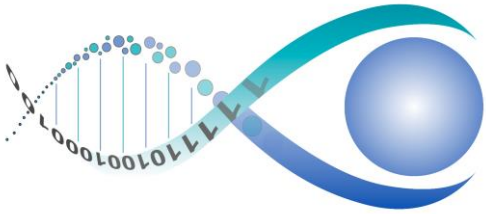


H2020-MSCA-ITN- 2018- 813545

HELICAL

Health Data Linkage for Clinical Benefit

DATA MANAGEMENT PLAN

Submitted x June 2019.

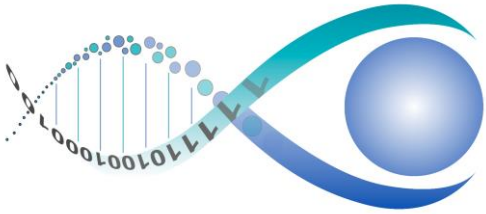
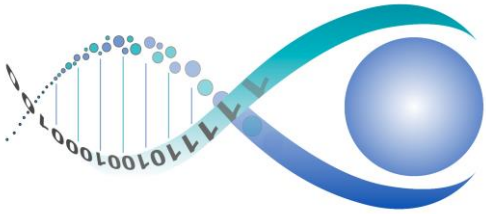


Table of Contents



1.Governance framework for the conduct of research using data and samples within HELICAL

1.1 Strategy for developing the HELICAL governance framework

The instruments (policies, codes, rules, template agreements) that will be defined for governing the conduct of research across the consortium partners will be designed to facilitate the adoption of consistent and legally-compliant practices for all of the research studies, across all work packages, giving confidence to those conducting research and those making data and samples available for it. The instruments will ensure that the autonomy and decision-making of each data/sample source is respected, including adherence to any local governance arrangements the source may be obliged to follow.

The development of the necessary governance instruments will be led by work package 4, will involve consultation with all of the partners and subsequent endorsement by our Information Governance Board (to be appointed). These rules will take as their starting point the assurances given in the Project Description of Action. The understanding and adoption of these instruments will be supported with guidance and training provided by i~HD, for partners and especially for the newly appointed ESR students.

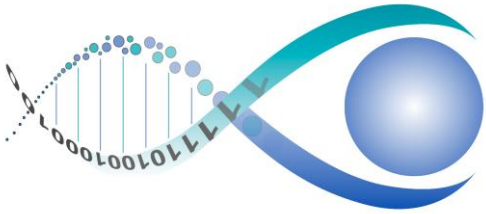
Common (project-wide) framework instruments will be developed to be as generic and all-encompassing as possible, but individual research studies, with specific data and sample access requirements, may need to add supplementary clauses on a case-by-case basis. The newly appointed ESR students will be introduced to these instruments as part of their induction training in information governance and information security. They will be invited to identify any necessary adaptations and special provisions needed for their particular research, which they will use to create customised instrument versions (if necessary) for approval by the relevant data and sample sources as well as their employing research organisation.

Experience of use, and novel areas of requirement, will be monitored throughout the project and the instruments updated as needed.

1.2 Governance principles

The following over-arching principles will apply to the governance framework.

- Data and sample access, sharing and research use will always comply with applicable national legislation, with the EU GDPR, and with any other applicable European Regulations and Directives relating to the conduct of research and the rights of citizens.



- Data and samples provided by partners within Helical may only be used for research undertaken as part of the project, by students, staff or contractors of consortium partners and specified through ethically approved protocols.

- Data/sample sources will always have autonomy over which assets are made accessible through Helical, and for which research studies.

- Data/sample sources will always determine the ethical acceptability and scientific validity of research studies that may be conducted on the assets they hold, and will formalise that access - including rules about how any shared data must be protected and governed - through signed data/sample sharing agreements with each data user, for each research study.

- Data/sample sources must be transparent about the data that are available, including any known quality issues with the data or samples.

- Helical will require all research users to respect and adhere to the ethical rules and privacy protection policies applicable to each source of the data/samples they use. All parties - sources and users - are expected to apply state of the art information security measures, de-identification techniques and privacy enhancing techniques to ensure that the privacy of the data subjects is always maximally protected.

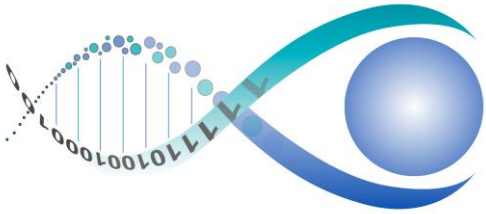
- Research users must acknowledge the sources they have used, and Helical, when publishing research findings, and include authors from the sources if this has been agreed in advance.

- Data that has been shared or made accessible through Helical may only be used for the specific agreed research purposes, and copies of any provided data be retained only for the duration necessary to conduct and validate that research.

1.3 Roadmap for developing the governance framework

Most of the generic instruments necessary to enable the conduct of research by the ESRs will be developed in the first several months of the project, hopefully also endorsed by each partner organisation by then.

- This work will begin at the start of the project, led by i~HD, to describe and catalogue the inventory of data and sample assets that are being contributed by the partners for Helical research. These descriptions will include a summary of the existing data/sample access rules at each site that would apply to the research.



- In parallel, the outline Helical research studies will be examined to identify the data and sample accesses and processing that are likely to be needed for their conduct.

- These activities will result in a generic Data Protection Impact Assessment that will be constructed for the project as a whole. Individual partners will be invited to take this generic DPIA and personalise it to their local environments, providing feedback on any issues that might affect the conduct of research using their assets. Partners will also need to review and possibly update transparency notices to indicate any additional processing that is taking place through Helical.

- The DPIA and awareness of the local access rules across the partners will inform a Code of Conduct relating to the expectations on data and sample providers, and research users, also reflected in a boilerplate data access and sharing agreement that can be tailored for specific uses (See Annex 1).

- A recommended minimum set of information security measures will also be specified as a Helical information security policy, to provide a baseline assurance across the consortium of the protections that will apply to all shared assets (See Annex 2).This may also need to be customised for specific assets and accesses.

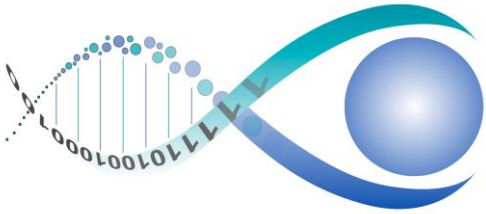
- Additional details will be developed on recommended practices for anonymisation and pseudonymisation, also to ensure a baseline project-wide assurance of standards.

- A Data Management Plan will be developed that includes recommendations for the kinds of output that should be made available open access, subject to partner level agreement on a case by case basis.

- As the project progresses, i~HD and WP4 will give support to the development of research governance instruments such as patient information leaflets, consent forms etc.

Existing instruments and policies will be requested from consortium partners, hoping that we can find the best alignment across them and minimise inter-partner difficulties with the sharing of assets and the conduct of the research. Partners with a special interest in these areas will be invited to contribute to authoring and reviewing draft versions of these instruments, prior to full consortium partner endorsement.

It is expected that some aspects of research on rare diseases will give rise to governance issues for which there is not yet legislation or established consensus. These are likely to be the areas that are investigated as part of the research undertaken by the WP4 ESR during the project.



2. Data protection and protection of privacy

2.1 General considerations

HELICAL will implement procedures to protect data and privacy of individuals consenting to participate in the studies. No data that might identify a patient as individual will be held on PCs or computers linked to a network. The project will include analysis of human tissue (aortic and kidney), human leukocytes and human biological fluids (serum, plasma and urine).

2.2 Tissue samples

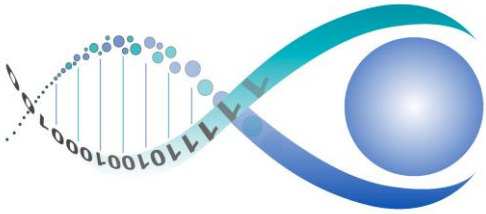
Tissue samples for immunohistochemistry or in situ hybridization will be coded by assigning consecutive numbers to patients and serum samples when consenting to the study and the information linking patient identity and code (assigned number) will be held as paper copy only by the investigators at a secure place. Only the principal researchers in the study(ies) will be able to link the code with the experimental results. These linkage tables and algorithms will be held securely. No information which might identify a patient as an individual will ever be made available. Subsequently, serum samples or clinical data required will be identified by patients' numbers only. Where medical records will be inspected during the studies, no personal data will be stored. Data of interest for the analysis of the results, such as age distribution of participants, diagnosis, treatment or laboratory results will be anonymised.

2.3 Anonymisation procedures

Random numbers will be assigned to patient's tissue, serum or blood samples when consenting to the study and a second list will contain numbers that identify certain disease entities, for instance number 1 = Wegener's Granulomatosis. The code will thus contain a set of two numbers and subsequently, serum samples or clinical data required will be identified by these numbers only. A paper copy of the code linking patients' identities with the assigned numbers will be held at a secure place to identify if necessary, individual subjects. Only anonymised research data will be stored on a computer. No data that might identify a patient as individual will be held on PCs or computers linked to a network. Additional Privacy Enhancing Techniques will be applied to relevant datasets, in accordance with state of the art and under guidance of the Ethics and Information Governance Officer.

No experiments will commence before the consent of relevant local and national Ethics committees and Regulatory Authorities has been obtained. The collection of personal data will be conducted under the applicable international, EU and national laws and regulations and requires previous written informed consent by the individual.
HELICAL researchers commit to the highest standards of data security and protection in order to preserve the personal rights and interests of study participants. They will adhere to the provisions set out in the

- Directive 2006/24/EC of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communication services or of public communications networks.



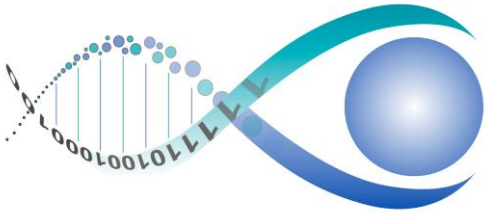
- Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications) and
- Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

Prior to collecting, storing and processing sensitive health data the consortium will seek consent of the applicable local and/ or national data protection authorities and work within the processes recommended in the e-Health Task Force Report "Redesigning Health in Europe for 2020".

To secure the confidentiality, accuracy and security of data and data management, the following measures will be taken:

- All clinical information and biological samples and tissues obtained from the clinical networks will be transmitted to partners within the consortium only after anonymisation/blinding. Keys to identification numbers will be held confidentially within the respective clinical units. In situations were re-identification of study participants becomes necessary (unblinding), for example the collection of additional data or in the case of incidental findings, this will only be possible through the treating unit and in cases where informed consent for such cases has been given. Policies and protocols detailing under which conditions unblinding may be requested will be developed at the project outset, aiming at maximum confidentiality, and which will guide the wording of relevant unblinding clauses within consent forms.
- Clinical data are entered to secure websites, for example in the case of electronic case report forms (eCRFs), observing state-of-the-art encryption technology. Collected data that cannot be entered directly or originates in patient files will be archived securely according to local data protection standards.
- eCRFs and other forms needed for the collection of patient data will be unified and reviewed by the relevant authorities as well as the Ethics Advisory Board to ensure only adequate and relevant information will be recorded
- Data are processed only for the purposes outlined in the patient information forms of the respective studies. Use for other purposes will require explicit patient approval. Also, data are not transferred to any laboratories or places out-side the consortium without patient consent.

It is currently not foreseen that any personal data will be imported from or exported to Non-EU countries. Access to anonymised experimental biomaterial (serum samples or tissue) will be granted to partners in Australia and the US for restricted use within the HELICAL project in line with the scientific cooperation set out in the Consortium agreement. Necessary authorisations will be obtained (including informed consent in case of samples from human subjects) for materials identified during the project to be exported for research purposed to non-EU countries, and export procedures recorded. Data handling in will be fully conforming to national laws and regulations and the European Directive 95/46/EC.



Access to anonymised experimental data will be granted to partners in to partners in Australia and the US for restricted use within the HELICAL project. Data handling in to partners in Australia and the US will be fully conforming to national laws and regulations and the European Directive 95/46/EC. In cases of contradiction the tighter regulation shall prevail.

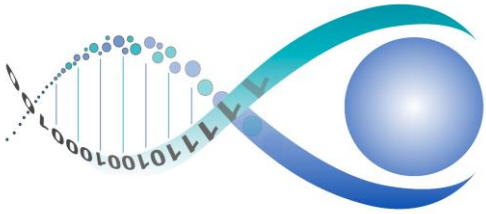
2.4 Commercial use

No personal data will be used for commercial purposes, but the knowledge derived from the research using the personal data may be brought forward to such use as appropriate, and this process will be regulated by the Grant Agreement and the Consortium Agreement, in accordance with any generally valid legislation and regulations.

3. Information Governance and Information Security

3.1. *Information governance and information security codes of practice*

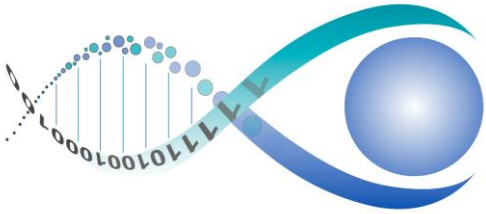
HELICAL aims to conduct novel bioinformatics research, primarily making use of existing clinical research databases and registries held by the consortium partners across Europe. The research will mainly be anchored at each of these sites, using the locally held clinical and life sciences data. All such new research studies will be required by HELICAL to comply with the data subject consent that exists at each data source, and to obtain and to comply with local ethical approval. This may at times require fresh consent to be elicited for requesting additional data from the database subjects, or for new analyses to be performed on bio-samples. It is anticipated that HELICAL research at each centre will largely comply with its existing ethical, information governance and privacy protection policies and measures. However, HELICAL will draw on existing such instruments that are being collated and further developed by i~HD at a European level, which will set a default standard across the consortium and may at times be adopted at individual sites, where the research processing of data and samples goes beyond the scope of local policies. Support will be provided to assist individual centres with strengthening their information security practices, where necessary.



At times cross-centre research collaborations will require data sharing, potentially between European countries. We foresee no requirement for identifiable data to be transferred between sites, but there is likely to be a need to transfer anonymised subject level data. This consortium will balance identification and protection of data identifying an individual's data when working with these data. In addition to ethical approval being sought from each of the relevant data sources, overarching HELICAL policies and measures will be applied to such data transfers (or to distributed querying). In order to ensure that the necessary research can be undertaken across HELICAL, and yet to minimise, control and audit the disclosure of subject-level data, HELICAL will conduct a Privacy Impact Assessment to inform this overarching information governance policy. An information security policy will be implemented across all of the cross-centre information flows and repositories. A uniform base standard for information security will apply across the consortium (designed to interface with the existing policies and practices at each organisation), to provide a consistent assurance of privacy protection across all HELICAL research. The WP4 members will have oversight of the site-based information security practices, and the remit to investigate any issues or concerns that arise during the project. We will support and have oversight of consent and ethical approvals in place and newly made across sites, and support good practices in capturing any new subject consent.

Good practices in data sharing, including a standard data sharing agreement, will be used across the consortium which defines in advance the data to be shared for each collaborative study, the data access mechanisms, the de-identification and security measures to be applied, and also includes predefined agreements about intellectual property, publication and authorship etc.

Consortium partner i~HD already has a portfolio of policies and governance instruments, including a data sharing template, that have been developed and refined through previous EC and IMI projects (including EHR4CR and EMIF) which respond to the obligations in the new European GDPR and conform to the IMI Code of practice on secondary use of medical data in European scientific research projects. Its experts will lead this task and refine the instruments to meet the needs of HELICAL, and then promote any improved versions within other consortia across Europe. If needed, we will develop a Trusted Third-Party policy to govern the involvement, activities and accountability of any external parties involved in performing linkage on behalf of the partners or their data sources.



3.2. Information Governance Board

Consortium partner i~HD will co-ordinate the constitution and operation of an Information Governance Board, largely comprised of experts in research ethics, data protection, information security. It will also include stakeholders with an important interest in the protection of privacy and good practices in the reuse of information for research, especially patient representatives. We will nominate and appoint external experts and internal project representatives, manage their meeting logistics including any appropriate honoraria and travel reimbursements, manage appropriate channels to inform and provide transparency to the Board members of the activities of the project, its policies and protection measures, its information handling activities and any issues that arise. Board approval will be sought for all of the instruments that constitute the overarching project governance framework. The activities and opinions of the Board will be made appropriately transparent through the project web site.

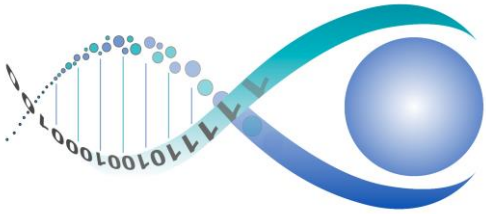
3.3. Open Data

The partners of the HELICAL consortium are strongly supportive of open access. HELICAL is committed to the principles of open data, data sharing, reusing data resources and research transparency. However, the partners cannot currently commit to an unconditional open access policy, but agree on a policy of “as open as possible, as restricted as necessary”.

Wherever possible HELICAL will create data sets that are sufficiently de-identified and/or aggregated to enable them to be published and reused. For this purpose, HELICAL will develop a Data Management Plan and template that will include the curation of standard metadata describing the data set and the mechanisms for gaining access to it by external researchers.

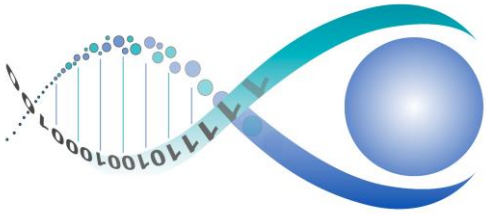
However, two important constraints need to be highlighted for transparency here. Firstly, most of the data accessed and used will be pre-existing data sources that have been curated and are being held by existing research centres, who will continue to have ownership of those data. HELICAL will therefore have limited data holdings of its own, which will limit the scope of what it can make available in the future as open research data. Nevertheless, every effort will be made to make research data subjects available, with permission from the originating data sources. Secondly, much of the novel research will be undertaken using subject level data, which will be inevitably difficult to de-identify. In compliance with national and European laws, HELICAL will primarily respect conformance to data protection legislation above any wish to make research data openly accessible.

In situations where potentially useful data sets are under the custodianship of HELICAL and which cannot be made openly accessible, HELICAL will publish a data sharing specification which makes clear how other external research teams may formulate a data sharing request in order to be able to undertake specified research using the data sets. The members of the consortium are acutely aware of the sensitivity of personal data and the need for its protection, and will implement an appropriate Data Management Plan given this context. Consortium partner i~HD has undertaken previous surveys of data management plans across research groups, and has experience of drafting data management plans, and will also make use of https://dmponline.dcc.ac.uk/. HELICAL partners have kept tack of the openAIRE pilot, and will align with the policies and metadata specifications of its successor open data service.



3.4. Good Practice in Data linkage

The overarching HELICAL policies and measures defined above will be applied to data transfers (or, probably more often, to remote/distributed querying). Drawing on work recently undertaken across the EMIF project consortium (not yet published), this will develop good practice guidelines and educational resources about how data linkage should be undertaken in ways that maximise the protection of data privacy and optimise the accuracy of the linkage.



Annex 1. i~HD Data access agreement template

i~HD Data access agreement template

This template should be used to specify the mutually agreed terms under which a research user may access a data asset in order to conduct their research.

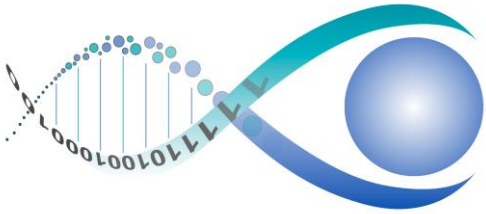
The terms data access and data sharing are used here to refer to the access provided to external researchers and research organisations by the custodian of a data asset. This access might be granted through the direct provision of a relevant data set extract, by enabling on-site querying of the data asset, access through an intermediary safe haven repository or as part of a federated network that supports distributed querying.

Specifically in HELICAL, this might most commonly be through on-site research conducted through a study placement at the location of a data asset and its academic team. However, at times the research may require a combination of data from multiple data assets (through linkage), in which case some data extraction and transfer may be necessary.

This template has been derived from work led by Dipak Kalra undertaken in the IMI EMIF project.

Template

1. **Parties (Name, designation, organisation, contact information)**
    1.1. Party providing the data intended for shared access: the data custodian
    1.2. Party wishing to undertake the specified research using the shared data: the research user
    1.3. Party to act as the trusted party on behalf of the research user, if any
2. **Date interval covering the intended use of the accessed data**
3. **The research purposes covered by this agreement**
    3.1. Specification of the research purpose(s), including therapeutic area
    3.2. Reference to a specific research protocol that has been approved by the data custodian or is under consideration



3.3. Confirmation by the data custodian that relevant permissions exist for the data disclosure and intended use by the research user (including research ethics approval, participant consent if this is the GDPR legal basis, and any specific permissions relating to sample access and analysis)

4. Arrangements for accessing and/or transferring the data

4.1. Specification of the subject population to be included in the shared data

4.2. Specification of the data items (variables) to be provided

4.3. Format(s) in which the data will be made available (e.g. file types, APIs)

4.4. Route of access or method of transferring the data

5. Arrangements for accessing and/or transferring samples

5.1. Specification of samples to be provided for the research

5.2. Specification of how the samples are to be accessed and analysis undertaken

5.3. Details of how analysis results are to be handled and which party will assume long term custodianship and controllership of the analysis results

6. Data quality

6.1. Will any metrics or assessments be provided by the data custodian regarding the quality of the data to be shared?

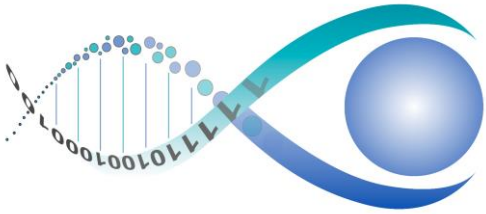
6.2. Details of the data quality information to be provided

7. Measures to protect privacy

7.1. To what extent have the data been de-identified?

7.1.1. Details of the anonymisation or pseudonymisation processes that have been undertaken, including which fields have been modified and what arrangements have been made for the handling of pseudonym keys

7.1.2. Confirmation of whether or not pseudonyms will be reused for supplementary data releases, enabling linkage between the releases



7.2. Details of any particular security conditions defining how the data custodian must protect the data, such as the use of a data safe haven, restrictions on access and/or restrictions on access to query result sets

7.3. May the research user transfer part or all of the dataset to other locations and countries within their organisation (in particular if outside the EU), and/or to other contracted third parties undertaking part of the research?

8. Data enhancing and interpretation services to be provided

8.1. Will any data enriching activities such as cleaning, variable derivation and/or analysis be undertaken by the data custodian on behalf of the research user?

8.1.1. Details of the data enriching activities to be performed

8.2. Will data linkage be performed by the data custodian prior to access?

8.2.1. Details of the datasets to be linked, how and when

8.3. Will updates to the dataset be provided during the above specified date interval?

8.3.1. Details of the anticipated data items to be updated, the nature of the updates and the frequency with which they will be provided

8.4. Will additional data collection or samples be provided during the specified date interval?

8.4.1. Details of the additional data and/or samples, and when these will be provided

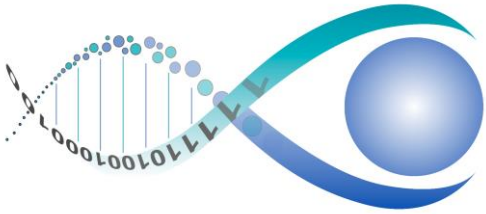
8.5. Will support be available from the data custodian for interpreting and/or analysing the data?

8.5.1. Details of the available support

9. Handling of analysis results

9.1. Details of the intended transparency and/or utilisation of the results, for example if the knowledge derived from the use of the data is intended for direct publication or other form of open access, or is intended to be used within products or services for the public good, but will not necessarily be published

9.2. Details of the handling of intellectual property, publication, authorship and acknowledgement, including the grounds on which a data custodian team member or



members should be included as authors in publications, and how a data custodian should be acknowledged within research outputs irrespective of whether a team member is a co-author

9.3. Is the research user expected to return any cleaned or derived variables?

9.3.1. Details of the variables to be enriched and/or returned

9.4. Details of the procedure to be followed in the event of the discovery of clinically significant findings (research results or incidental findings) that may have implications for the data subjects within the dataset

10. Data retention

10.1. Details of the trigger event (allowing time for the completion of the analysis and publication of the research etc.) for data destruction of any extracted data sets or analysis tables, who should perform this and what evidence the data custodian will require that the data destruction has been carried out to a required standard

10.2. Is a copy of the released data set permitted to be held as an archive for a longer period?

10.2.1. If so, for what duration, in what format, how should it be protected, what will be the grounds for access, are any audit arrangements required?

11. Demonstrating compliance

11.1. Details of the records that the data custodian must maintain in order to be able to demonstrate conformance to the terms of this agreement, such as lists of authorised users and audit logs

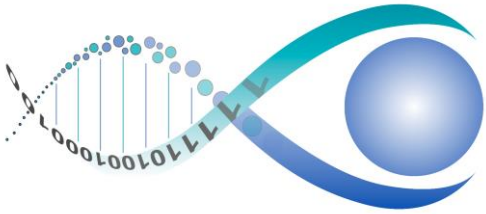
11.2. Details of the records that the research user must maintain in order to be able to demonstrate conformance to the terms of this agreement, such as lists of authorised users and audit logs

11.3. Procedures to be followed in case of a suspected breach of the terms of this agreement, and any sanctions or penalties to be applied

12. Signatories

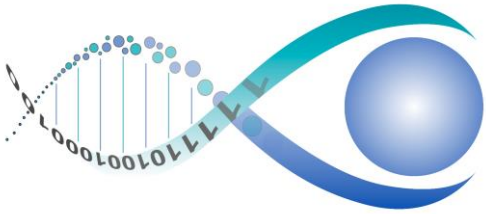
12.1. Name and designation of the signatories on behalf of the research user(s)

12.2. Name and designation of the signatories on behalf of the data custodians



12.3. Date(s) and location(s) of signature

12.4. Witnesses (if applicable)



Annex 2. i~HD Data Sharing Asset Register

i~HD Data Sharing Asset Register

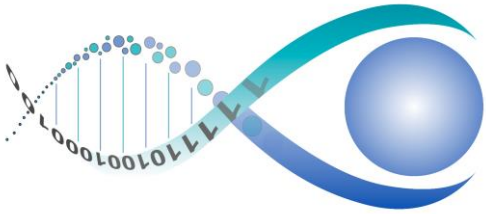
This template has been developed to capture and share information about data assets that are available for shared use within a collaborative initiative or as part of a federated network of data sources, for research. It focuses primarily on the governance aspects of data sharing.

The term data sharing is used here to refer to the access provided to external researchers and research organisations by the custodian of a data asset. This access might be granted through the direct provision of a relevant data set extract, by enabling on-site querying of the data asset, access through an intermediary safe haven repository or as part of a federated network that supports distributed querying.

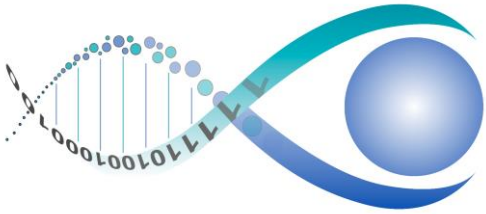
This asset register aligns with the FAIR principles, which promotes that data assets should be:

Findable: easily discovered, normally through an online catalogue or internet searching
Accessible: there is relevant metadata to support understanding of the data asset
Interoperable: adopting standards for the data and for the descriptive metadata
Reusable: the terms of data reuse are transparently provided

However, this template elaborates on certain aspects that are needed to give information and assurance to potential research users about the legal basis (including GDPR compliance) and permissions (including ethical approvals) that are in place to permit data sharing. It does not go into detail on the data or metadata describing the suitability of the data for different kinds of research, since this is the role of other data cataloguing templates. It also does not seek to capture evidence of the GDPR compliance of the data asset itself, but only in relation to the reuse of the asset for research by external organisations.



Question	Asset response
Asset descriptors	
What is the name of the asset?	
What are the overall asset objectives (the purposes for holding the data)?	
What type of cohort or population asset is it? *Options: Research study, Cohort, National registry, Healthcare provider registry, EHR extract*	
How long will the asset continue to be available for external research use?	
Organisational contacts	
What is the name of the organisation with legal responsibility for holding and maintaining the asset?	
What kind of organisation is the asset owner? *Options: Academic, Governmental, Charity, Other not for profit, Industry, other entity*	
How is the asset primarily funded?	
Administrative contact name	
Administrative contact address	
Administrative contact email	
Administrative contact phone	
Scientific contact name	
Scientific contact address	
Scientific contact email	
Scientific contact phone	
Technical contact / data manager contact name	
Technical contact / data manager contact address	
Technical contact / data manager contact email	
Technical contact / data manager contact phone	
URL of the asset website or Unit/Centre Web site	
Data subject population	
What are the main inclusion characteristics of the asset population?	
What is the approximate number of subjects in the repository?	
What is the approximate number of subjects for whom data collection is ongoing	
What is the geographical area where are the data subjects located?	
What was the age range of your data subjects at the time of recruitment?	
What are the overall time periods of the data held?	
Are you recruiting new participants? *Options: Recruiting new participants, Continuing to collect data on existing participants, No active data collection, but continuing analysis/research, Frozen and held for data sharing*	
Which care setting data sources are incorporated? *Options: Birth and child health services, Primary care, Outpatient care, Inpatient care, Pharmacy, Management reporting, Billing, Disease registry, Health event registry, Procedure registry, Bio-Bank, Other*	
Do you hold data about family or other people related to the subjects?	
Data set and metadata	
What main categories of health data are held?	
What kinds of genetic information are held?	
What kinds of bio-samples are held?	
Does the asset have a summary of the main categories of variables held?	
Does the asset have a formally documented data dictionary?	
In what data or database format are the data primarily held?	
What data standards are adopted for organising, storing, managing or protecting the asset data sets?	
Does the asset have a standard format or computing language in which data extraction queries are constructed on the data set?	
Does the asset have a process for documenting and saving locally-run queries for audit purposes and for potential future re-use?	
GDPR compliance	
Data Controller organisation details	
Data processor organisation details (if applicable)	
Main purpose of processing	
Legal basis for processing under GDPR Article 6	



Question	Asset response
Legal basis for processing under GDPR Article 9 (if applicable)	
Geographic location of personal data (including pseudonym keys, if applicable)	
Data Protection Impact Assessment (DPIA) status, with regard to data sharing	
Are there any stipulations or obligations on data sharing arising from the DPIA?	
Informed consent (if applicable)	
Has informed consent been given by the data subjects?	
Is participant consent required for new research uses (e.g. new research topics)?	
Is participant consent required for external organisations to reuse the asset data?	
How does the asset custodian handle withdrawal requests by data subjects?	
Approvals	
Which regulatory bodies have approved the collection and use of the data set?	
What additional approvals are required for external organisations to reuse the data?	
What research purposes have been approved for the asset?	
Is ethics board approval required for every additional study, only for new research areas, or is there broad approval for many kinds of research?	
Is ethics board approval required for new studies that only use an anonymous extract of the data?	
Are there any areas of research that are explicitly excluded?	
Are there specific limitations on the time during which research can be done?	
Is there a requirement for significant findings to be notified to data-subjects, and by whom?	
How frequently and in what form must studies from external organisations report to the asset's internal board, ethics committee or other bodies?	
What requirements are there for archiving or destruction of shared datasets?	
Linkage and de-identification	
Are any standard demographic identifiers retained e.g. a national health number?	
Does the asset hold names, addresses and contact information of the data subjects?	
Does the asset hold the details of external data providers e.g. GP details?	
To which external data sources does this repository have linkage?	
Are consistent pseudonymous IDs issued to permit longitudinal linkage of data sets?	
What methods are applied to generate an anonymous data set for external users?	
Which variables are normally aggregated before sharing (e.g. by age-banding)?	
Which variables are normally blurred before sharing?	
Is the data filtered in any other way (e.g. removal of rare diseases or conditions)?	
Are combinations of variables checked in case they may act as a pseudo-identifier?	
Are small-number restrictions applied to query results?	
Is there a tracking of serial queries to detect 'triangulation' attempts?	
Data sharing processes	
Is there a published procedure for submitting a data sharing request?	
Are any data access fees normally required from a data sharing party?	
Are external requesters permitted to come to the asset site and view the data before finalising their data set request, or to perform queries themselves?	
Are external parties permitted to access the data remotely, provided this is secure?	
Does the asset have documented procedures for the methods by which data set extracts may be sent to data sharing third parties, including required safeguards?	
Are there particular measures necessary in order to share genetic data?	
Are there particular measures necessary in order to share biological samples?	
Is a code of practice specified that parties reusing the data must adhere to?	
Are there specified measures or standards for data protection and information security that the research user must adopt?	
Does the asset keep a formal record of all disclosures made, including a reproducible definition of the data set that was disclosed?	
Does the asset require data sharing collaborators to provide back any newly derived or quality-improved variables?	
Is there a policy on how long after disclosure a collaborator may access or retain keep a data set?	
Is there a publication policy or guidance indicating the kinds of authorship or acknowledgement that should be included in publications derived from the asset?	

Entry completed by: Name Role Date

Annex 3. Predicted ESR Data and Sample requirements