wake cycle and activity patterns are collected who agree to wear an actiwatch for a minimum of 120 consecutive days from study entry.

The second R-LiNK clinical study is implemented in two centres (Paris and Besançon). We evaluate the feasibility, safety and patients' acceptability of a home-based device to monitor salivary Li levels.

These two studies are independent in terms of study sample.

---

2. FAIR data

2.1 Making data findable, including provisions for metadata:

- **Outline the discoverability of data (metadata provision)**
- **Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?**
- **Outline naming conventions used**
- **Outline the approach towards search keyword**
- **Outline the approach for clear versioning**
- **Specify standards for metadata creation (if any). If there are no standards in your discipline describe what**

metadata will be created and how

The R-LiNK dataset is continuously augmented and improved as acquisition centres send data, curators discover and fix errors, data are processed and published. We plan on regularly releasing new versions of the R-LiNK dataset to take into account these changes while providing some stability to end users.

We will not only assign a Digital Object Identifier (DOI) to the R-LiNK dataset as a whole, but also a distinct DOI to each successive release of the dataset. A public landing page will be created for the dataset and each formal release will be described on this landing page or on its own landing page. To avoid excessive storage cost, we might not keep all successive releases of the R-LiNK dataset.

We will strive to use metadata and data organization standards whenever possible, such as the BIDS format for neuroimaging data. In other cases metadata will be available as free text in the accompanying documentation.

At this point we do not plan on providing a facility to search data by keywords.

Each recruitment site is identified by a 2-digit number and a standardized short name. Each patient enrolled in R-LiNK is identified by a 5-digit pseudonym, starting with the 2 digits of the recruitment site followed by a 3-digit incremental number.

We will use a two levels pseudo-anonymization procedure: the first level will occure at the recruiting center and the second before publication of the data.

2.2 Making data openly accessible:

- **Specify which data will be made openly available? If some data is kept closed provide rationale for doing so**
- **Specify how the data will be made available**
- **Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?**
- **Specify where the data and associated metadata, documentation and code are deposited**
- **Specify how access will be provided in case there are any restrictions**

R-LiNK data will be initally shared between members of the consortium and their collaborators. One reason is that we need time to curate data, typically from 6 months to 1 year. Another reason is we need to address privacy concerns in the context of the GDPR and additional national rules. Opening medical research data has always been a legal challenge for research institutions and we need to investigate the **costs of enforcing necessary restrictions in the long term**. We plan on opening the R-LiNK data to the scientific community, upon request, as soon as we have addressed these issues.

We keep data files sent by acquisition sites or processed by working groups on a storage system located in CEA server rooms. These files could be indexed and/or imported into a database to allow queries. The benefit of using a database needs to be further discussed, and balanced with the costs of database maintenance, especially in the long term beyond the end of the R-LiNK project.

Data release for scientific production shall take into account the type consent each patient has given (use of the data within the consortium only, also in collaboration with external academic partners only, also in collaboration with external private partners). Principal investigator in charge of the analysis and the draft of the scientific production will have to submit a “Data request form” to the Executive Committee. After approval, the data release will be organized.

Data will be published on a server with direct access to a storage system in a CEA server room. The server allows only authenticated and encrypted access, such as HTTPS or SSH/SFTP. The exact modalities of access to the server are decided by project management. Once access has been granted, we create an account and assign access rights in accordance with project policy and the needs of the user.

The infrastructure used to collect most of the data, operated by the CATI for imaging and AP-HP for clinical data, is not open source. On the other hand we would like scripts used to process data to be published on GitHub or GitLab.

2.3 Making data interoperable:

- **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.**
- **Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?**

AP-HP export clinical data from the eCRF as tables. We are not aware of standards for table or column names in our research domain, therefore tables and columns need to be explained in a specific document. CEA integrate clinical data with the rest of the data and publish them as tables available in CSV or TSV format, accompanied by their documentation.

Imaging data are sent by acquisition centres to CEA in DICOM format. CEA will convert files to NIfTI format and leverage the BIDS emerging standard, specific to neuroimaging studies, to organise NIfTI files and decorate them with metadata in associated JSON files.

Genomics data are produced by the IRCSS from the biological samples. They are sent to CEA as files in standard formats, such as:

- PLINK for genotypic data,
- language-independant self-describing file formats, typically NetCDF for array-oriented data and HDF5 for more complex data structures.

2.4 Increase data re-use (through clarifying licenses):

- **Specify how the data will be licenced to permit the widest reuse possible**
- **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed**
- **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why**
- **Describe data quality assurance processes**
- **Specify the length of time for which the data will remain re-usable**

Due to the sensitive nature of medical research, sharing R-LiNK data outside members of the consortium and their collaborators is not straightforward. We will evaluate if and how part of the R-LiNK data can be opened more widely. Minimal requirements include further data de-identification and authorization from the Ethics Advisory Board and data protection authorities.

---

3. Allocation of resources

Explain the allocation of resources, addressing the following issues:

- **Estimate the costs for making your data FAIR. Describe how you intend to cover these costs**
- **Clearly identify responsibilities for data management in your project**
- **Describe costs and potential value of long term preservation**

Question sans réponse.

---

4. Data security

Address data recovery as well as secure storage and transfer of sensitive data

We will use a two levels pseudo-anonimization procedure: the first level will occure at the recruiting center and the second

before publication of the data. Only recruitment sites can convert between first level pseudonym and identifying data. Recruitment sites ship only pseudnymized data. Only, the centralized database can convert between first and second level pseudonyms.

Therefore, even if a link in the processing chain is compromised, it will be impossible to trace the identity of the individual.

Acqusition centres and data processing centres shall follow their local IT rules. However we typically expect data to be kept on secure storage, either in a server room or on encrypted disks, and not on a unencrypted USB keys, laptops or external disks.

The CEA implements its own IT security policy (*PSSI*) which is directly derived from the state IT security policy (PSSI-E). The R-LiNK server, the storage system and all other parts of the CEA information system implement this *PSSI*. All communication with the R-LiNK server at CEA use authenticated/encrypted protocols (SFTP, HTTPS). Data are not encrypted on disk as this would be technically challenging, however access to the server room is restricted to authorized IT personnel. The storage system performs regular backups and duplicates data in two server rooms in distinct buildings of the CEA/Saclay centre.

---

5. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

---

6. Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

We collect and process data according to the MR-001 methodology published by CNIL, the French data protection and privacy commissioner.

CEA follow their own IT security policy (*PSSI*) which is directly derived from the IT security policy of the French state (PSSI-E).