

DELIVERABLE 11.6

POPD – REQUIREMENT NO. 7

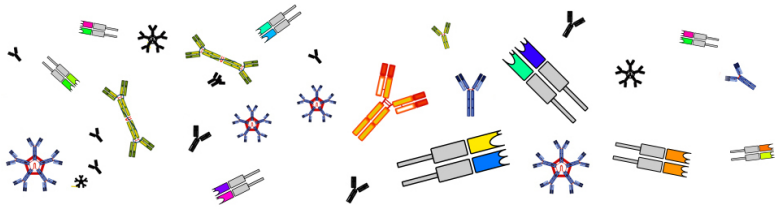
WORK PACKAGE NUMBER: 11

WORK PACKAGE TITLE: ETHICS REQUIREMENTS

ETHICS



This project is funded by the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 825821 and Canadian Institutes of Health Research (CIHR)



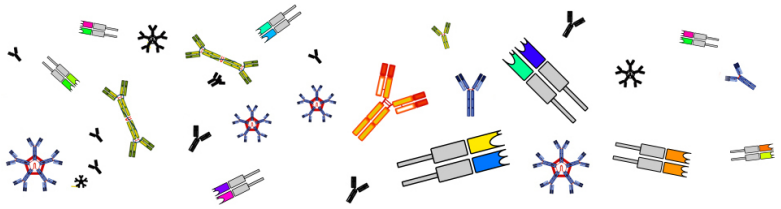
Document Information

iReceptor Plus Project Information	
Project full title	Architecture and Tools for the Query of Antibody and T-cell Receptor Sequencing Data Repositories for Enabling Improved Personalized Medicine and Immunotherapy
Project acronym	iReceptor Plus
Grant agreement number	825821
Project coordinator	Prof. Gur Yaari
Project start date and duration	1 st January 2019, 48 months
Project website	http://www.ireceptor-plus.com

Deliverable Information	
Work package number	WP11
Work package title	Ethics requirements
Deliverable number	D11.6
Deliverable title	POPD – Requirement No. 7
Description	The beneficiary must explain how all of the data they intend to process is relevant and limited to the purposes of the research project (in accordance with the ‘data minimisation ‘principle). This must be submitted as a deliverable.
Lead beneficiary	Bar Ilan University
Lead Author(s)	Jos Dumortier, Liesa Boghaert
Contributor(s)	



This project is funded by the European Union’s H2020 Research and Innovation Programme under Grant Agreement No. 825821 and Canadian Institutes of Health Research (CIHR)



Revision number	
Revision Date	
Status (Final (F), Draft (D), Revised Draft (RV))	F
Dissemination level (Public (PU), Restricted to other program participants (PP), Restricted to a group specified by the consortium (RE), Confidential for consortium members only (CO))	CO

Approvals				
	Name	Organisation	Date	Signature (initials)
Coordinator	Prof. Gur Yaari	Bar Ilan University	30.06.2020	GY
WP Leaders	Prof. Gur Yaari	Bar Ilan University	30.06.2020	GY



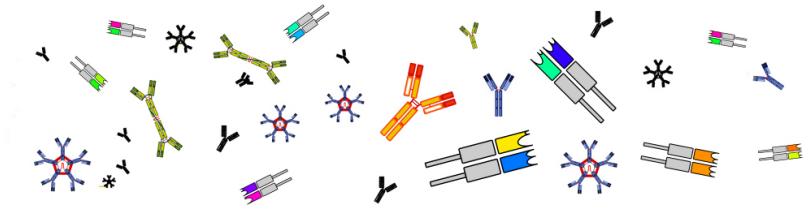
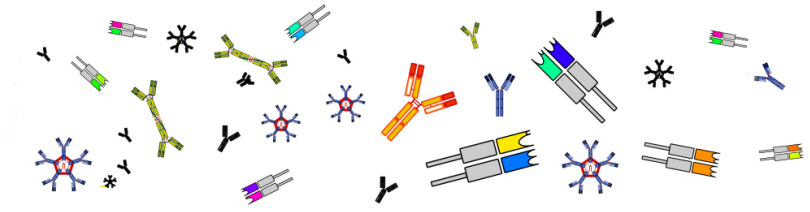


Table of Contents

Executive Summary.....	5
Introduction	6
iReceptor Plus' objectives	7
The principle of data minimisation	8
Data processing activities in iReceptor Plus	10
Data processed in the context of iReceptor Plus.....	10
AIRR Data Commons approach	11
Conclusion.....	14





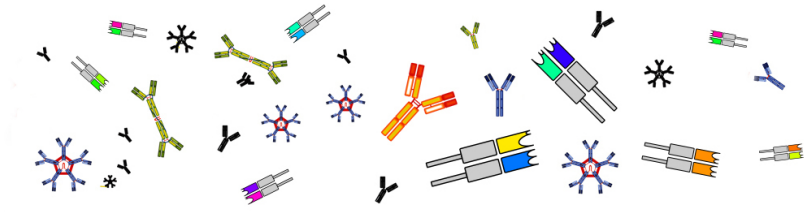
Executive Summary

D11.6 POPD – Requirement No. 7 concerns the ethics requirements of the iReceptor Plus project.

The goal of this deliverable is to detail how all of the data the consortium intends to process is relevant and limited to the purposes of the research project (in accordance with the ‘data minimisation ‘principle).



This project is funded by the European Union’s H2020 Research and Innovation Programme under Grant Agreement No. 825821 and Canadian Institutes of Health Research (CIHR)



Introduction

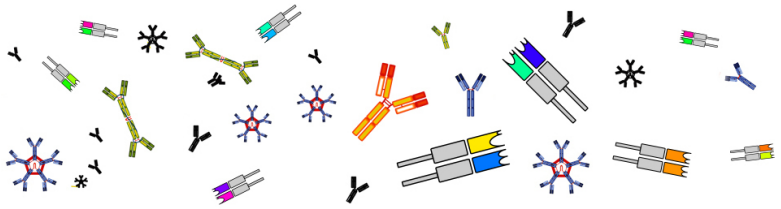
As a part of Work Package 11, POPD Requirement No. 7 expects the beneficiaries to explain how all of the data they intend to process is relevant and limited to the purposes of the research project (in accordance with the ‘data minimisation principle’).

This deliverable will satisfy this requirement by first reiterating the objectives of the iReceptor Plus project, then clarifying the data minimisation principle as it is enshrined in the General Data Protection Regulation (GDPR) and finally, explaining which datasets are processed in the context of the iReceptor Plus project and confirming that these data processing activities respect the principle of data minimisation.

This deliverable should be read in line with D10.2 on the Data Management Plan (DMP).



This project is funded by the European Union’s H2020 Research and Innovation Programme under Grant Agreement No. 825821 and Canadian Institutes of Health Research (CIHR)

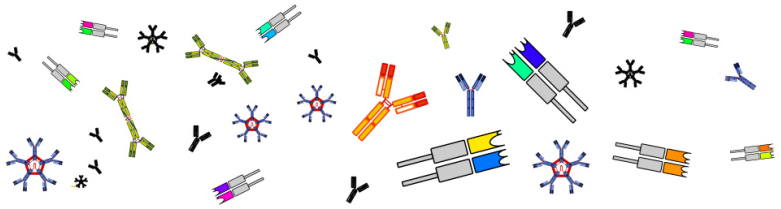


iReceptor Plus' objectives

The objective of iReceptor Plus as a research and innovation action is to build a common scalable platform to integrate distributed repositories of Adaptive Immune Receptor Repertoire sequencing data (AIRR-seq data) for enabling improved personalized medicine and immunotherapy for diseases with an immune component. iReceptor Plus will be designed as a network of federated repositories that facilitates data queries and advances analyses through a centralized web portal (the iReceptor Plus Scientific Gateway). In essence, this means that iReceptor Plus will be a software platform, which enables (1) the querying of multiple repositories at once by common metadata searches, (2) the analysis of federated data aggregated from multiple repositories and (3) the integration of these data with other types of large scale human health and genomic data.

It is thus clear that data processing activities have a pivotal role in the project, as iReceptor Plus in essence intends to lower the barrier to share, access and analyse large sets of AIRR-seq data from around the world and to ease the availability of these AIRR-seq data to academia, industry and clinical partners. Ultimately, this increased availability of AIRR-seq data will advance the understanding of immune responses and may lead to the discovery of biomedical interventions (such as vaccines and other immunotherapies) that manipulate the adaptive immune system. Such advancements will enable improved personalized medicine and immunotherapy in cancer, inflammatory an autoimmune diseases, allergies, and infectious diseases.





The principle of data minimisation

The principle of data minimisation, as referred to in POPD Requirement No.7, is laid down in Article 5.1 (c) of the General Data Protection Regulation (GDPR).

Article 5 GDPR sets out the ‘principles relating to the processing of personal data’. As such, the provision puts forward 7 key principles that should lie at the heart of any organisation’s approach to processing personal data. The principles are not hard and fast rules, but rather embody the spirit of the general data protection regulation. Consequently, there are very limited exceptions. Compliance with the spirit of these key principles is therefore a fundamental building block for good data protection practice. These principles are:

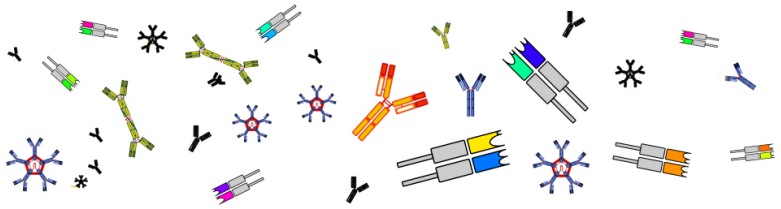
1. Lawfulness, fairness and transparency
2. Purpose limitation
3. Data minimisation
4. Accuracy
5. Storage limitation
6. Integrity and confidentiality
7. Accountability

It is important to emphasize that Article 5 (and the entire GDPR for that matter) only governs the processing of *personal data* (i.e. data relating to an identified or identifiable natural person) and does as such not apply to anonymous, non-personal data.

The third principle set forth in Article 5.1 (c) GDPR, namely ‘the data minimisation principle’ is central to this deliverable. Pursuant to this provision, “*personal data shall be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed*”.

The principle of data minimisation is not a fundamentally new concept, as Article 6, para 1, lit. c of Directive 95/45/CE stated that national legislation should guarantee that controllers process only personal data that is “not excessive in relation to the purposes for which they are collected and/or further processed”. Essentially this means that data processing must be limited to what is necessary to fulfil a legitimate purpose, and it should only take place when the purpose of the processing cannot be reasonably fulfilled by other means. Also, data processing may not disproportionately interfere with the interests, rights and freedoms at stake.





According to the United Kingdom's Supervisory Authority – the Information Commissioner's Office, the three key elements of data minimisation are to be understood in the following manner:

- **adequate** means that the processed data is to be sufficient to achieve the purpose pursued;
- **relevant** means that the processed data is to be related to the achievement of that purpose; and
- **limited** to what is necessary means that the controller is to process only the minimum amount categories of data and data sets to achieve that purpose.

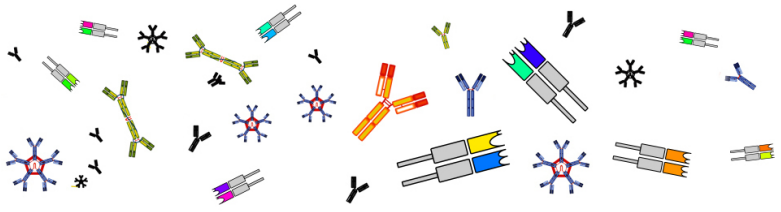
The data minimisation principle entails that the controller should strictly limit the collection of data to such information that is strictly necessary for the specific purpose pursued by the processing.

This principle is also closely related to the principle of purpose limitation which requires that personal data is to be processed with the aim to attain a specific and legitimate purpose, and not further processed in a way that contradicts this initial purpose (Article 5.1 (b) GDPR). Additionally, in order to observe this principle, the controller is to process personal data only in the cases where the purpose sought could not be achieved in any other way.

Another principle which relates to data minimisation is the principle of storage limitation (Article 5.1 (e) GDPR). The latter refers to the form used to safekeep the data so it is guaranteed that the when no longer needed, and the purpose for which the data has been collected and processed is already achieved, the controller would proceed with secure data erasure. Alternatively, if data erasure is impossible at the current stage, the controller is to take any measures available to safeguard the identity of the natural persons concerned.

Additionally, it should be noted that data minimisation requirements also refer to whom the respective personal data is available. The principle calls for information sharing on a 'need to know' basis, so only the relevant experts process the data, and not the whole partner/project team.





Data processing activities in iReceptor Plus

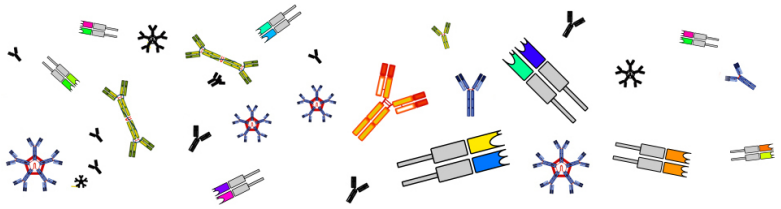
Data processed in the context of iReceptor Plus

To achieve the objectives of iReceptor Plus mentioned above, namely to develop a software platform, which enables (1) the querying of multiple repositories filled with AIRR-seq data at once by common metadata searches, (2) the analysis of federated AIRR-seq data aggregated from multiple repositories and (3) the integration of these AIRR-seq data with other types of large scale human health and genomic data the iReceptor Plus consortium will have to process certain data. In fact, data sharing (and thus data processing) is central to the project.

As was indicated in the project's Data Management Plan (D10.2), four fundamental categories of datasets will be processed in the project, more particularly:

- **Study Metadata:** These research data concern the metadata relating to a given research study that is included in a repository that is part of the iReceptor Plus network of distributed repositories. These data typically consist of data about the study, the subjects in the study, the biological samples in the study as well as the way in which these samples were obtained and processed.
- **AIRR-seq data:** These research data are the sequences that result from the processing of biological samples from research studies, annotated with a set of features critical to understand the Adaptive Immune Response.
- **External Metadata:** These research data encompass metadata that are linked to the 'core' data components above but originate from outside the iReceptor Plus Platform. Examples of External Metadata linked to Study Metadata are other immunologically relevant data such as cell phenotype, microbiome, clinical, gene expression, HLA-typing, quantified autoantibodies, cytokine and chemokine data. Genetic associated data will be provided as quantitative measures without public access to the raw sequence (except in regards of point 2.4). External Metadata linked to AIRR-seq data are for example epitopes (binding target).





- **Analysis Metadata:** These research data are generated from analysing the combination of a set of Study Metadata, AIRR-seq data and External Metadata and as such constitute 'data about other data'.

Apart from these categories, some other types of data might be collected, processed and/or generated in the course of the project, such as:

- Documents relating to meetings, focus groups, workshops...
- Academic literature and knowledge materials
- Contact details of people working for the consortium partners

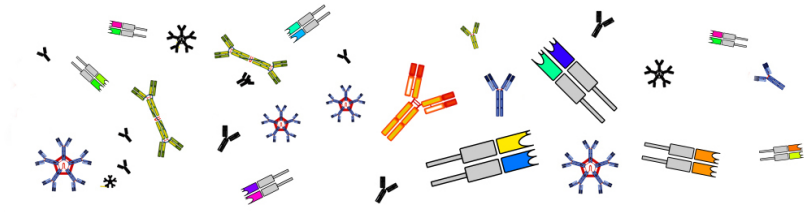
It should be emphasized however that except for the contact details of people working for the consortium partners, all of the data processed by the iReceptor Plus consortium is non-personal data, given that it is either anonymized (and thus does no longer relate to an identified/identifiable natural person) or never related to a person in the first place. As such, the data minimisation principle as enshrined in the GDPR is in principle not applicable to these data, as they are not personal. Nevertheless, due to the inherent risk of re-identification of the data processed in iReceptor Plus, the consortium aims to uphold a strict data minimisation approach throughout all its processing activities.

[AIRR Data Commons approach](#)

That is why, as one of its key characteristics, the iReceptor Plus projects builds on the achievements of the Adaptive Immune Receptor Repertoire (AIRR) Community which developed a minimal standard for repertoire metadata and sequence annotation data for studies involving AIRR-seq data (the MiAIRR standard). This standard lists the minimum data elements to be included with an AIRR study for publication according to the AIRR Community Minimal Standards Working Group.

By endorsing this standard, iReceptor Plus envisages to respect the data minimisation principle to the widest extent possible. Since the MiAIRR standard sets out the minimum data elements to be included with an AIRR study for publication, all repositories taking part in iReceptor Plus will be made MiAIRR-compliant, meaning that they will only publish those AIRR-seq data and metadata elements as foreseen by the MiAIRR standard or less if including all minimum data elements of the MiAIRR would render the data concerned identifiable.



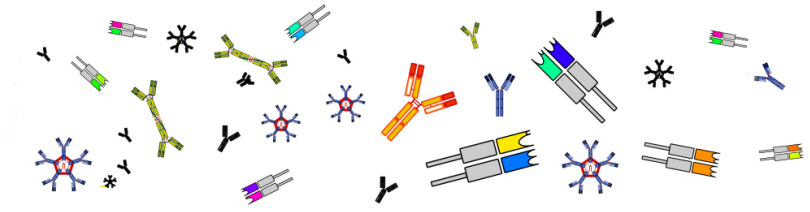


This means that each of the researchers or clinicians adding data to a repository that is part of the iReceptor Plus network of distributed repositories will decide on which AIRR-seq data and metadata he/she will add so the data is non-identifying, but these will at the very most be all data elements prescribed by the MiAIRR standard.

The data elements specified in **the MiAIRR standard** respect the principle of data minimisation, given that these data elements are considered by the AIRR community as the **minimum, adequate and relevant elements that researchers/clinicians need to perform the research that the iReceptor Plus project aims to advance**. In conclusion, the iReceptor Plus project will only process those AIRR-seq data and metadata elements that are adequate, relevant and limited to what is necessary in relation to the project's purpose of enabling researchers to perform research that advances the understanding of immune responses and may lead to the discovery of biomedical interventions (such as vaccines and other immunotherapies) that manipulate the adaptive immune system.



This project is funded by the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 825821 and Canadian Institutes of Health Research (CIHR)



Below are figured the MiAIRR standard data elements:

Study

- [Contact info \(collection\)](#)
- [Contact info \(deposition\)](#)
- [Grant funding agency](#)
- [Inclusion/exclusion criteria](#)
- [Lab address](#)
- [Lab name](#)
- [Relevant publications](#)
- [Study ID](#)
- [Study title](#)
- [Study type](#)

Subject

- [Age](#)
- [Age event](#)
- [Ancestry population](#)
- [Ethnicity](#)
- [Organism](#)

Diagnosis

- [Diagnosis](#)
- [Disease stage](#)
- [Immunogen/agent](#)
- [Intervention definition](#)
- [Length of disease](#)
- [Medical history](#)
- [Prior therapies](#)
- [Study group](#)

Cell Processing

- [# cells/experiment](#)
- [# cells/sequencing reaction](#)
- [Cell isolation procedure](#)
- [Cell quality](#)
- [Cell storage](#)
- [Cell subset](#)
- [Cell subset phenotype](#)
- [Processing protocol](#)

- [Race](#)
- [Relation to other subject](#)
- [Relation type](#)
- [Sex](#)
- [Strain name](#)
- [Subject ID](#)
- [Synthetic library](#)

Sample

- [Anatomic site](#)
- [Biomaterial provider](#)
- [Collection time event](#)
- [Sample collection time](#)
- [Sample disease state](#)
- [Sample ID](#)
- [Sample type](#)
- [Tissue](#)

- [Single-cell sort](#)
- [Tissue processing](#)

Nucleic Acid Processing

- [Complete sequences](#)
- [Fwd PCR primer target](#)
- [Library generation method](#)
- [Library generation protocol](#)
- [Linkage of loci](#)
- [Protocol IDs](#)
- [Rev PCR primer target](#)
- [Target locus for PCR](#)
- [Target substrate](#)
- [Target substrate quality](#)
- [Template amount](#)

Sequencing Run

- [Batch number](#)
- [Date of sequencing run](#)

- [Read lengths](#)
- [Reads passing QC](#)
- [Sequencing facility](#)
- [Sequencing kit](#)
- [Sequencing platform](#)

Software Processing

- [Collapsing method](#)
- [Data protocols](#)
- [Paired read assembly](#)
- [Primer match cutoffs](#)
- [Quality thresholds](#)
- [Software tools/versions](#)

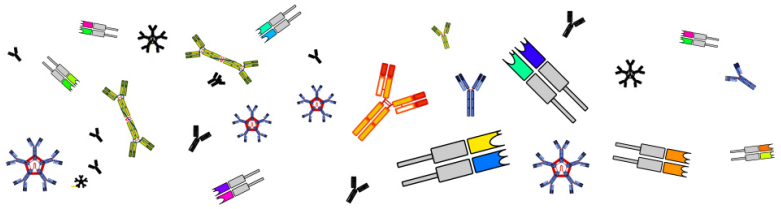
Other

- [Full-text search](#)
- [Repository](#)
- [Sequences](#)



This project is funded by the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 825821 and Canadian Institutes of Health Research (CIHR)

Conclusion



According to the principle of data minimisation, *personal data shall be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed*. Even though the amount of personal data processed in iReceptor Plus is limited, the consortium will strictly respect the data minimisation principle in all its data processing activities, by processing only the data elements put forward by the MiAIRR standard which represents the data elements that are adequate, relevant and limited to what is necessary for the research purposes envisaged in iReceptor Plus, or even less data elements, if such is required to preserve the anonymous nature of the data concerned.



This project is funded by the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 825821 and Canadian Institutes of Health Research (CIHR)