ECRIN
EUROPEAN CLINICAL RESEARCH INFRASTRUCTURE NETWORK

Paediatric Clinical Research Infrastructure Network PedCRIN

CSA_ H2020-INFRADEV-2016-2017/H2020-INFRADEV-2016-1 (Individual support to ESFRI and other world-class research infrastructures)
Grant Agreement # 731046

Deliverable D3.3
Procedure for access to individual patient clinical trial data

Date of preparation: 18th Sep 2017

Working Package: WP3 (ECRIN)

Contact: Jacques Demotes
ECRIN-ERIC Director General
European Clinical Research Infrastructure Network
BioPark, 5-7 rue Watt
75013 Paris, France
Tel: +33 180058646
Jacques.demotes@ecrin.org
www.ecrin.org

Table of Contents

Abbreviations

ECRIN	European Clinical Research Infrastructure Network
PedCRIN	Paediatric Clinical Research Infrastructure Network
CORBEL	Coordinated Research Infrastructures Building Enduring Life-science Services
IPD	Individual Participant Data
IPR	Intellectual property rights
YPAGs	Young Person Advisory groups
GDPR	General Data Protection Regulation
HIPPA	Health Insurance Portability and Accountability Act

1. Introduction

Individual Patient Data (IPD) are the data recorded within a clinical study dataset associated with individual participants. This may be in the form of measurements of patient characteristics (weight, blood pressure, heart rate, etc.), a description of a patient's medical history and data collected about an individual participant's clinical outcome during the trial. It can also include clinical laboratory results or images such as X-rays, details of randomisation and treatment received, and any adverse event information. Sharing IPD from clinical studies has many challenges, such as Intellectual property rights (IPR) related issues, ensuring patient confidentiality and appropriate re-use of data by third parties. In recent years, several major organisations have called for increased sharing of the data generated by publicly funded research, including the Organisation for Economic Co-operation and Development[1], the European Commission[2], the National Institutes of Health in the US[3] and the G8 science ministers[4]. This trend reflects the growing recognition that: "Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner"[5]. With the issue of sharing data from clinical trials in the spotlight, there has been considerable focus on how industry, regulatory bodies, clinical trial funders and sponsors can amend their practices to facilitate clinical trial data sharing. Barriers to data sharing are often presented without discussion of how to overcome them, and although several stakeholders are actively encouraging data sharing activities there are still more opportunities to increase the provision of data to other researchers for further use[6] . This document is aimed at illustrating the approach that will be followed in PedCRIN trials regarding the data sharing matters.

2. Broad Consent

The process of informing trial participants about possible sharing of their data, and then gaining their explicit consent to it, is of fundamental importance, and is normally a prerequisite for the sharing of pseudonymised data (i.e. data that has been de-identified but which can still be linked back to individuals using additional but separately stored material). Data sharing activities that are an integral part of a trial (for instance data transfer between collaborating groups) can be anticipated and described in the information given to participants, and so can be included within the informed consent for trial participation. But the nature, purpose and destination of IPD sharing that may occur after the trial completes are impossible to predict. By definition, therefore, any consent for this secondary use of data cannot be fully 'informed'. Instead what should be sought from the participant is a 'broad' consent to their data being shared, with the caveat that it should be shared only for scientific purposes.

Broad consent should still be given with as much information as is possible, for instance about the reasons for data sharing (in general, not as it might relate to their own data) and the nature of any preparation of the data prior to it being shared (for instance a statement saying that it will be de-identified) along with the data sharing procedure. Like all consent, to be meaningful it must also be given without coercion, however unintended that coercion might be. In particular, the consent should be explicit and clearly separate from any other consent. It cannot be implied by the consent to participate in the trial, because it is a separate activity and not part of that trial. Nor can consent to data sharing be used as an inclusion criterion for the trial, as this implies coercion.

Gaining explicit broad consent is the only simple way to avoid the legal complexities of attempting to share data where such consent does not exist. Even though, in some jurisdictions, explicit consent for the secondary use of fully anonymized clinical trial data may not be legally necessary, there are problems with what 'fully anonymised' might mean in practice. In addition, the legal context continues to evolve, for instance with the introduction of the General Data Protection Regulation (GDPR)[7] in Europe, and future national modifications and judicial interpretations of that regulation, and it is difficult to predict possible limitations on the use of data without consent. Beyond

this pragmatic requirement for gaining consent, there is also an ethical imperative to be open and transparent with participants about the possible use of their data, which should make seeking explicit consent for data sharing mandatory.
The broad consent given should allow the future scientific use of the data. Restricting future secondary use to research in particular disease areas or types of research, for example, should be avoided, because it will be impossible to predict the source of requests for data access and how they might be categorized.
Importantly a broad consent for data sharing for research purposes is required to share pseudonymized data. However, anonymized can be shared without broad consent.

An appropriate consent process for secondary use of data should ensure the following:

1. The reasons for asking about data sharing, and the general benefits of data sharing in clinical research, should be made clear to the trial participant.
2. The nature of data preparation, storage and access should be explained to the trial participant, in detail at the time the patient documents are produced. It will also be important to describe, in broad terms, how and where the data will be stored, and how confidentiality will be maintained (e.g. by de-identification measures). Even though consent for data sharing cannot be fully informed, because the nature, purpose and destination of data sharing that may occur after the trial completes are impossible to anticipate, efforts should be made to describe the measures that will be used to protect participant privacy, the type of requests that will be considered and the scrutiny to which they will be subjected, etc. In other words the consent should be as informed as possible.
3. The information provided should be clear and concise, and couched in vocabulary understood by the trial participants (or if applicable their legal representatives). The information given to children should be clear, concise and comprehensible. Moreover, it is very important to define how the information should be conveyed to the children of various age groups (5 years and under, 6-10 years; young people 11-18 years). Comics and cartoons could be used for providing information to very small children. While establishing data sharing procedures for paediatric clinical trials special consideration should be given to the accreditation of the roles of young person advisory groups (YPAGs). Review of the content by YPAGs (e.g. eYPAGnet) on the language, terminology, design etc. could be useful. It is arguable that children are capable of being partners in research and that they have rights to receive information, to be listened to, have their wishes and feelings taken into account.
4. The explicit consent for data sharing should be reflected in the layout of the consent forms. A request for consent to secondary use of data must be clearly distinguishable from any other matters in the informed consent document.
5. Data participants should have the right to withdraw their consent for data sharing, the practical difficulties in implementing this should be made clear. In legal terms, the need for a consent is normally coupled with a corresponding right to withdraw that consent, and this is acknowledged (GDPR (Article 7.3)[7]. As long as the stored data is still pseudonymized (i.e. a participant's data can be identified), a participant's request that their data be removed from the dataset can be honored. This might involve providing new versions of datasets to repositories, and be supported by including clauses about the management of withdrawn consents in data use agreements[8]. However, the withdrawal of informed consent should "not affect the results of activities already carried out, such as the storage and use of data obtained on the basis of informed consent before withdrawal[9]. Any limitations to withdrawing consent for data sharing should be made clear in any explanatory material in the patient information sheets.

6. Long term validity of the board consent when given by children needs to be addressed clearly

3. Data de-identification, anonymization & pseudonymisation

Shared IPD from clinical trials used for further scientific research should always be de-identified and either pseudonymised or anonymized.

3.1. De-identification

De identification is not defined under the GDPR but is defined in the US, for example in the Health (Insurance Portability and Accountability Act) HIPAA regulations[10]. It means removing or recoding 18 identifiers in the HIPPA rule, removing or redacting free text verbatim terms, and often removing explicit references to dates. Participants' identification code numbers are de-identified by replacing the original code number with a new random code number. Before data can be shared, it should be de-identified removing possible identifiers to minimize the risk of re-identification. Adequate de-identification is one of the key determinants of successful protection of study participants from re-identification.

The level of de-identification required for both pseudonymised and anonymised data is the same. In all cases it should provide a high level of assurance that the data content, in and of itself, cannot be used to re-identify the individuals within the dataset. Other policies and procedures (e.g. the use of a data use agreement) also provide protection against re-identification, but de-identification is a necessary pre-requisite and should be applied to all data made available for secondary use.

Thus, if a de-identified dataset is pseudonymised the participants in it can be identified only by those who possess the relevant 'additional information'. If a de-identified dataset is fully anonymised the participants cannot be identified by anyone (leaving aside the theoretical possibility of matching against the original clinical data). If a de-identified dataset is effectively anonymised there remains only the very small possibility of matching the data against a corresponding but pseudonymised set, if it is accessible (it should not be), but the matching cannot be guaranteed, especially if the participants share many of the same data values.

3.1.1. An assessment of the residual risks for re-identification

An assessment of the residual risks for of participants in de-identified datasets should be performed. Under the GPDR, at least in Europe, there is obligation on the data controller to carry out a data privacy impact assessment, to "evaluate… the origin, nature, particularity and severity" of the "risk to the rights and freedoms of natural persons" before processing personal data. The impact assessment "should include the measures, safeguards and mechanisms envisaged for mitigating" the identified risks. This implies that the initial de-identification of data, for instance prior to its deposition in a repository, should be accompanied by such an impact assessment, ideally included within the record of de-identification. In addition, at least in a managed access environment, assessments of re-identification risk should be made when data are requested for secondary use, because a full risk assessment will be sensitive to the particular context of the planned usage, in particular any data use agreement. If the data has already been adequately de-identified, such a risk assessment may be relatively light, and in some cases may be delegated to the repository managers.

Attempted re-identification of data subjects should be explicitly prohibited in any formal data use agreement. Even when a binding agreement does not exist, attempting re-identification is likely to be illegal, and in any case should be subject to sanction. The sanctions that might be applied could be organisational (e.g. for serious misconduct) and financial (e.g. loss of access to further funding) as well as legal (e.g. for breach of contract).

3.2. Pseudonymisation

Pseudonymisation means processing personal data in such a way that the data can no longer be attributed to a specific data-subject without the use of additional information, (e.g. a dataset linking trial identifiers to identified or identifiable persons) provided that such additional information is kept separately and under controlled access, to prevent the data being identifiable in isolation. Though theoretically such information could be used to match against a clinical trial dataset and identify individuals, this would be very difficult in practice and could only occur if there was a major breach of security.
Sharing of pseudonymous data is recommended and should be the normal expectation. Clinical trial data is pseudonymous when collected, or can be easily turned into pseudonymous data within the research unit, by processing of the data set and splitting off the identifying data points. It would be rare for trial data to become fully anonymised, or at least not until many years have elapsed after data collection. There are legal obligations on sponsors to maintain the pseudonymised dataset, as collected, for many years, the exact time depending on national regulations. In addition, the original investigators, or their institution, may want to use the pseudonymising key in case they wish to return to the same participants to carry out further investigations (assuming they have the ethical approval and / or explicit consent to do so).
The advantage of sharing pseudonymised data is that, if the secondary user discovers good reasons for clarifying, expanding or matching some of the data, or even for further investigations with some of the source population, they can contact the holders of the pseudonymous data and discuss if and how this might be achieved, because the individual participants are still (indirectly) identifiable. This does not mean that identifiable or identifying information would be transferred to a secondary user, unless there was explicit consent from the participant for this to happen (though this seems unlikely to be given). It only means that if a case can be made for identifying the individuals in the data set it is at least possible to discuss the possibilities of doing this, including possibly returning to the individuals concerned to request additional consent.

3.3. Anonymisation

Anonymisation is a technique applied to personal data to make it, in practice, unidentifiable. Full (complete, or irreversible) anonymisation involves de-identification and the destruction of any link to an identified or identifiable person via a pseudonym. Effective anonymisation can be applied to a specific dataset, by de-identification and removal of the link to a pseudonym, coupled with the use of new identifiers for individuals. There is no link maintained between these new internal identifiers and any others that might exist, for example in another pseudonymised data set, (e. g pseudonymised data set of the sponsor).

4. Data access

Access to individual-participant data and trial documents should be as open as possible and as closed as necessary, to protect participant privacy and reduce the risk of data misuse.
Depending on several factors (e.g., the nature of the consent obtained, risk of re-identification, concerns about stigmatization, misuse of information, incorrect analysis etc.), access models may range from publicly accessible web based systems, with the possibility of downloading datasets, through various types of request/review mechanisms that may or may not allow data download. A granularity of access may also be applied on different parts of the same datasets, as some pieces of information may be more sensitive or difficult to handle than others.
All secondary data users should acknowledge and agree to some basic rules of data use. For instance, they should identify themselves (including validating their email address using a call-back and confirmation process), not attempt to re-identify participants, make the results of any secondary analyses public, and cite the data source correctly in any published work. The definition of international standard practice for data sharing would usefully clarify these basic rules, and help to alleviate the fears of researchers about possible problems. At its simplest compliance with the basic

rules of re-use could be signaled by completing a web-based form. More detailed attestation or formal agreement is likely to be needed in some situations, for example if the original consent to secondary use mention possible restrictions, data sensitivity is high, or the data generators are concerned over misinterpretation.
Access to data should not be limited to a specific type of requester or professional profile. The requesters or their team would, however, normally need to demonstrate the ability to draw scientifically literate conclusions from the data. If access is formally managed, the data requester may need to provide a research protocol and analysis plan, including information on data management, data storage, and plans for publication of the results of the re-analysis. The requester should also provide information on his/her (or team) expertise, possibly making use of persistent digital identifier systems (e.g. ORCID).
Depending on the informed consent status, pseudonymisation/anonymisation of data or the investigator or sponsor policy, various options could be considered for data sharing.
This includes for instance

- Access to downloadable anonymised datasets (possible even without broad consent)
- Access to non-downloadable pseudonymised data with a broad consent, for aggregation by virtual machines
- Managed controlled access through request for data sharing policy submitted to a data custodian board etc.

5. Tools for data management and repositories

Data and trial documents made available for sharing should be transferred to a suitable data repository, to help ensure that the data objects are properly prepared, are available in the longer term, are stored securely and are subject to rigorous governance.

Data repository

A designated data repository can be dedicated to clinical research data and documents on a global or regional level, a general scientific repository, or one specialised in storing data objects related to a specific disease area. It may be a repository established by the researchers' own institution for 'their' research.
A dedicated repository is very useful for transfer of data being shared because of the following reasons:

1. The original research team (or collaboration) will change its composition, or may even cease to exist, and it may then become difficult or impossible for data to be managed and requests to be properly considered.
2. The transfer of data to a third-party repository makes it more likely that preparation of the data for sharing (e.g. de-identification, provision of metadata) will occur, and help ensure that the data and related documents are properly described.
3. Planning for transfer to a repository helps to explicitly identify data preparation and sharing costs at an early stage of the trial.
4. It helps to makes the data and trial documents more easily discoverable.
5. It can relieve the original research team / sponsor of the need to review requests and even (depending on the arrangements made with the repository) of the need to make the decisions about agreeing to such requests.

Repositories for clinical data and data objects should be compliant with defined quality criteria. In order to give its users confidence that their data and documents will be stored securely and in accordance with the specific data transfer agreements they have agreed. Some generic standards and criteria for trustworthy digital repositories have been developed and are being applied (e.g. Data Seal of Approval[11], ICSU World Data Systems[12], DIN 31644[13]) and several instruments for certification of repositories have been implemented[11;12;14;15].

The necessity for collaboration and harmonization of these different activities has been acknowledged[16] and proposals for a unified core set of requirements for trustworthy data repositories have recently been made (ICCSU/WDS, DAS[17]). The available standards, requirements and certification instruments for trusted data repositories need to be examined and their applicability to clinical research data objects needs to be checked. If necessary, extensions or adaptions should be provided.

There will also be a need to develop or adapt sustainable systems to assess repositories for clinical data and data objects against these standards. This is all work still to be undertaken but, given the likely variety of repositories that will be available to researchers, we see it as a necessary part of any acceptable data sharing environment. Research infrastructure organisations can play a key role in developing and disseminating both the standards and the assessment systems.

Information about the different repositories that hold clinical research objects should be made available to data generators so that they can make an informed choice, so far as local policies allow. This information should include costs as well as the features and access options available, and any assessment against the quality standards described above. The purpose is simply to assist the data generators in their decision on where to store data objects, as well as to encourage some healthy competition between repositories. Information about the different repositories that hold clinical research objects should be made available to data generators so that they can make an informed choice, so far as local policies allow. Moreover, the transfer of any data objects to repositories (including those within the same institution) should be subject to a formal agreement that set out the roles, rights and responsibilities of the data generators and the repository managers.

6. CORBEL Principles of patient data sharing

Ten principles emerged from the CORBEL consensus process (Table 1), representing what the task force saw as the fundamental requirements for any framework for the sharing and re-use of clinical trials data[6].

Table 1: Principles of Data Sharing in Clinical Trials[6]

No.	Principles of data sharing
1.	The provision of individual-participant data should be promoted, incentivised and resourced so that it becomes the norm in clinical research. Plans for data sharing should be described prospectively, and be part of study development from the earliest stages.
2.	Individual-participant data sharing should be based on explicit broad consent by trial participants (or if applicable by their legal representatives) to the sharing and re-use of their data for scientific purposes.
3.	Individual-participant data made available for sharing should be prepared for that purpose, with de-identification of datasets to minimise the risk of re-identification. The de-identification steps that are applied should be recorded.
4.	To promote inter-operability and retain meaning within interpretation and analysis, shared data should, as far as possible, be structured, described and formatted using widely recognised data and metadata standards.
5.	Access to individual-participant data and trial documents should be as open as possible and as closed as necessary, to protect participant privacy and reduce the risk of data misuse.
6.	In the context of managed access, any citizen or group that has both a reasonable scientific question and the expertise to answer that question should be able to request access to individual-participant data and trial documents.

7.	The processing of data access requests should be explicit, reproducible, and transparent but, as far as possible, should minimise the additional bureaucratic burden on all concerned.
8.	Besides the individual-participant datasets, other clinical trial data objects should be made available for sharing (e.g. protocols, clinical study reports, statistical analysis plans, blank consent forms), to allow a full understanding of any dataset.
9.	Data and trial documents made available for sharing should be transferred to a suitable data repository, to help ensure that the data objects are properly prepared, are available in the longer term, are stored securely and are subject to rigorous governance.
10.	Any dataset or document made available for sharing should be associated with concise, publicly available and consistently structured discovery metadata, describing not just the data object itself but also how it can be accessed. This is to maximise its discoverability by both humans and machines.

7. Procedure for access to IPD for paediatric clinical trial through PedCRIN

Providing the scientific community with access, upon request, to individual patient-level clinical trial data is part of the European Clinical Research Infrastructure Network (ECRIN) eligibility criteria, assessed during the evaluation of projects asking for support to trial management. This eligibility criterion has been used for the selection of projects for Paediatric Clinical Research Infrastructure Network (PedCRIN) funding. One objective of PedCRIN project is to develop a framework in which, ultimately, all patient level data from paediatric and neonatal clinical trials become available to those who can demonstrate they can make appropriate use of it. However, there is currently no established procedure (who should request data, what should be the content of the protocol for re-analysis/meta-analysis, who should be the data custodian providing access etc.).

To overcome this barrier PedCRIN members (WP3) are establishing a procedure for patient level data sharing from paediatric and neonatal trials with special considerations to (i) the observational studies as they represent a large part of the data collected in children, (ii) patient consent and (iii) accreditation of the roles of YPAGs.

8. Conclusion

Recommendations set by the Horizon2020-funded project CORBEL would not only assist PedCRIN members towards establishing a procedure for patient level data sharing from neonatal and paediatric trials but will also provide a window of opportunity for implementing and testing in practice. Well established data sharing procedures will accelerate new discoveries by avoiding duplicative trials, stimulating novel ideas for research, and permitting the maximal scientific knowledge and benefits to be gained from the efforts of clinical trial participants and investigators. However, due to the lack of special considerations within the CORBEL document, PEDCRIN is proposing few but relevant points to be considered when paediatric data are concerned.

9. Reference List

(1) The Organisation for Economic Co-operation and Development. OECD Principles and guidelines for access to research data from public funding http://www.oecd.org/science/sci-tech/38500813.pdf. 2007.

(2) C(2012) 4890 final Commission recommendation of 17/7/2012 on access to and preservation of scientific information. European Commission. Available at http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf. 2012.

(3) G8 Science Ministers Statement. Available at https://www.gov.uk/government/news/g8-science-ministers-statement. 2013.

(4) G8 Science Ministers Statement. Available at https://www.gov.uk/government/news/g8-science-ministers-statement. 2013.

(5) RCUK Common Principles on Data Policy. Research Councils UK. Available at http://www.rcuk.ac.uk/research/datapolicy/. 2015.

(6) Christian Ohmann, Rita Banzi, Steve Canham, Serena Battaglia, Mihaela Matei[4], Chris Ariyo, Lauren Becnel, Barbara Bierer, Sarion Bowers, Luca Clivio[2], Monica Dias, Christiane Druml, Hélène Faure, Martin Fenner, Jose Galvez, Davina Ghersi, Christian Gluud, Trish Groves, Paul Houston[6], Ghassan Karam, Dipak Kalra, Rachel Knowles, Karmela Krleza-Jeric, Christine Kubiak[4], Wolfgang Kuchinke, Rebecca Kush, Ari Lukkarinen[5], Pedro Marques, Andrew Newbigging, Jennifer O'Callaghan, Philippe Ravaud, Irene Schlünder, Daniel Shanahan, Helmut Sitter, Dylan Spalding, Catrin Tudur Smith, Peter Van Reusel[6], Evert-Ben Van Veen, Gerben Rienk Visser, Julia Wilson, Jacques Demotes-Mainard[4]: Sharing and re-use of individual participant data from clinical trials: principles and recommendations. CORBEL (Coordinated Research Infrastructures Building Enduring Life-science Services). *Manuscript submittted for publication*. 2017.

(7) General Data Protection Regulation. Available at http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32016R0679, accessed. 10-7-2017.

(8) UK Data Service. Consent for data sharing. Available at https://www.ukdataservice.ac.uk/manage-data/legal-ethical/consent-data-sharing/withdrawing-consent, accessed . 10-7-2017.

(9) Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on clinical trials on medicinal products for human use; Available at https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-1/reg_2014_536/reg_2014_536_en.pdf accessed. 10-7-2017.

(10) HIPAA Privacy Rule, Code of Federal Regulations, 45CFR164.514. Available at https://www.ecfr.gov/cgi-bin/text-idx?tpl=/ecfrbrowse/Title45/45cfr164_main_02.tpl, accessed. 10-7-2017.

(11) Data seal of approval: Certification of sustainable and trusted data repositories. Available at https://datasealofapproval.org/en, accessed. 10-7-2017.

(12) International Council for Science (ICSU). World Data System (WDS): Trusted data services for global science. Available at http://www.icsu-wds.org, accessed. 10-7-2017.

(13) DIN 31644: Information and documentation - criteria for trustworthy digital archives. 2017.

(14) Nestor certification Working Group: NestorSeal for Trustworthy Digital Archives, 2013. Available at http://files.dnb.de/nestor/materialien/nestor_mat_17_eng.pdf, accessed. 10-7-2017.

(15) International Organization for Standardization (ISO): 16363. 2012. Space data and information transfer systems -- Audit and certification of trustworthy digital repositories. 2017.

(16) TrustedDigitalRepository.eu: a collaboration between Data Seal of Approval, the Repository Audit and Certification Working Group of the CCSDS and the DIN Working Group "Trustworthy Archives - Certification". Available at http://trusteddigitalrepository.eu/Memorandum%20of%20Understanding.html, accessed. 10-7-2017.

(17) Repository Audit and Certification DSA-WDS Partnership WG Recommendations. 2016. Available at https://www.rd-alliance.org/group/repository-audit-and-certification-dsa%E2%80%93wds-partnership-wg/outcomes/dsa-wds-partnership accessed. 10-7-2017.