RaDiCo, the French National Program on Rare Disease Cohorts

Serge Amselem
  INSERM

Sonia Gueguen
  INSERM

Jérôme Weinbach
  Ministere de la Santé et de l'Action Sociale: Ministere de la Sante et de l'Action Sociale

Annick Clement
  Assistance Publique Hopitaux de Paris: Assistance Publique - Hopitaux de Paris

Paul Landais (✉ paul.landais@umontpellier.fr)
  Universite de Montpellier   https://orcid.org/0000-0002-4166-8432

Research

Abstract

Background: Rare diseases (RDs) affect nearly 3 million people in France and at least 26-30 million people in Europe. These diseases, which represent a major medical concern, are mainly of genetic origin, often chronic, progressive, degenerative, life threatening and disabling, accounting for more than one third of all deaths occurring during infancy. In this context, there are needs for coordinated information on RDs at national /international levels, based on high quality, interoperable and sharable data. The main objective of the RaDiCo (Rare Disease Cohorts) program, coordinated by Inserm, was the development of RD e-cohorts via a national platform. The cohort projects were selected through a national call in 2014. The e-cohorts are supported by an interoperable platform, equivalent to an infrastructure, constructed on the "cloud computing" principle and in compliance with the European General Data Protection Regulation. It is dedicated to allow a continuous monitoring of data quality and consistency, in line with the French Health Data Hub.

Results: Depending on cohorts, the objectives are to describe the natural history of the studied RD(s), establish phenotype-genotype correlations, decipher their pathophysiology, assess their societal and medico-economic impact, and/or identify patients eligible for new therapeutic approaches. Inclusion of prevalent and incident cases started at the end of 2016. As of April 2021, 5558 patients have been included within 13 RD e-cohorts covering 67 diseases integrated in 10 European Reference Networks and contributing to the European Joint Program on RDs. Several original results have been obtained in relation with the secondary objectives of the RaDiCo cohorts. They deal with discovery of new disease genes, assessment of treatment management, deciphering the underlying pathophysiological mechanisms, diagnostic approaches, genotype-phenotype relationships, development and validation of questionnaires relative to disease burden, or methodological aspects.

Conclusion: RaDiCo currently hosts 13 RD e-cohorts on a sharable and interoperable platform constructed on the "cloud computing" principle. New RD e-cohorts at the European and international levels are targeted.

Background

Rare Diseases (RDs) represent a major health care issue. A disease is called rare in Europe when it affects fewer than 5 in 10,000 persons. Depending on the definition used, it is estimated that there are 7,000 to 8,000 RDs. Taken as a whole, RDs affect at least 26-30 million people in Europe. These diseases involve children in about 75% of cases. They are often chronic, progressive, degenerative, life threatening and disabling, accounting for more than one third of all deaths occurring during infancy. A large number of these diseases lead to a significant decrease of life expectancy and most of them have a significant impact on patients' quality of life and health care systems. This is a major medical concern since, for most of these diseases, there is no available cure. RDs are mainly of genetic origin. At the date of writing this article, 4,425 genes underlying 6,868 disease phenotypes have been discovered (https://www.omim.org/statistics/geneMap), but the genes underlying, and additional Mendelian

conditions have yet to be recognized. Thus, there are needs for high quality, interoperable and sustainable creation and monitoring of (inter)national cohorts of patients with rare diseases.

Coordinated care and research on RDs appear critical for patients and families, in order to better describe the natural history of the disease, to improve the diagnostic procedures, to decipher the underlying pathophysiological mechanisms and to better stratify patients for targeted clinical trials and treatments through personalized medical approaches. It is also expected that research on RDs will increase our understanding of the pathophysiology of common chronic diseases, as RDs often represent a "model of dysfunction" severely affecting a limited number of biological pathways.

Identifying the locks is essential. Yet, RD clinical management and research in France and Europe have been hampered by a lack of resources at several levels: few scientists work on only one given disease; few patients per disease and patients are scattered over large geographic areas, causing difficulties in gathering data on their disease; existing databases as well as biological collections, when existing, are usually local, small, incomplete, not always quality-controlled, of heterogeneous formats and contents, and are rarely accessible or standardized for allowing interoperability; phenomes are often complex and partially described along with time, with insufficient interdisciplinary cooperation.

The need to promote networks of expertise for RDs in order to improve both RD clinical care and research has been considered a priority in France since 1995. Indeed, RDs occupy an important place in public health in France: Orphanet (an informational RD website and a directory of expert services, launched in 1997) [1] was created by the Health Ministry; the Alliance Maladies Rares (French federation of patients' organizations) was launched in February 2000 [2]; and the GIS (*Groupement d'Intérêt Scientifique*) *Maladies Rares* was created to fund research in the field of RDs. The creation of the first French National Rare Disease Plan (PNMR, "*Plan National Maladies Rares*") (PNMR1, 2005-2008) [3] allowed access to high quality care, and treatment was facilitated by the creation, at the national level, of RD Reference Centers (RDRCs) and Competence Centers. The PNMR2 (2011-2016) consolidated previous achievements, aiming at reinforcing national and international cooperation [4]. Twenty University genetic laboratories were equipped with the Next Generation Sequencing technology for clinical use. In 2014, the RDRCs and RD Competence Centers were grouped into 23 thematic RD Healthcare Networks (RDHNs or "*Filières de Santé Maladies Rares*"). This move anticipated the 23 RD European Reference Networks (ERNs) that were launched at the end of 2016. The RDHNs coordinate diagnosis, provision of health, social care and training; they collect healthcare data, develop research programs and write the National Protocols for Diagnosis and Care ("PNDS" standing for "*Protocoles Nationaux de Diagnostic et de Soins*").

One main objective of the PNMR3 (2018-2022) [5], in line with the International Rare Diseases Research Consortium (IRDIRC) [6] vision, was to provide an accurate diagnosis within a year of the first specialty medical consultation. The PNMR3, which identified 109 RD Coordinating Centers (RDCCs) associated to 386 constitutive centers and 1840 competence centers, reinforced the links with European research

initiatives on RDs. It also aims at strengthening the role of RaDiCo in integrating research data for RDs (Action 11.4).

A need for launching RD cohorts. In this fast-mutating context, RD professionals highlighted the critical need to implement nation-wide, multidisciplinary, high-quality cohort studies in order to address key scientific and medico-economic questions. In this context, the request was high to get access to appropriate resources, methods and tools for collecting RD data at the (inter)national level. Given the specificities of RDs - limited number of patients per country, scarcity of relevant knowledge and expertise, and fragmentation of research – they have been considered as a distinctive domain of very high national and European added value. A supporting research program was therefore required to provide essential information on disease history and characteristics, and to foster the identification of underlying molecular mechanisms, genotype/phenotype correlations. New knowledge would ultimately lead to better targeted care and treatments. The structuring boost given by the PNMR2 was considered as an opportunity to make collective efforts for building RD cohorts and therefore to propose the RaDiCo (Rare Disease Cohort) project to the national call on cohorts of the first Investments for the Future Program (*Programme d'Investissements d'Avenir* – PIA) launched by the French Ministry of Research in 2010 [7].

The objective of the RaDiCo project was twofold. On the one hand, the scientific objective was to set up several RD e-cohorts with the following aims according to the idiosyncrasy of each cohort: Describe the natural history of the targeted RDs; Establish genotype-phenotype correlations; Decipher the underlying pathophysiological mechanisms; Identify new therapeutic avenues; Estimate their societal and medico-economic impact; Identify patients eligible for new therapeutic approaches; Define a methodological strategy of analysis for cohorts including both prevalent and incident cases (with respect to modeling and bias analyses).

On the other hand, to reach these scientific objectives, another goal was to build a national operational platform, equivalent to an Infrastructure as a Service (IaaS), for implementing a potentially unlimited number of e-cohorts consisting of prevalent and incident cases.

Such cohort projects had to be closely articulated with the above-mentioned established networks on RDs. The PNMR2 also fostered the development of the National RD Data Bank (BNDMR standing for "*Banque Nationale de Données Maladies Rares*", Figure 1). The BNDMR was built as part of CEMARA [8] ("*Centre Maladies Rares*"), which started in 2007 at Necker Enfants-Malades Hospital (AP-HP). The BNDMR, which is dedicated to public health issues, aims at collecting general epidemiological and public health data on all patients with a RD in France, on the basis of a common RD minimum data set [9] and a unique identifier [10] for each RD patient.

The RD cohorts also had to integrate, whenever appropriate, non-French RD expert centers and patients, in order to overcome the low number of patients in one single country with respect to the sample sizes needed for proper statistical power, thereby anticipating the emergence of the future RD European Research Networks (ERNs).

Results

Thirty-three letters of intent were received after the publication of the RaDiCo call for RD cohorts (see Material and Methods). This call led to select 16 national and/or European RD cohort projects on July 15th, 2014. Among these, 3 have been discontinued, secondary to decisions of the Scientific and Plenary committees of the RaDiCo program (see Material and Methods): after demand from the principal investigators (PIs) for two of them since the cohorts had not started, and another in 2019 since it could not start in due time. The groups of RDs targeted by the 13 current cohorts appear in Table 1.

The general framework of the RaDiCo cohorts is presented below (Table 2), with or without associated biocollections at each site, together with the planned inclusion and follow-up period. The start of the inclusions and the inclusion targets are mentioned. All cohorts are multicentric, mainly national but also European (SEDVasc, ECYSCO) and international (GenIDA). For all the cohorts, a total of 11,650 included patients are expected at the end of the inclusion period (July 2027). As of April 2021, 5558 patients had been included into 13 RD e-cohorts, covering 67 diseases from ~300 affiliated expert centers (Figure 2). Each implementation step of the e-cohorts is presented in Figure 3.

These cohorts were supported by a platform enabling information flow and communication, which has been set up in the framework of the RaDiCo program. This platform, which fulfills the necessary requirements of an Information System (IS), has been assessed by an independent external auditor. Moreover, it contributed to the design of the IS from the Inserm "France Cohorts" program designed to support not only RD e-cohorts, but also cohorts of patients with common multifactorial disorders, as well as population-based epidemiological cohorts.

Primary objectives of the RD cohorts are not yet achieved; however, several results have already been obtained in relation with the secondary objectives of several RaDiCo cohorts. They deal with discovery of new disease genes, assessment of treatment management, deciphering the pathophysiology and diagnostic approaches, genotype-phenotype relationships, development and validation of questionnaires relative to the diseases burden, or methodological aspects (Table 3).

Discovery of new disease genes - Four genes, i.e. *TTC12* [11], *GAS2L2* [12], *DNAH9* [13] and *DNAJB13* [14], whose mutations are responsible for primary ciliary dyskinesia (PCD) have been identified through molecular and cellular studies performed in the framework of the RaDiCo-DCP cohort. This cohort has been built on the deep phenotyping of the patients, which includes the ultrastructural defects of the microtubule-based structure of motile cilia and sperm flagella (i.e. the ciliary and flagellar axonemes). These organelles contain several dynein arms, each of them consisting of multiprotein complexes that carry an ATPase activity required for ciliary/flagellar motility. Specific phenotypes have been associated with mutations in these new genes that code for different classes of proteins. TTC12 is believed to be a co-chaperone involved in the cytoplasmic pre-assembly of dynein arms [11]; the lack of GAS2L2 causes PCD by impairing cilia orientation and mucociliary clearance [12], whereas DNAH9 encodes one of the axonemal dynein chains [13]. As for *DNAJB13*, it encodes an HSP40 family member involved in the

proper building of the ciliary and flagellar axoneme [14]. In addition, besides the identification of those new molecular causes of PCD, one of the key results obtained through functional studies performed on both patients' primary airway epithelial cells (AECs) and CRISPR-Cas9-edited human primary AECs is the existence of distinct dynein assembly mechanisms in human motile cilia versus flagella [11]. As for the developmental eye defects reported in patients from the RaDiCo-AC-OEIL cohort, *de novo* missense variants have been identified in *FBXW11*, a gene that encodes an F-box protein involved in ubiquitination and proteosomal degradation [15].

Assessment of treatment management - From a therapeutic viewpoint, as shown in the RaDiCo-SEDVasc cohort, which is dedicated to patients with a rare genetic connective tissue disorder called vascular Ehlers-Danlos syndrome, the assessment of treatment management has revealed the impact of different therapies on morbidity and mortality. Indeed, in this disease condition due to mutations in *COL3A*, it has been shown that patients treated with celiprolol —a beta blocker— had a better survival than those not treated with celiprolol and that the observed reduction in mortality was dose-dependent [16].

Pathophysiology and diagnostic approaches - Several aspects were explored: Accuracy of clinical diagnostic criteria for patients with vascular Ehlers-Danlos syndrome in a tertiary reference center for the RaDiCo-SEDVasc cohort [17]; In the RaDiCo-PID cohort, functional assessment and phenotypic heterogeneity of *SFTPA1* and *SFTPA2* mutations in interstitial lung diseases and lung cancer [18]; Health-related quality of life in infants and children with interstitial lung disease [19]; Pulmonary fibrosis in children [20]; Diagnosis approaches of chronic interstitial lung diseases in children: [21,22,23]; or genetic causes and clinical management of pediatric interstitial lung diseases [24].

Genotype-phenotype relationships – For the RaDiCo-DCP cohort, infertility in an adult cohort with primary ciliary dyskinesia [25]; Identification of a major Mediterranean allele in primary ciliary dyskinesia gene [26]. For the RaDiCo-EURBIO Alport cohort, a unified classification of genetic disorders of collagen IV α345 in Alport syndrome [27]. For the RaDiCo AC-OEIL, review of the genetics of non-syndromic anophthalmia/microphthalmia [28].

Development and validation of burden questionnaires - Two questionnaires were validated in order to support the RaDiCo–FARD cohort: the burden of albinism and development and validation of a burden assessment tool [29], and the burden of adult neurofibromatosis 1, development and validation of a burden assessment tool [30].

Methodological aspects - Several tools for harmonizing our approaches in our own country and in Europe have been developed. We defined a way to federating RD patients identities [10]; to develop Cerberus, an access control scheme for enforcing least privilege in patient cohort study platforms for the RaDiCo-GENIDA cohort [31]; to propose an overview of the current situation and experiences of the national RD registries in Europe: [32]; with other European colleagues to share recommendations for improving the

quality of RD registries [33,34] the two latter papers were used for the design of our e-Health approach. Moreover, RaDiCo is involved in the French PNMR3 and contributes to the EJP-RD program [35].

Discussion

The RaDiCo program is intended to promote, through the support of RD e-cohorts' projects, the collection of phenotypic data for epidemiological and clinical research purposes in connection with basic and translational research. To this end, we implemented an operational team (cf. Material and Methods section) with the mission to establish a centralized platform of expert services and tools to ensure installation and follow-up of the RD e-cohorts, via consulting services, for instance legal & regulatory services, or clinical research quality tools, development and provision of innovative information technology tools (electronic case report forms (eCRF), interoperability solutions, or access to e-health tools).

The preparatory period of the cohorts was slowed down by the uncontrollable delays required to obtain all the ethical and regulatory authorizations, as well as the evolving regulations both at the EU level, with the new European regulation on personal data safety and security (GDPR) [36], and at the national level, with modifications of the law on research involving human subjects (French «Jardé law»)[37]. Moreover, a consortium agreement had to be signed between the partners for formalizing legal links between them, setting the modes of governance of the project on scientific, strategic and operational plans, and outlining the scientific valorization in terms of intellectual property, publications and citations of the project. A set of 19 Key Performance Indicators (KPIs) was implemented to enforce the cohorts' follow-up. Once initiated, the 13 ongoing RD cohorts progressed appropriately and the first results related to the secondary objectives of several cohorts have been published in international peer-reviewed journals.

The costs and management of informatics providers and Clinical Research Organization (CRO) are important. Public-Private Partnerships helped to build our sustainability plan. Legal contracts were set up for consortium and collaboration agreements, enabling data collection and sharing.

At the cohort level, we industrialized our production processes, as well as tooling methods such as standardized operating procedures or documents shared online. Recurring difficulties are the shortage of manpower for data entry, the difficulties in identifying cases in the different hospital information systems, or the legal barriers to collect data from deceased children. We implemented solutions through identifying key data and prioritization of the fields related to the main objectives to be completed by the clinical research technicians with support from the residents for filling the medical information. Clinical research technicians have been recruited by the RDHNs and specific resources have been obtained through targeted research projects supported by Public-Private Partnerships or applications to specific grants.

We explored the impact of RaDiCo for the medical community. Based on current information provided by all the cohorts' participating teams, the program should contribute to significant improvements in patients' care and outcome. Indeed, the design of the cohorts is made in such a way that it allows

structured collection of disease symptoms at various stages of the pathological processes. Expected benefits include better knowledge of the natural history of all the investigated RDs, recognition of relevant comorbidities, novel proposals for healthcare management, and, ultimately, better quality of life for the patients and their families. In line with the objectives specifically defined for each cohort, other expectations are: identification of relevant disease biomarkers for diagnosis, disease severity and exacerbations; patient selection for clinical and therapeutic trials; and production of novel quality of life questionnaires. The development of electronic health tools is also among the secondary objectives of several cohorts. These tools are designed to help patients to self-report their symptoms as well as their medical management, and ultimately to assist the health care providers to make appropriate changes to medication use.

Altogether, the cohort studies should help progressing towards a more personalized medicine, particularly for appropriate investigations and therapeutic strategies. They should contribute to decrease the burden of these chronic diseases and to improve their socio-economic impact.

The RD cohorts are multidisciplinary and, for most of them, include molecular diagnostics and research laboratories. Organization and standardization of biological sample collections (biobanks) are part of the program. Strong interactions and collaborations between clinicians and basic scientists are developed to identify genetic and environmental determinants in well-defined groups of patients, to establish genotype /phenotype correlations, and to progress in the understanding of the underlying molecular and cellular mechanisms.

Rare diseases carry high morbidity and mortality; they represent a significant burden for the health care systems. Evaluation of the economic impact of RDs is therefore critical. This implicates a dedicated focus on the cost-effectiveness analyses, which evaluate both the costs and results of the health care systems and organizations applied in RDs. The final goal is to appropriately allocate the health care resources to the RD patients all over the country. This implies that RD are adequately traceable in the national health information systems. In line with these needs, the Ministry of Health has set up a national bank for rare disease data (BNDMR). All the reference centers and their network of teams in France have the obligation to implement this data bank. Standardized and detailed patient information collected through the cohort programs support the production of economic indicators. Moreover, access to the French Health Data Hub [38] that is dedicated to cross-referencing the health databases and facilitating their use by research teams with full respect of user privacy of the health system, will open access to patients' health expenditures from three main databases: the National Health Insurance Fund ("*Caisse Nationale d'Assurance Maladie*") [39], hospital stays from the French hospital discharge database (PMSI standing for "*Programme de Médicalisation du Système d'Information*") [40] and to the causes of death from the Inserm Centre for Epidemiology on the Medical Causes of Death, in charge of the national statistics of medical causes of death ("CepiDC") [41]. Economic indicators aim at monitoring the development and implementation of national strategies for RDs, which imply homogeneity and standardization of patient management by the medical structures dedicated to RDs.

All the teams of the RaDiCo cohorts are members of the RDCCs in France. Among their tasks is the production of the RDs National Protocols for Diagnosis and Care (PNDS) [42]. Therefore, through their organizations and their multidisciplinary teams, all the cohorts actively participate in the production of PNDS.

The ERNs are networks of RD centers of expertise, healthcare providers and laboratories that are organized across borders. Most of the medical teams associated with the RaDiCo program are involved in the activities of the ERN in their field. They participate in the implementation of European data registries, in the production of clinical practice recommendations, and in the setting of epidemiological and economic indicators.

Conclusion

The RaDiCo program promotes the collection of RD phenotypic data of different types for epidemiological and clinical research purposes in connection with basic and translational research. The RaDiCo platform offers the cohorts an information system based on the cloud principle (Information as a Service) together with a common core of services and specific procedures for some cohorts to ensure installation and follow-up of the RD e-cohorts. RaDiCo IS can drive a virtually unlimited number of RD e-cohorts, a major strength compared to other IS that usually deal with only one cohort. Currently, 13 RD e-cohorts are implemented and other national, European and international cohorts are targeted.

Material And Methods

RaDiCo's national call for implementing RD cohorts

In 2014, RaDiCo launched a national call for RD Cohort proposals organized as a two-stage procedure (letter of intention / full application). Guidelines and templates for candidate cohorts were designed and made accessible on line to applicants. The full dossiers were evaluated by independent international experts (at least 3 per project), with pre-established evaluation criteria.

RaDiCo's governance

An Institutional Committee (IC) is composed of the founding institutions including the following: Inserm (coordinating institution), 6 Universities: Paris Descartes and Paris Diderot now merged in the University of Paris, as well as Sorbonne, Paris-Est Créteil, Aix-Marseille and Montpellier universities; and the « Entreprises du Médicament » (LEEM), together with the Alliance for Research and Innovation of the Industries of Health (ARIIS). These bodies are linked to the RaDiCo program by an agreement established with the National Research Agency (ANR).

Among other missions, the IC regularly follows the implementation of the project including the budget implementation. It validates, on proposal of the Scientific Committee, the annual program of activities. It decides the creation, composition, missions and the functioning of all *ad-hoc* governing bodies.

The RaDiCo Scientific Committee (SC) for the whole Program is composed of 18 members with complementary expertise associating coordinators of RDRC, RDHN, epidemiologists, biostatisticians, experts in Information Systems, molecular geneticists, directors of Inserm Research Units in the field of RDs, and representatives of Inserm Thematic Institutes. Its role is to support the Executive Committee concerning strategic issues and scientific governance of the program, including management of the call for projects and specific follow-up of cohorts; and to contribute to the national call for RD cohort projects launched by RaDiCo in 2014.

An Executive Committee (EC) is shared by all the cohorts. It is composed of 4 members with complementary expertise (SG, AC, PL and SA). Its role is to ensure the implementation, development and monitoring of the entire program. Its main missions are operational management and deployment planning of the sustainability and internationalization of the RaDiCo platform, the links with each RD cohort, and a factorization of the shares. The EC is responsible for the scientific management of the program.

The management of the scientific program is implemented at different levels: by the Executive Committee, the Scientific Committee, the Inserm Thematic Institutes of "Public health", "Technology for Health", "Genetics, Genomics and Bioinformatics", as well as by the 3 scientific governance bodies of each RD cohort. Indeed, each cohort has its own governance, on a scheme comparable to that of the whole program, with a Plenary Committee, a Scientific Committee and an Executive Committee.

The RaDiCo team is directed by a Scientific & Operational Director (SOD), in close collaboration with the Executive Committee. The SOD leads 24 fully dedicated employees. The team acts within 3 different sub-units: Clinical research, Biometry, and Information System & e-Health.

The RaDiCo platform

The RaDiCo platform uses exchange format and data security in compliance with the European directive on the General Data Protection (GDPR). The RaDiCo work plan for each cohort is structured to achieve four levels of interoperability: technological, semantic, syntactic, and institutional. This allows networking and optimizing the use of existing RD patient cohorts at the EU level, while allowing integration of new types of data and technologies. The data of each RaDiCo cohort and associated services have to be compliant with the FAIR principles (Findable, Accessible, Interoperable and Reusable) by both people and computers. An important step in the FAIR data approach is to publish existing and new datasets from RaDiCo cohorts in a semantically interoperable format that can be understood by computer systems.

Declarations

Availability of data and materials: The data of the cohorts are not available presently since they are still running and primary objectives have not been analyzed yet. The datasets analyzed will be available from the PIs of the corresponding cohorts on reasonable request.

Competing interests: The authors declare that they have no competing interests.

Ethics approval and consent to participate are given for each of the13 RaDiCo cohorts :

* Ethics approval : an agreement has been requested for each RaDiCo cohort. Concerning the consent, the need for approval was waived.

* The ethics committee that approved each study is given in the table below.

* The committee's reference number was delivered for each cohort as presented below.

* Moreover, we obtained the agreement of the Commission Nationale Informatique et Libertés (CNIL) and from the Comité consultatif pour le traitement des données en Santé. The agreement was obtained for each cohort as stated below.

Consent for publication : not applicable

Cohorts acronyms	Ethical Committee agreement per cohort (agreement number)	Date of agreement	CCTIRS	Date of agreement	CNIL
AC-ŒIL	CEEI (IRB00003888)	08/09/2015	CCTIRS	07/01/2016	CNIL
AcoStill	CPP Ile de France I (IRB IORG0008367)	02/04/2016	CCTIRS	12/05/2016	MR001
COLPAC	CPP Ile de France I (IRB IORG0008367)	12/07/2016	CCTIRS	19/05/2016	MR001
DCP	CEEI (IRB00003888)	06/10/2015	CCTIRS	12/11/2015	CNIL
ECYSCO	CEEI (IRB00003888)	08/09/2015	CCTIRS	12/09/2015	CNIL
EURBIO Alport	CPP Ile de France I (IRB IORG0008367)	31/03/2016	CCTIRS	11/02/2016	CNIL
FARD	CPP Ouest V	04/07/2017	not appl.	not appl.	MR003
GenIDA	CEEI (IRB00003888)	15/11/2016	not appl.	not appl.	CNIL
IDMet	CPP Ile de France I (IRB IORG0008367)	31/03/2016	CCTIRS	11/02/2016	CNIL
MPS	CPP Ile de France I (IRB IORG0008367)	06/09/2016	CCTIRS	12/09/2016	MR001
PID	CEEI (IRB00003888)	06/09/2015	CCTIRS	17/03/2016	CNIL
PP	CPP Ile de France I (IRB IORG0008367)	12/09/2017	not appl.	not appl.	MR003
SEDVasc	CEEI (IRB00003888)	08/09/2015	CCTIRS	12/11/2015	CNIL

CEEI: Comité d'évaluation éthique de l'Inserm; CPP: Comité de Protection des Personnes; CCTIRS: Comité Consultatif pour le Traitement de l'Information en Recherche en Santé;

CNIL: Commission Informatique et Libertés;
 MR001: cf https://www.cnil.fr/sites/default/files/atoms/files/mr-001.pdf ;
 MR003: cf https://www.cnil.fr/sites/default/files/atoms/files/mr-003.pdf;

not appl. : not applicable according to the new « Loi Jardé » in force as of Nov 18, 2016

References

1. https://www.orpha.net/consor/cgi-bin/index.php?lng=EN (accessed on May 30, 2021).

2. http://alliant.cluster028.hosting.ovh.net/ (accessed on May 30, 2021).

3. https://solidarites-sante.gouv.fr/IMG/pdf/plan_national_maladies_rares_2005-2008.pdf (accessed on May 30, 2021).

4. https://solidarites-sante.gouv.fr/IMG/pdf/pnmr2_version_anglaise.pdf accessed on May 30, 2021).

5. https://solidarites-sante.gouv.fr/IMG/pdf/pnmr3_-_en.pdf (accessed on May 30, 2021).

6. https://irdirc.org/ (accessed on May 30, 2021)

7. https://www.enseignementsup-recherche.gouv.fr/cid51358/cohortes.html (accessed on May 30, 2021).

8. Landais P, Messiaen C, Rath A, Le Mignot L, Dufour E, et al for the CEMARA task force. CEMARA an information system for rare diseases. Stud Health Technol Inform. 2010;160 (Pt 1):481-5.

9. Choquet R, Maaroufi M, de Carrara A, Messiaen C, Luigi E, Landais P. A methodology for a minimum data set for rare diseases to support national centers of excellence for healthcare and research. J Am Med Inform Assoc. 2015 Jan;22(1):76-85. doi: 10.1136/amiajnl-2014-002794.

10. Maaroufi M, Landais P, Messiaen C, Jaulent MC, Choquet R. Federating patients identities: the case of rare diseases. Orphanet J Rare Dis. 2018 Nov 12;13(1):199. doi: 10.1186/s13023-018-0948-6. PMID: 30419918.

11. Thomas L, Bouhouche K, Whitfield M, Thouvenin G, Coste A, Louis B, Szymanski C, Bequignon E, Papon JF, Castelli M, Lemullois M, Dhalluin X, Drouin-Garraud V, Montantin G, Tissier S, Duquesnoy P, Copin B, Dastot F, Couvet S, Barbotin AL, Faucon C, Honore I, Maitre B, Beydon N, Tamalet A, Rives N, Koll F, Escudier E, Tassin AM, Touré A, Mitchell V, Amselem S, Legendre M. TTC12 Loss-of-Function Mutations Cause Primary Ciliary Dyskinesia and Unveil Distinct Dynein Assembly Mechanisms in Motile Cilia Versus Flagella. Am J Hum Genet. 2020 Feb 6;106(2):153-169. doi: 10.1016/j.ajhg.2019.12.010. Epub 2020 Jan 23.

12. Bustamante-Marin XM, Yin WN, Sears PR, Werner ME, Brotslaw EJ, Mitchell BJ, Jania CM, Zeman KL, Rogers TD, Herring LE, Refabert L, Thomas L, Amselem S, Escudier E, Legendre M, Grubb BR, Knowles MR, Zariwala MA, Ostrowski LE. Lack of GAS2L2 Causes PCD by Impairing Cilia Orientation and Mucociliary Clearance. Am J Hum Genet. 2019 Feb 7;104(2):229-245.

13. Fassad MR, Shoemark A, Legendre M, Hirst RA, Koll F, le Borgne P, Louis B, Daudvohra F, Patel MP, Thomas L, Dixon M, Burgoyne T, Hayes J, Nicholson AG, Cullup T, Jenkins L, Carr SB, Aurora P, Lemullois M, Aubusson-Fleury A, Papon JF, O'Callaghan C, Amselem S, Hogg C, Escudier E, Tassin AM, Mitchison HM. Mutations in Outer Dynein Arm Heavy Chain DNAH9 Cause Motile Cilia Defects and Situs Inversus. Am J Hum Genet. 2018 Dec 6;103(6):984-994. doi: 10.1016/j.ajhg.2018.10.016. Epub 2018 Nov 21. PMID: 30471717

14. El Khouri E, Thomas L, Jeanson L, Bequignon E, Vallette B, Duquesnoy P, Montantin G, Copin B, Dastot-Le Moal F, Blanchon S, Papon JF, Lorès P, Yuan L, Collot N, Tissier S, Faucon C, Gacon G, Patrat C, Wolf JP, Dulioust E, Crestani B, Escudier E, Coste A, Legendre M, Touré A, Amselem S. Mutations in DNAJB13, Encoding an HSP40 Family Member, Cause Primary Ciliary Dyskinesia and Male Infertility. Am J Hum Genet. 2016 Aug 4;99(2):489-500. doi: 10.1016/j.ajhg.2016.06.022

15. Holt RJ, Young RM, Crespo B, Ceroni F, Curry CJ, Bellacchio E, Bax DA, Ciolfi A, Simon M, Fagerberg CR, van Binsbergen E, De Luca A, Memo L, Dobyns WB, Mohammed AA, Clokie SJH, Zazo Seco C, Jiang YH, Sørensen KP, Andersen H, Sullivan J, Powis Z, Chassevent A, Smith-Hicks C, Petrovski S,

Antoniadi T, Shashi V, Gelb BD, Wilson SW, Gerrelli D, Tartaglia M, Chassaing N, Calvas P, Ragge NK. De Novo Missense Variants in FBXW11 Cause Diverse Developmental Phenotypes Including Brain, Eye, and Digit Anomalies. Am J Hum Genet. 2019 Sep 5;105(3):640-657. doi: 10.1016/j.ajhg.2019.07.005.

16. Frank M, Adham S, Seigle S, Legrand A, Mirault T, Henneton P, Albuisson J, Denarié N, Mazzella JM, Mousseaux E, Messas E, Boutouyrie P, Jeunemaitre X. Vascular Ehlers-Danlos Syndrome: Long-Term Observational Study. J Am Coll Cardiol. 2019 Apr 23;73(15):1948-1957. doi: 10.1016/j.jacc.2019.01.058.

17. Henneton P, Albuisson J, Adham S, Legrand A, Mazzella JM, Jeunemaitre X, Frank M. Accuracy of Clinical Diagnostic Criteria for Patients with Vascular Ehlers-Danlos Syndrome in a Tertiary Referral Centre. Circ Genom Precis Med. 2019 Mar;12(3): e001996.

18. Legendre M, Butt A, Borie R, et al. Functional assessment and phenotypic heterogeneity of SFTPA1 and SFTPA2 mutations in interstitial lung diseases and lung cancer. Eur Respir J. 2020 Aug 27:2002806. doi: 10.1183/13993003.02806-2020.

19. Lauby C, Boelle PY, Abou Taam R, Bessaci K, Brouard J, Dalphin ML, Delacourt C, Delestrain C, Deschildre A, Dubus JC, Fayon M, Giovannini-Chami L, Houdouin V, Houzel A, Marguet C, Pin I, Reix P, Renoux MC, Schweitzer C, Tatopoulos A, Thumerelle C, Troussier F, Wanin S, Weiss L, Clement A, Epaud R, Nathan N. Health-related quality of life in infants and children with interstitial lung disease. Pediatr Pulmonology. 2019 Jun;54(6):828-836.

20. Nathan N, Sileo C, Thouvenin G, Berdah L, Delestrain C, Manali E, et al. Pulmonary fibrosis in children. J Clin Med. 2019 Aug 26;8(9):1312.

21. Nathan N, Berdah L, Borensztajn K, Clement A. Chronic interstitial lung diseases in children: diagnosis approaches. Expert Rev Respir Med. 2018 Oct 22:1-10. PMID: 30345849

22. Alimi A, Taytard J, Abou Taam R, Houdouin V, Forgeron A, Lubrano Lavadera M, et al. Pulmonary hemosiderosis in children with Down syndrome: a national experience. Orphanet J Rare Dis. 2018 Apr 20;13(1):60.

23. Nathan N, Sileo C, Calender A, Pacheco Y, Rosental PA, Cavalin C, Macchi O, Valeyre D, Clement A; French Sarcoidosis Group (GSF); Silicosis Research Group. Paediatric sarcoidosis. Paediatr Respir Rev. 2019 Feb; 29:53-59.

24. Nathan N, Borensztajn K, Clement A. Genetic causes and clinical management of pediatric interstitial lung diseases. Curr Opin Pulm Med. 2018 May;24(3):253-259.

25. Vanaken GJ, Bassinet L, Boon M, Mani R, Honore I, Papon JF, Cuppens H, Jaspers M, Laurent N, Coste A, Escudier E, Amselem S, Maitre B, Legendre M, Christin-Maitre S. Infertility in an adult cohort with primary ciliary dyskinesia: phenotype- gene association. Eur Respir J. 2017 Nov 9;50(5).

26. Mani R, Belkacem S, Soua Z, Chantot S, Montantin G, Tissier S, et al. Primary ciliary dyskinesia gene contribution in Tunisia: identification of a major Mediterranean allele. Hum Mutat. 2020 Jan;41(1):115-121.

27. Kashtan CE, Ding J, Garosi G, Heidet L, Massella L, Nakanishi K, Nozu K, Renieri A, Rheault M, Wang F, Gross O. Alport syndrome: a unified classification of genetic disorders of collagen IV α345: a position paper of the Alport Syndrome Classification Working Group. Kidney Int. 2018, 93:1045-1051.

28. Plaisancié J, Ceroni F, Holt R, Zazo Seco C, Calvas P, Chassaing N, Ragge NK. Genetics of anophthalmia and microphthalmia. Part 1: Non-syndromic anophthalmia /microphthalmia. Hum Genet. 2019 Sep;138(8-9):799-830.

29. Morice-Picard F, Taïeb C, Marti A, Gliksohn A, Bennani M, Bodemer C, Ezzedine K; Filière Maladies Rares en Dermatologie: FIMARAD. Burden of albinism: development and validation of a burden assessment tool. Orphanet J Rare Dis. 2018 Sep 18;13(1):162.

30. Armand ML, Taieb C, Bourgeois A, Bourlier M, Bennani M, Bodemer C, Wolkenstein P; along with the French national network on rare skin diseases (FIMARAD). Burden of adult neurofibromatosis 1: development and validation of a burden assessment tool. Orphanet J Rare Dis. 2019 May 3;14(1):94. doi: 10.1186/s13023-019-1067-8.

31. P. Parrend, T. Mazzucotelli, F. Colin, P Collet, Mandel J.-L. Cerberus, an Access Control Scheme for Enforcing Least Privilege in Patient Cohort Study Platforms A Comprehensive Access Control Scheme Applied to the GENIDA Project - Study of Genetic Forms of Intellectual Disabilities and Autism Spectrum Disorders. J Med Syst 2018 42: 1: 1.

32. Taruscio D, Vittozzi L, Choquet R, Ketil Heimdal B, Iskrov G, Kodra Y, Landais P, Posada M, Stefanov R, Steinmueller C et al. National Registries of Rare Diseases in Europe: an overview of current situation and experiences. Public Health Genomics. 2015;18(1):20-5.

33. Kodra Y, Weinbach J, Posada-de-la-Paz M, Coi A, Lemonnier SL, van Enckevort D, Roos M, Jacobsen A, Cornet R, Ahmed SF, Bros-Facer V, Popa V, Van Meel M, Renault D, von Gizycki R, Santoro M, Landais P, Torreri P, Carta C, Mascalzoni D, Gainotti S, Lopez E, Ambrosini A, Müller H, Reis R, Bianchi F, Rubinstein YR, Lochmüller H, Taruscio D. Recommendations for Improving the Quality of Rare Disease Registries. Int J Environ Res Public Health. 2018 Aug 3;15(8). pii: E1644.

34. Kodra Y, Posada de la Paz M, Coi A, Santoro M, Bianchi F, Ahmed F, Rubinstein YR, Weinbach J, Taruscio D. Data Quality in Rare Diseases Registries. Adv Exp Med Biol. 2017; 1031:149-164.

35. https://www.ejprarediseases.org/index.php/about/ (accessed on May 30, 2021).

36. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679 (accessed on May 30, 2021).

37. https://solidarites-sante.gouv.fr/systeme-de-sante-et-medico-social/recherche-et-innovation/recherches-impliquant-la-personne-humaine/ (accessed on May 30, 2021).

38. https://www.health-data-hub.fr/ (accessed on May 30, 2021).

39. https://travail-emploi.gouv.fr/emploi/emploi-et-handicap/prevention-et-maintien-dans-l-emploi/cnamts (accessed on May 30, 2021).

40. https://www.atih.sante.fr/mco/presentation (accessed on April 30 2021).

41. https://www.cepidc.inserm.fr/ (accessed on May 30, 2021).

42. https://www.has-sante.fr/jcms/c_1340879/fr/protocoles-nationaux-de-diagnostic-et-de-soins-pnds (accessed on May 30, 2021).

Tables

Table 1. The 13 ongoing cohorts (acronyms in alphabetic order).

Cohorts'acronyms	Diseases	RDHN[1]	Corresponding ERN[2]
AC-ŒIL	Congenital eye defects	SENSGENE AnDDI-Rares	ERN-Eye
AcoStill	Still's disease	FAi2R	ERN-Rita
COLPAC	LPAC syndrome	FILFOIE	RARE-LIVER
DCP	Primary ciliary dyskinesia	RESPIFIL	ERN-Lung
ECYSCO	Cystinosis	ORKiD	ERKnet
EURBIO-Alport	Alport syndrome	ORKiD	ERKnet
FARD	Rare skin diseases burden	FIMARAD	ERN Skin
GenIDA	Intellectual deficiency and Autism spectrum disorders	DéfiScience AnDDi-Rares	ERN ITHACA
IDMet	Imprinting disorders	FIRENDO OSCAR DéfiScience	ENDO-ERN
MPS	Mucopolysaccharidoses	G2M	MetabERN
PID	Idiopathic interstitial pneumonia	RESPIFIL	ERN-Lung
PP	Periodic paralysis	FILNEMUS	ERN euro-NMD
SEDVasc	Vascular Ehlers-Danlos syndrome	FAVA-Multi	VASCERN

[1] RDNH: Rare Disease Healthcare Network; [2] ERN: European Research Network

Table 2. RaDiCo cohorts: general framework

RaDiCo-Cohort Acronyms	Biocollections	Inclusion period (years)	Follow-up period (years)	Start of inclusions	Inclusion targets
AC-ŒIL	Diagnosis	10	10	Jul.-2017	800
AcoStill	Yes	5	5	Dec.-2017	500
COLPAC	Yes	3	5	Nov.-2017	550
DCP	Diagnosis	5	5	May-2017	700
ECYSCO	Diagnosis	2	2	Apr.-2017	400
EURBIO Alport	Yes	2	3	May-2017	1000
FARD	Yes	4	5	Mar.-2018	900
GenIDA	No	5	5	Nov.-2016	1000
IDMet	Yes	5	10	Mar.-2018	2000
MPS	Yes	2	5	Dec.-2017	1000
PID	Diagnosis	5	5	Nov.-2017	2400
PP	No	1	1,5	Jun.-2019	60
SEDVasc	Diagnosis	4	3	Dec.-2016	340
				Expected total inclusions	11650

Table 3: Publications so far associated to the RaDiCo program

Discovery of new disease genes	• *TTC12* loss-of function mutations cause primary ciliary dyskinesia and unveil distinct dynein assembly in motile cilia vs. flagella [11] • Lack of *GAS2L2* causes primary ciliary dyskinesia by impairing cilia orientation and mucociliary clearance [12] • Mutations in outer dynein arm heavy chain *DNAH9* cause motile cilia defects and situs inversus [13] • Mutations in *DNAJB13*, encoding an HSP40 family member, cause primary ciliary dyskinesia and male infertility [14] • *de novo* missense variants in *FBXW11*, a gene that encodes an F-box protein involved in ubiquitination and proteosomal degradation [15]
Assessment of treatment management	• Vascular Ehlers-Danlos syndrome – Long-term observational study [16]
Pathophysiology and diagnostic approaches	• Accuracy of clinical diagnostic criteria for patients with vascular Ehlers-Danlos syndrome in a tertiary referral centre [17] • Functional assessment and phenotypic heterogeneity of *SFTPA1* and *SFTPA2* mutations in interstitial lung diseases and lung cancer [18] • Health-related quality of life in infants and children with interstitial lung disease [19] • Pulmonary fibrosis in children [20] • Chronic interstitial lung diseases in children: diagnosis approaches [21] • Pulmonary hemosiderosis in children with Down syndrome: a national experience [22] • Paediatric sarcoidosis [23] • Genetic causes and clinical management of pediatric interstitial lung diseases [24]
Genotype-phenotype relationships	• Infertility in an adult cohort with primary ciliary dyskinesia: phenotype-gene association [25] • Primary ciliary dyskinesia gene contribution in Tunisia: Identification of a major Mediterranean allele [26] • Alport syndrome: a unified classification of genetic disorders of collagen IV α345 [27] • Genetics of anophthalmia and microphthalmia. Part 1: Non-syndromic anophthalmia/microphthalmia [28]
Development and validation of burden questionnaires	• Burden of albinism: development and validation of a burden assessment tool [29] • Burden of adult neurofibromatosis 1: development and validation of a burden assessment tool [30]

| Methodological aspects	• Federating patients identities: the case of rare diseases [10]
	• Cerberus, an access control scheme for enforcing least privilege in patient cohort study platforms: [31]
	• National registries of rare diseases in Europe: an overview of current situation and experiences [32]
	• Recommendations for improving the quality of rare disease registries [33]
	• Data quality in rare diseases registries [34]

Figures
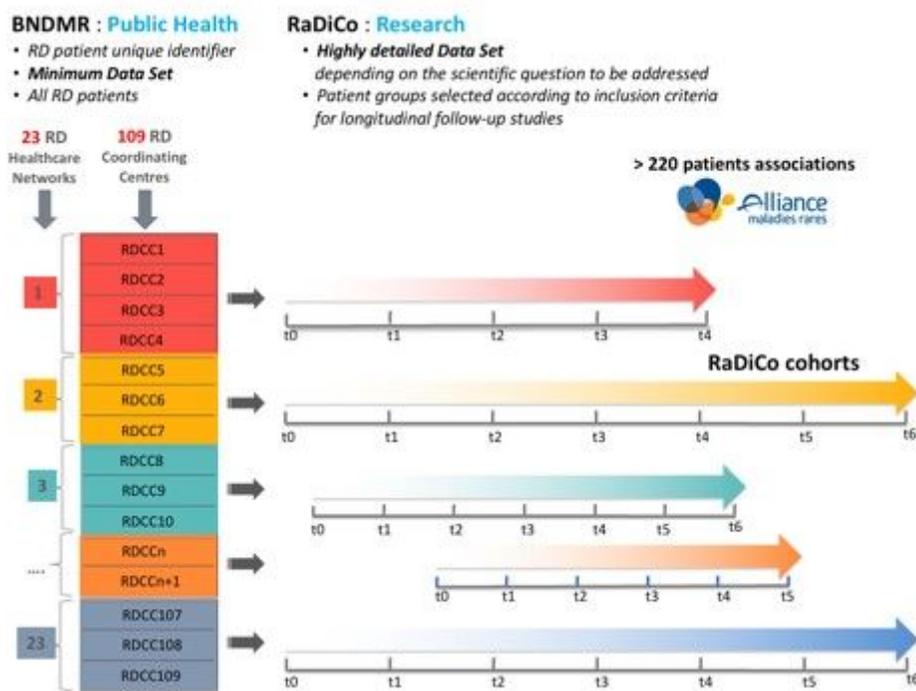
Figure 1

The French RD Healthcare Networks, the RD Coordinating Centres (RDRC) and the two related research programs of Public Health and Research: BNDMR (RD National Databank) for Public Health and RaDiCo
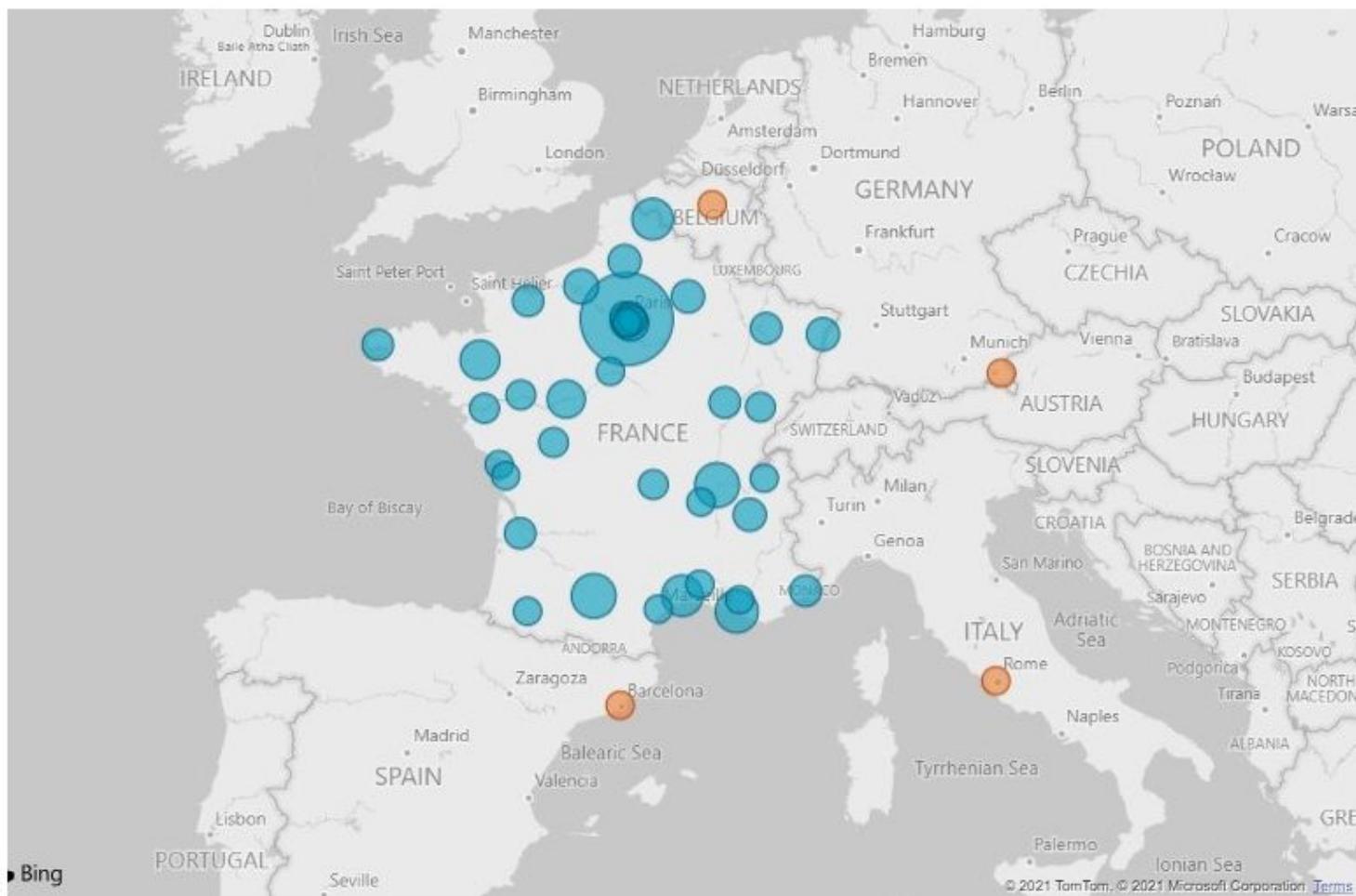
Figure 2

Geographical distribution of RDs centres contributing to the RaDiCo cohorts in France (blue) and other European countries (orange). Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.
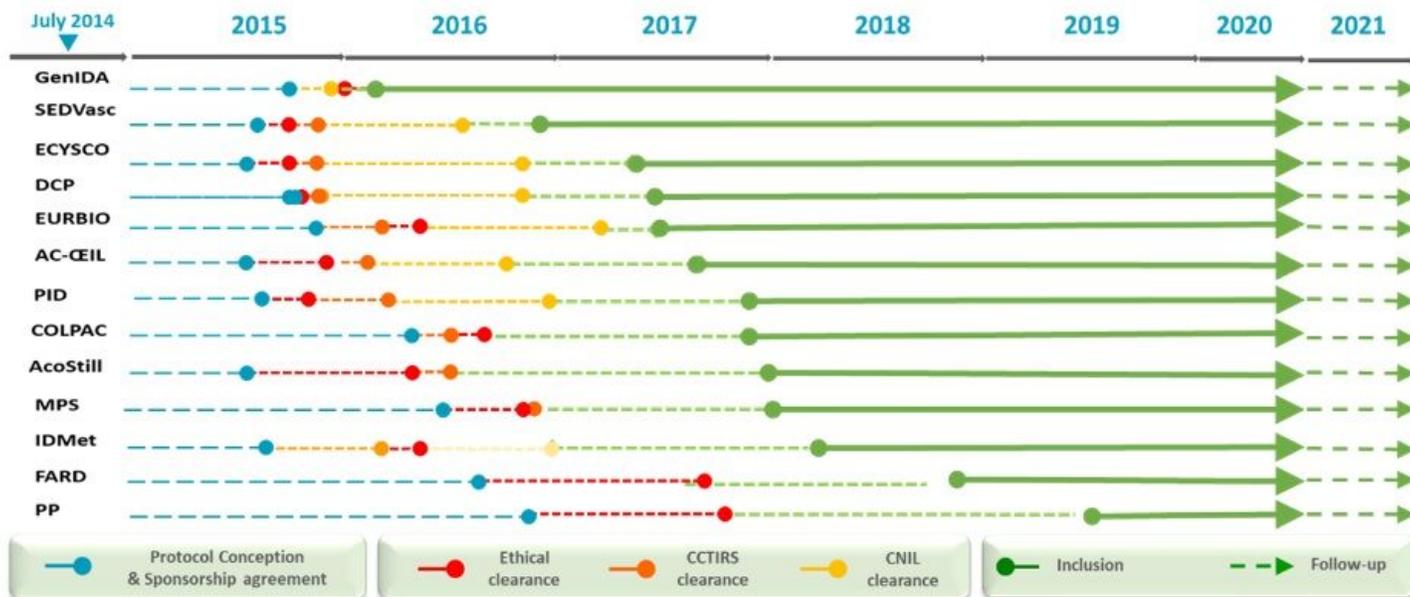
Figure 3

RaDiCo cohorts' preparatory work and 2014-2021 follow-up (CCTIRS, « Comité Consultatif pour le Traitement de l'Information en Recherche en Santé » ; CNIL « Commission Nationale Informatique et Libertés »)