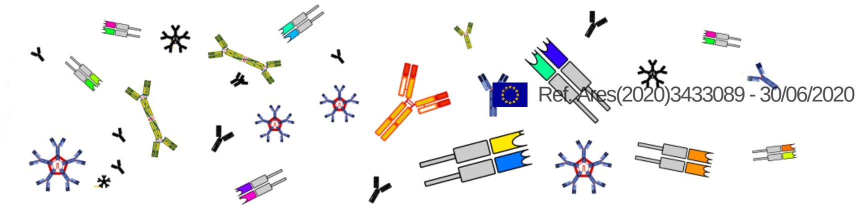


DELIVERABLE 11.7

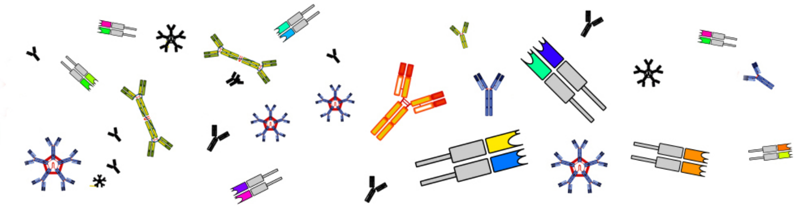
POPD – REQUIREMENT NO. 8

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	1st January, 2019, 48 months
Project website	http://www.ireceptor-plus.com

Deliverable Information	
Work package number	WP11
Work package title	Ethics requirements
Deliverable number	D11.7
Deliverable title	POPD – Requirement No. 8
Description	A description of the technical and organisational measures that will be implemented to safeguard the rights and freedoms of the data subjects/research participants must be submitted as a deliverable.
Lead beneficiary	Bar Ilan University
Lead Author(s)	Jos Dumortier, Liesa Boghaert
Contributor(s)	Artur Rocha
Revision number	1.0

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 Date	29/06/2020
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

Document History

Revision	**Date**	**Modification**	**Author**
1.0	29/06/2020	Contributions of INESC TEC	Artur Rocha

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

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)

Table of Contents

Executive Summary

Deliverable D11.7 describes the technical and organisational measures that will be implemented in iReceptor Plus to safeguard the rights and freedoms of the data subjects/research participants.

Introduction

Deliverable D11.7 aims to address POPD Requirement No. 8 which requires a description of the technical and organisational measures that will be implemented to safeguard the rights and freedoms of the data subjects/research participants.

This deliverable will first reiterate the objectives of iReceptor Plus and the related data processing activities. Then the deliverable will clarify the GDPR's requirement of implementing appropriate technical and organisational measures and will describe the appropriate measures that will be implemented in the iReceptor Plus project.

Objectives of iReceptor Plus and related data processing activities

The iReceptor Plus project essentially intends to lower the barrier to share, access and analyze large sets of Adaptive Immune Receptor Repertoire sequencing data (AIRR-seq data) from around the world and to ease the availability of these AIRR-seq data to academia, industry and clinical partners. This increased availability of AIRR-seq data will advance researcher's 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.

To achieve this objective, the iReceptor Plus aims to build a common scalable platform to integrate distributed repositories of AIRR-seq data. 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). This Scientific Gateway will allow researchers to pose complex queries about AIRR-seq data, their metadata and annotated sequence data. The Gateway then, on behalf of the end-user, will send the query to each of the repositories (using the REST API), will federate the results from each repository and present these federated results to the end-user. But the Gateway will go even further, as it will be able to stage federated data resulting from a query to an advanced analysis tool that uses computational methods on the aggregated data, such as relational datamining algorithms and deep learning techniques (AI), to facilitate complex analysis of the federated data and integrate it with other types of human health and genomic data.

The data in the distributed repositories originate from scientific and clinical studies that were undertaken by academia, industry or clinical partners. In accordance with common practice in research, these academia, industry and clinical partners deposit all data supporting their research findings in an open access repository upon publication of their research findings in scientific journals. The data concerned will be deposited in these repositories in an anonymous form, since this is required by most data protection laws and will in any case be obliged by the competent supervisory ethics committees.

A key characteristic of the project is thus that it will **facilitate a secondary use of data that was previously collected and deposited in a repository by other researchers/clinicians or industry partners**. As such, the iReceptor Plus project will not actively collect data from research or clinical trial participants. Rather, it will only provide access to data of research participants that already sits in a repository set-up by researchers/clinicians or industry partners. Those researchers and clinicians are the only ones that will have been in contact with the research or clinical trial participants at the stage of data collection, but neither they nor the iR+ consortium partners will still be able to retrace or contact those participants, given that the data have been anonymised in the meantime. The data that will be accessible through the iReceptor Plus platform will thus only be data that have undergone anonymisation techniques. As such, these data cannot be considered as personal data in the sense of the GDPR, since they are anonymised and thus no longer relate to an identified or identifiable natural person.

It is not yet clear if the iReceptor Plus consortium will at a later stage in the project also enable the sharing of AIRR-seq data that researchers have not yet uploaded to a repository in an anonymous format and that are thus, personal data. **In any case, due to the inherent risk of re-identification of the data processed in iReceptor Plus, the consortium aims to uphold the highest standards of data protection and will thus, to the extent possible, foresee appropriate technical and organisational security measures, to limit the risk of re-identification.**

The GDPR requirement of implementing appropriate technical and organisational measures

Concept

The GDPR includes many references to the implementation of “appropriate technical and organisational measures”.

For example, in Article 89.1, which relates to the safeguards and derogations related to processing for scientific research purposes, the GDPR states that:

> *“Processing for scientific research purposes shall be subject* ***to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject****. Those safeguards shall ensure that* ***technical and organisational measures are in place in***

> ***particular in order to ensure respect for the principle of data minimisation**. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner."*

Article 32.1 GDPR, on the security of processing, states that:

> *Taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall **implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk**, including inter alia as appropriate: (a) the pseudonymisation and encryption of personal data; (b) the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services; (c) the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident; (d) a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing.*

But also Article 24 (on the responsibility of the controller) and 25 GDPR (on the principles of data protection by design and by default) respectively state that the **controller should implement appropriate technical and organisational measures to ensure and to be able to demonstrate that processing is performed in accordance with the Regulation; and which are designed to implement data protection principles such as data minimisation in an effective manner**.

When it comes to defining what exactly these measures are, the GDPR is a lot less clear. In essence, technical and organisational measures are the functions, processes, controls, systems, procedures and measures taken to protect and secure the personal data that an entity processes and to facilitate compliance with data protection obligations. 'Appropriate' in this context refers to the fact that the measures taken should be suited to the intended purpose and will thus directly relate to the type and volume of personal data processed.

A technical or organisational measure can be anything from the use of advanced technical solutions to the basic training of personnel. There is no requirement to the sophistication of a measure as long as it is appropriate for implementing the data protection principles effectively.

Technical and organisational measures are moreover strongly connected to the principle of data minimisation. According to these principle, only personal data that is adequate, relevant and limited to what is necessary for the purpose of processing shall be processed.

As a result, first of all it must be determined whether personal data even need to be processed to achieve the relevant purposes. It should be verified whether technology, processes or procedures exist that could make the need to process personal data obsolete. Such verification could take place, in a particular point of the processing activity or even throughout the processing lifecycle.

Minimising can also refer to the degree of identification. If the purpose of the processing does not require the final set of data to refer to an identified or identifiable individual (such as in statistics), but the initial processing does (e.g. before data aggregation), then the controller shall anonymize personal data as soon as identification is no longer needed. Or, if continued identification is needed for other processing activities, personal data should be pseudonymized to mitigate risks for the data subjects' rights.

Technical and organisational measures in iReceptor Plus

General approach

The iReceptor Plus consortium is committed to a strong data minimisation approach. This means that it will limit the data processing in iReceptor Plus as much possible to what is necessary in relation to the purposes for which they are processed.

For example, in respect of repositories that do more than aggregating public data from open access repositories, access to data will be given in tiered levels: in first instance, public access will only be given to for example study-level metadata; in second instance, researchers can request for access to aggregate data, without access being given to the actual sequences; lastly, researchers could request for full access to sequences on a study-level basis.

This data minimisation approach will also be upheld by (as much as possible) avoiding to transfer raw sequences, but rather transferring rearrangements, which will already significantly lower the risk of re-identification, while still remaining valuable and sufficient for performing research.

Moreover, the iReceptor Plus will implement analysis endpoints for 'exploratory data analysis (EDA)' in the different repositories connected to the iReceptor Plus network of distributed repositories, which enables the repositories to extract specific data that is requested via the iReceptor Plus Gateway and sending only this data, rather than sending a larger amount of data to a central processing node from which the relevant data still need to be extracted.

Technical measures

Security values		Technical security measures in iReceptor Plus
CONFIDENTIALITY	**Access control (system authorisation)** *Use of data processing systems by unauthorized persons is prevented by:*	
	Access authorisation control	Access control is one of the main components defined in the Layered Security Framework of iReceptor Plus, set by Task 3.2. Accessibility to the main APIs used in the project is controlled using security standards based on OpenID Connect/OAuth 2.0. The current implementation of these security standards is based on an Open Source platform, the Keycloak framework. It supports federated authentication and identity brokering.
	Logging of access	Implementing adequate auditing mechanisms. A blockchain-based auditing approach is also being prototyped.
	Access control/user control (data authorisation) *It is ensured that authorised persons can only access data covered by their access authorization and that they cannot read, copy, alter, delete data without authorisation, by:*	Data authorization along with access control are main components defined in the Layered Security Framework of iReceptor Plus, set by Task 3.2. The main APIs used in the project use a current security standard named User-Managed Access (UMA 2.0). This is an OAuth-based access management protocol for data authorization. It allows for a fine-grained access control to datasets as well as full control over the sharing of datasets by their respective owners. The current implementation uses the Keycloak framework.
	Separation control *It is ensured that data collected for different purposes can be processed separately, by:*	1) Setting different access rules for study datasets according to the purpose they serve. 2) Having a network of federated repositories and processing resources.

	Disclosure/communication control *It is ensured that bodies to which data have been or may be transmitted or made available can be verified and established, by:*	Implementing adequate auditing mechanisms and a data minimization approach (layered security).
INTEGRITY	**Transmission control/transfer control** *Unauthorised reading, copying, alteration or deletion in transmission is prevented, by:*	Using SSL/TLS protocols to implement authenticated encrypted communication.
	Input control / storage control *It is ensured that it can be checked and established whether and by whom personal data is entered, altered or deleted from the system, by:*	Implementing adequate access control and auditing mechanisms.
	Storage control *It is ensured that that the unauthorisaed input of data and the unauthorised inspection, modification or deletion of stored data is prevented, by:*	Implementing authentication and authorization mechanisms. Including Web Application firewalls in the repository distributions. Implementing security by design practices. Network isolation and VPN access to servers with critical data. Vulnerability and patch Management. Backup encryption.
AVAILABILITY, RECOVERABI	**Availability control** *It is ensured that personal data are protected against accidental destruction or loss or temporary unavailability, by:*	Implementing an adequate backup policy (encrypted backups in a different datacenter). Continuous monitoring of systems and services. Periodic system penetration testing. Enterprise Endpoint Security solution. Computer Security Incident Response Team.

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)

	Restoration *It is ensured that systems may be restored in case of interruption, by:*	Implementing an adequate backup policy (encrypted backups in a different datacenter). Computer Security Incident Response Team.

Organisational measures

Information collected through the platform and stored on INESC TEC's premises will implement technical and organisational measures in order to meet the following specific security requirements:

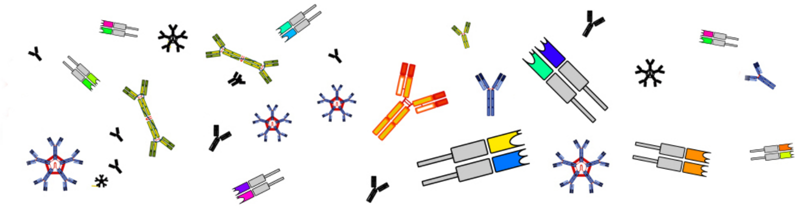
- Deny unauthorized persons access to data processing equipment used for processing personal data (equipment access control).
- Prevent the unauthorized reading, copying, modification or removal of data media (data media control).
- Prevent the unauthorized input of data and the unauthorized inspection, modification or deletion of stored personal data (storage control).
- Prevent the use of automated data processing systems by unauthorized persons using data communication equipment (user control).
- Ensure that persons authorised to use an automated data-processing system only have access to the data covered by their access authorisation (data access control).
- Ensure that it is possible to verify and establish to which bodies personal data have been or may be transmitted or made available using data communication equipment (communication control).
- Ensure that it is subsequently possible to verify and establish which personal data have been input into automated data processing systems and when and by whom the data were input (input control).
- Prevent the unauthorised reading, copying, modification or deletion of personal data during transfers of personal data or during transportation of data media (transport control).
- Ensure that installed systems may, in case of interruption, be immediately restored (recovery).
- Ensure that the functions of the system perform without fault, that the appearance of faults in the functions is immediately reported (reliability) and that stored data cannot be corrupted by means of a malfunctioning of the system (integrity).

Although the existing technical infrastructure is not ISO/IEC 27001:2013 certified, it shall enforce the aforementioned guarantees by implementing, in particular, the following security measures:

- Servers and storage are housed in data centers with biometric access control restricted to system
- administration employees, access logs, and 24x7 video surveillance.
- Virtual machines are hosted on server clusters with UPS backup power, and redundant network equipment.
- Directory for authentication and access control is linked to the Human Resources database to enforce
- contractual access only.
- Backups are encrypted and stored in a different datacenter.
- External but also internal accesses to servers are controlled and monitored with advanced firewall and IDS
- solutions.
- Network isolation and VPN access to servers with critical data.
- Vulnerability and patch Management.
- 24x7 monitoring of systems and services.
- Periodic system penetration testing.
- Enterprise Endpoint Security solution.
- Computer Security Incident Response Team.
- Procedures in the event of a security breach.

Please note that these organisational measures are the ones ensured for data stored or processed on the infrastructure of the project partner responsible for the security aspects of the project, INESC TEC. Each partner hosting data must check its own organisational measures.

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

The iReceptor Plus project is strongly committed to a data minimization approach and will therefore implement appropriate technical and organisational measures to safeguard the rights and freedoms of the data subjects/research participants. These measures are suited to the research purpose of the project and directly relate to the type and volume of data processed.