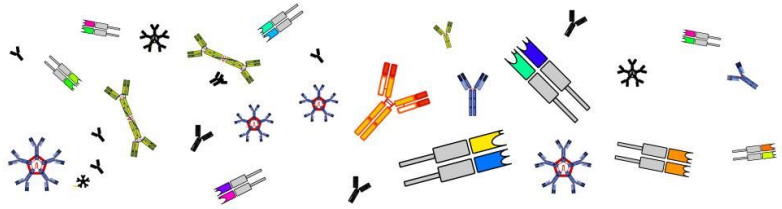


DELIVERABLE 1.2

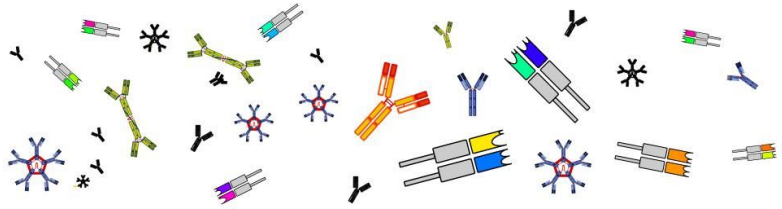
USE CASES IDENTIFIED AND DOCUMENTED FOR DEVELOPMENT AND FURTHER ITERATION

WORK PACKAGE NUMBER: WP1

WORK PACKAGE TITLE: END USER EXPERIENCE & USE CASE DEMONSTRATION

TYPE: REPORT

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



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.ireceptorplus.org

Deliverable Information	
Work package number	WP1
Work package title	End user experience & use case demonstration
Deliverable number	D1.2
Deliverable title	Use cases identified and documented for development and further iteration
Description	Document both the process used to elicit use cases from the iReceptor Plus partners as well as the use cases acquired through that elicitation process
Lead beneficiary	Ascora
Lead Author(s)	Ascora
Contributor(s)	Ascora, SFU, INESC TEC, UTSW, Interteam, BIU, Sorbonne, DKFZ, Leitat, Oslo, MedGenome, OUS
Revision number	
Revision Date	
Status (Final (F), Draft (D), Revised Draft (RV))	D

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

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))	PU

Document History			
Revision	**Date**	**Modification**	**Author**
v0.1	16-05-2019	Initial outline	Brian Corrie
v0.2	21-05-2019	Refined outline	Brian Corrie, Artur Rocha, Ademar Aguiar
v0.3	23-05-2019	Document revision	Brian Corrie
v1.0	28-05-2019 to 30-05-2019	Document revision	Brian Corrie
v1.1	01-06-2019 to 24-06-2019	Document revision	Tobias Hinz, Victor Greiff, Frances Breden, Ademar Aguiar, Jerome Jaglale
v2.0	26-06-2019	Final revisions	José Alexandre Teixeira, Artur Rocha, Brian Corrie, Amit Chaudhuri, Rui Barros

Approvals				
	Name	**Organisation**	**Date**	**Signature (initials)**
Coordinator	Prof. Gur Yaari	Bar Ilan University	30 June, 2019	
WP Leaders	Tobias Hinz	Ascora	30 June, 2019	

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

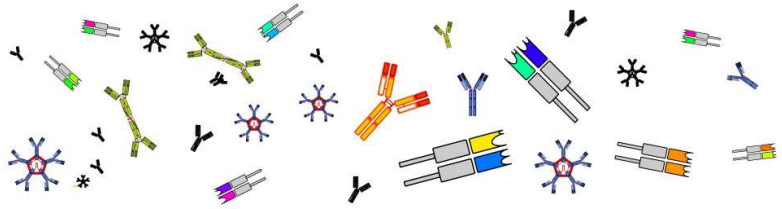
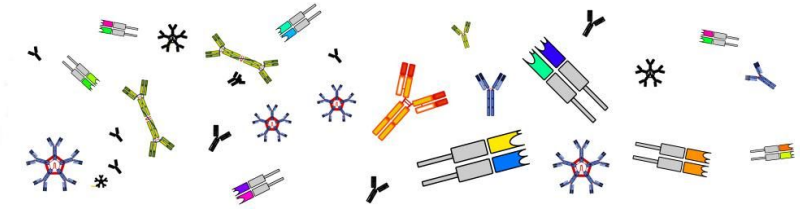


Table of Contents

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



Executive Summary

User-centered design is a fundamental component of the iReceptor Plus project and is the foundation of Work Package 1 (WP1). The objective of this WP is to establish a communication channel with the users and to ensure that the communication informs the design and implementation across all the work packages. In order to identify their requirements, use cases are elicited from the user community at the beginning of the project to guide initial development, with their feedback gathered continually throughout the project as the platform develops. Use cases and user feedback will be critical in the development of the tools in the platform, such as the Scientific Gateway (WP1), but all work packages will consider the fundamental use cases during platform development.

Deliverable description

This document is intended to satisfy WP1, Deliverable D1.2: Use cases identified and documented for development and further iteration. This document reports on both the process used to gather use cases within the project as well as describes the plan on how the use cases will be translated into requirements that will be implemented within the tasks and deliverables across the other iReceptor Plus work packages. It is important to point out that user-centered design is not a static thing, and it is intended that our use cases will evolve throughout the project and that the project will adapt to the changes in use cases in an agile and continuous manner.

This deliverable consists of:

- a description of the use case elicitation process used within the project
- a short description of each of the use cases that were obtained and from which partner the use case was obtained.
- a detailed description of each of the use case categorizations used to aggregate the individual use cases.

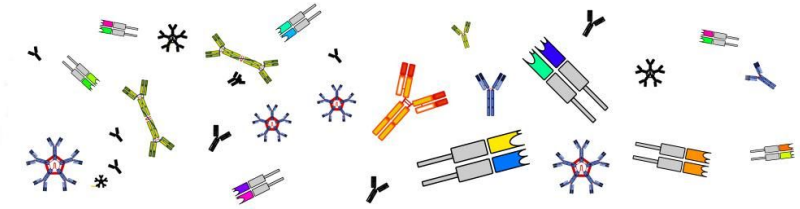
iReceptor Plus Use Case Process

In order to maximize the impact of the project deliverables, the project considers a broad range of user perspectives. We rely heavily on the range of partners within the project whose research focuses around the use of AIRR-seq[1] and related data to create a set of use cases that span a range of user domains. We are fortunate in this project in that we have partners that represent all of the key domains in this area. These domains include:

- scientific research,

[1] Data generated from using next generation sequencing technologies to sequence the Adaptive Immune Receptor Repertoire

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



- biomedical research,
- use within the biopharma sector,
- use within AIRR-seq data processing and analysis service providers to the clinical, hospital, biopharma sectors, and
- and use within infrastructure service providers that provide “cloud” (computation, storage) services to the clinical, hospital, and biopharma sectors.

iReceptor Plus follows an agile software development process, which often resorts to User Stories rather than Use Cases to elicit requirements. User stories consist of short descriptions of what a user intends to do when using the software (e.g.: as a *<type of user>*, I want to *<some goal>*, so that *<some purpose>*).

User stories capture the purpose, which is then further detailed in the scope of informal conversations with the stakeholders. For small-size projects, a user story driven process might be enough, however, in large projects such as iReceptor Plus, user stories may lack the necessary context (a sense of higher goal) and the notion of completeness of all the aspects required to achieve a goal.

For iReceptor Plus, the use case elicitation includes the purpose as a user story would, but also includes other information that helps capture context, assisting the Use Case Team in the process of distilling encompassed functionality in order to aggregate use cases into categories and identifying potential functional gaps.

The following sections list the methodology used to gather our use cases and how we intend to translate those use cases into actionable tasks within the iReceptor Plus work packages.

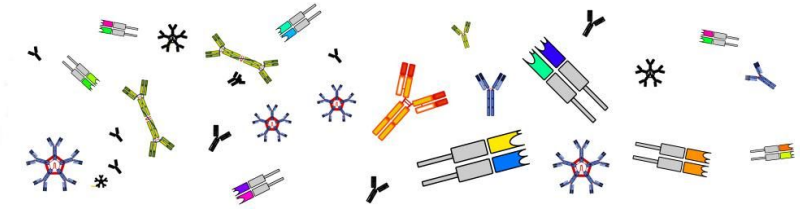
Use case team

The first task in establishing a set of use cases was to form a Use Case Team. This team was initially formed from the software engineering partners, including experts in User Interface and User Experience (UI/UX) design within the consortium (SFU, INESC TEC, and Ascora). The group was formed shortly after the iReceptor Plus kick-off meeting and was tasked with determining the process used to elicit the use cases from our partners. The Use Case Team was then expanded through a self-nomination process with the iReceptor Plus partners to include a wider range of stakeholders, including members with software development experience and scientific domain experience. This group developed a Use Case Elicitation document that would be sent to all partners to gather individual use cases.

Stakeholder engagement

In order to ensure the Use Case Elicitation document would meet the needs of the project, the document was initially presented to and discussed with senior researchers within the project. These researchers were walked through the elicitation process, and feedback was provided to the Use Case Team. Each of the senior researchers was asked to provide a Use Case as part of this process. The Use Case Elicitation document was refined as required.

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



In addition, consultations were carried out with all of the iReceptor Plus partners. EMail was used to present the elicitation plan to WP Leaders, and WP Leaders were asked to distribute the plan within their work packages for feedback. A consortium wide video conference was also held with all partners invited to attend. In this session, the Use Case process and the elicitation document were described to all of the attendees.

Use case elicitation

The consortium wide Use Case meeting was held on March 5th (using video conferencing), with attendance from a broad range of partners. All partners were asked to provide use cases following the protocol set out in the Use Case Elicitation document. Follow up elicitation was conducted via EMail, with WP Leaders asked to ensure that they communicated within their work packages to ensure that coverage across all work packages was sufficient. Partners and WP leaders were reminded on several occasions of the importance of use cases to the project and to ensure that their specific interests and needs from the project were covered by the use cases they provided.

Use case assessment

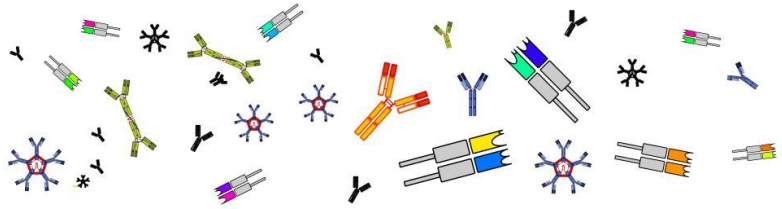
Use cases were assessed as they were provided by the partners. At the end of the first round of elicitation (end of April), 20 use cases were received from 8 iReceptor Plus partners. The Use Case Team used Miro, a digital storyboard platform to capture the critical elements of each use case. This allowed the internationally distributed Use Case Team to work together to understand the use cases and generalize the features that occurred in the use cases.

The Use Case Team then categorized the individual use cases into a small number of high-level functional groups. These categorizations consist of Discovery, Statistics, Analysis, Comparison, and Integration workflows (these categories are detailed below). Finally, the Use Cases were summarized in a Use Case Summary document. This document is grouped by the categories above and presents each use case in a summary paragraph (or two).

Use case feedback

Once the summary document was prepared, the document was sent to all of the iReceptor Plus project partners. WP leaders were asked to disseminate to their WP teams and asked to provide feedback on the document. The iReceptor Plus consortium held its first face-to-face intermediate project meeting on May 15 and 16 in Genoa, Italy. One of the key goals of this meeting was to present the Use Cases to the wider team, get feedback, and discuss the process of translating the Use Cases into actionable tasks against the deliverables for each WP in the project.

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



From use case to task and deliverable

At the time of this writing, the initial round of use cases has been gathered, using the process above. The remainder of this section contains a description of the plan on how the iReceptor Plus consortium will translate those use cases into actionable tasks that meet the deliverables for the project. These actions will be performed both through other tasks within WP1 (D1.3 - D1.7) as well as through tasks in other WPs (see below).

Each WP Leader has been provided with the Use Case summary document as well as the full text of each of the use cases. WP Leaders (and their WP teams) will assess each use case and identify if that use case has an impact on their work package. For those use cases that apply to a work package, the WP Leader will associate one or more tasks within that work package with the use case. This mapping will capture how each use case impacts all of the tasks for each work package. The mapping will initially be captured in the Use Case Summary document, giving the entire project a holistic view of how the use cases span the work packages. This will also provide the use case author with an opportunity to provide feedback on anything that they feel that WP Leaders have missed in the assignment of tasks.

Once use cases have been mapped to tasks, WP Leaders will map the use cases to deliverables (at a high level) within those tasks. Linking use cases directly to deliverables will help drive the generation of requirements for software development tasks.

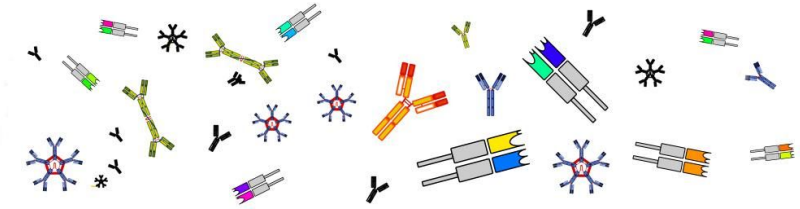
Requirements and development

Requirements, from an Agile software development perspective are "... a service, function or feature that a user needs. Requirements can be functions, constraints, business rules or other elements that must be present to meet the needs of the intended users."[2] We use our use cases to help define requirements to meet the needs of our users. Because of the dynamic nature of the use of AIRR-seq data, we expect user needs to change over time. The agile approach to software development addresses this problem by specifying high-level requirements early on in the project and drilling down to detailed requirements as late in the process as possible.

As a result, we expect WP Leaders to generate high-level requirements for their work packages and to refine those requirements iteratively as a better understanding of the problem and the solution space evolve. It is important for each WP Leader to understand the broad high-level requirements but to only derive detailed requirements when detailed work on those requirements is performed. In this way, we avoid having to have a detailed set of requirements for all tasks and deliverables (an error prone task) while at the same time allowing for the evolution of other requirements as we learn more about the use cases and their impact on the tasks within the work packages.

[2] https://www.agilebusiness.org/content/requirements-and-user-stories

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



As requirements become more detailed, they will evolve into explicit functional and nonfunctional requirements. These requirements will then be translated into development tasks for individuals against components of the iReceptor Plus platform. Care will be taken to have a clear definition of when the requirement has been implemented (a definition of done), including acceptance criteria that can be tested.

Although not a requirement of project members, iterative development, including continuous integration and deployment of changes to a staging platform will be encouraged by the work package teams. Our goal is to be able to deploy changes often, present them to stakeholders and users, get feedback, and iterate on our design and implementation.

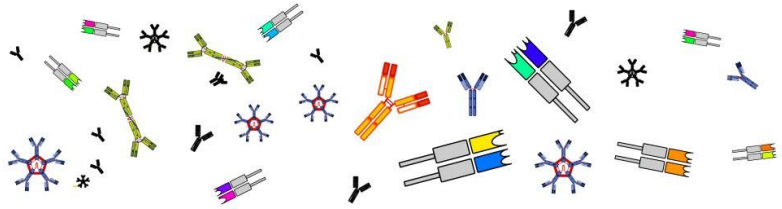
Use Case Overview

What is in a Use Case?

The use case elicitation document utilized to acquire use cases from the iReceptor Plus partners was structured in two parts. The first part explained the purpose of gathering use cases to the partners, followed by setting the context as to why use cases are important to the iReceptor Plus platform. The second part of the use case elicitation document provided a template for users to fill out. This template included the following components:

- Usage scenario name: a short, descriptive name, provided by the user, for their usage scenario.
- Role: a list of roles of the types of individuals that would trigger the scenario (e.g. researcher, clinician).
- Additional roles: a list of additional roles of individuals that might be involved in the usage scenario.
- Purpose: the purpose of the action being taken in the usage scenario. This is a critical component of the use case, and the provider of the use case was asked to provide the purpose statement in a *ROLE performs ACTION for a PURPOSE* format. An example was given.
- Context: the user was asked to provide some “big picture” context so that the use case team could understand the broad context in which the action was taken.
- Description: a more detailed description of the action undertaken in the usage scenario. It was left up to the creator of the use case as to how long this would be, but it was suggested that descriptions of no longer than a page were likely to be necessary at this stage of the use case elicitation process.
- References: use case authors were asked to provide references that would be beneficial to supporting the use case.
- Annexes: use case authors were asked to provide any other documents, diagrams, or pictures as they deemed appropriate.

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



In order to keep the burden low on use case authors, it was suggested that providing a use case should take no longer than 1 to 2 hours.

Who contributed?

We received 20 use cases from 8 iReceptor Plus partners in our initial use case elicitation phase (March 5th - April 30). We received use cases from research, clinical, and industry partners, with the bulk of our responses from our research partners. Two of our clinical partners provided use cases that were oriented towards research in a clinical setting and two of our industry partners provided use cases that were oriented towards an industry setting. Although our use case landscape is relatively rich, it will be important in moving forward to follow up with and gather more clinical and industry use cases. We have been in touch with partners in these areas and we do expect such use cases in the near future.

What was contributed?

Different partners responded to the use case request in different manners. Some partners provided a single use case. Other partners distributed the use case widely internally to their organization, which resulted in a relatively large number (up to 8) of varied use cases from a single partner. Some partners responded with complex use cases (there was more than one *ACTION* and *PURPOSE* in the use case). In these instances, the large, complex use cases were split up into several simple use cases.

All use cases that were gathered contained enough information to provide a basic understanding of the use case. In many cases, follow up will be required to ensure that the use case team understands each use case correctly. Those use cases that require follow up will be identified when the WP Leaders assess each use case and determine if it applies to their work package tasks. At that time, further consultation with the partner that produces the use case will be under carried out.

Who are the actors?

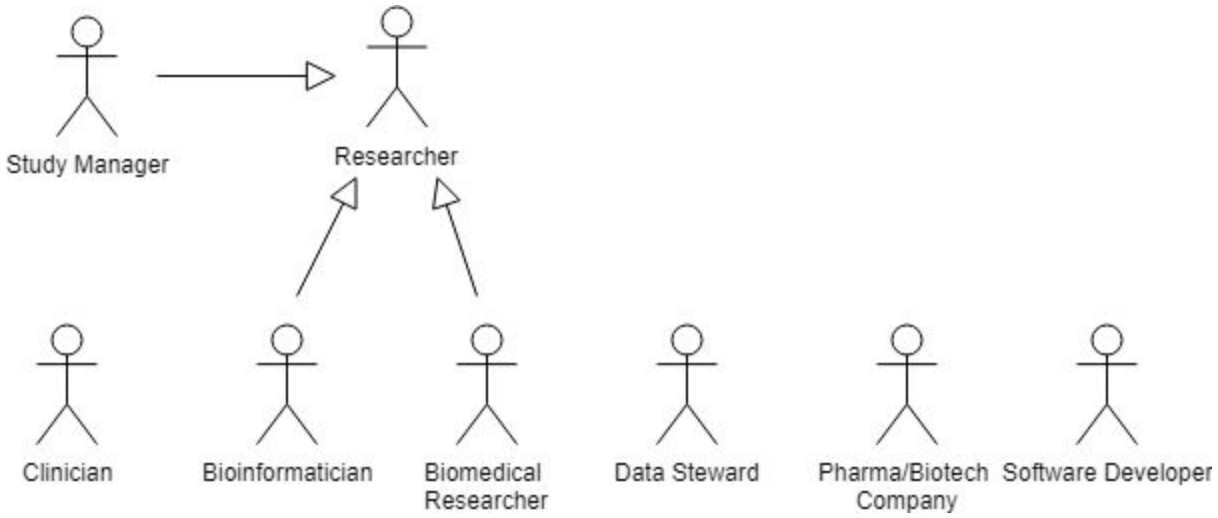
The actors that are identified through use case elicitation are critical to the process, as they represent the users of the iReceptor Plus platform. Our use cases identify the following actors as potential users of the platform.

- Scientific researcher
- Biomedical researcher
- Clinician
- Bioinformatician/computational immunologist
- Software developer
- Study manager
- Data steward
- Biotech/Pharma company

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

An actor's diagram depicting the identified user roles and relationships between them is provided below. An arrow pointing to another actor means that the latest is a generalization of the former. From a functional perspective, the more specific actor will inherit all the use cases from the generic one and has additional ones of its own.

Figure 1: iReceptor Plus actors diagram



Use case categorization

Once the use cases were tabulated, it became quite clear that each use case fit in to one or more use case categories. The use case team utilized Miro (https://miro.com) to allow the internationally distributed team to categorize the use cases, coming up with a relatively simple, but critical basic set of use case categories.

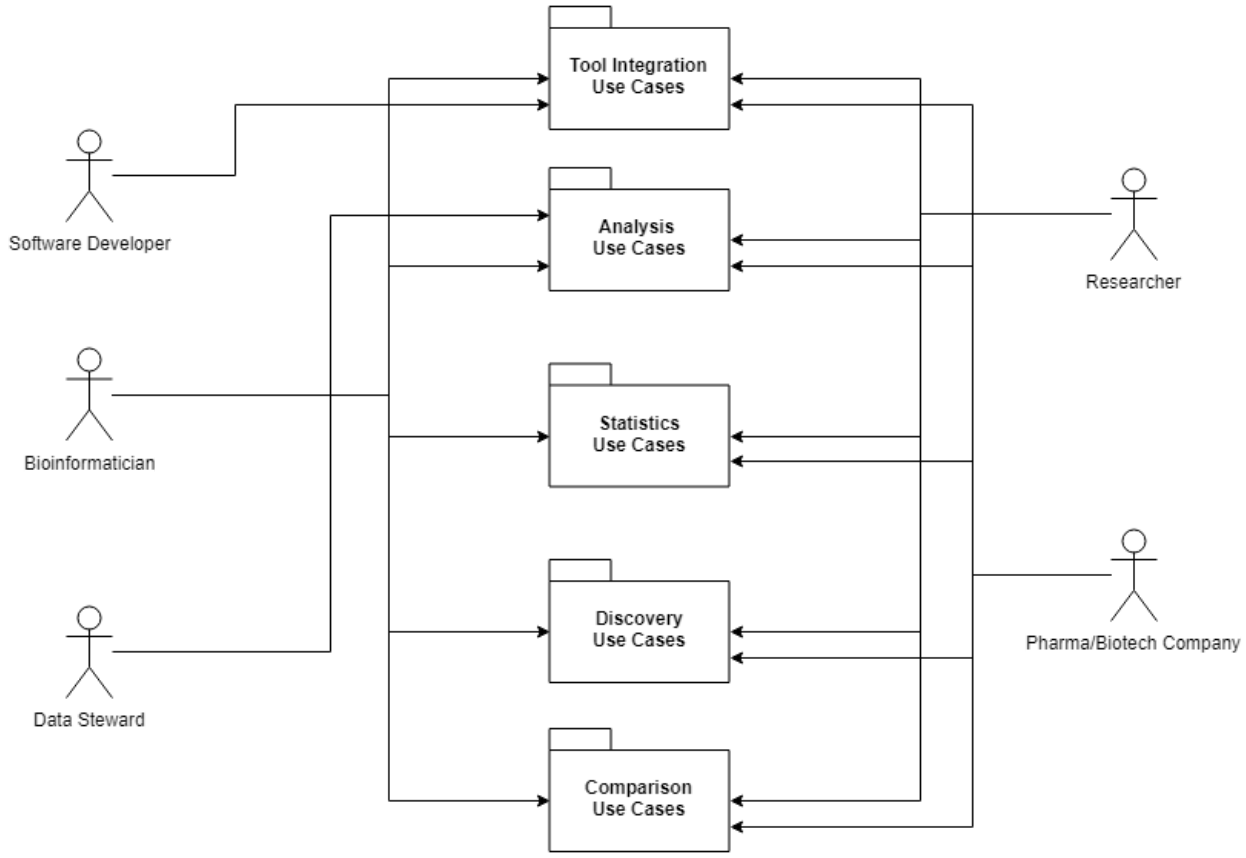
The categories included:

- Discovery: Use cases that involved finding AIRR-seq data that met specific search criteria.
- Statistics: Use cases that involved performing statistical analysis of AIRR-seq data at either the repertoire metadata or the rearrangement level. These statistical analyses are often performed on data that is found in Discovery use cases.
- Analysis: Use cases that involved performing complex analyses on AIRR-seq data. Analyses are differentiated from Statistics in that they are either computationally complex (and therefore take time to perform) or produce ancillary data that needs to be maintained and used in downstream analyses.

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

- Comparison: Use cases that involve comparing two or more AIRR-seq data sets. This may involve the comparison of relatively simple statistics about the data sets in question through to extremely sophisticated and computationally expensive comparative analyses.
- Tool Integration: Use cases where users want to integrate analysis tools into the iReceptor Plus platform to expand on the analysis capabilities of the platform.

Figure 2: Use case categories (Discovery, Statistics, Analysis, Comparison, and Tool Integration)



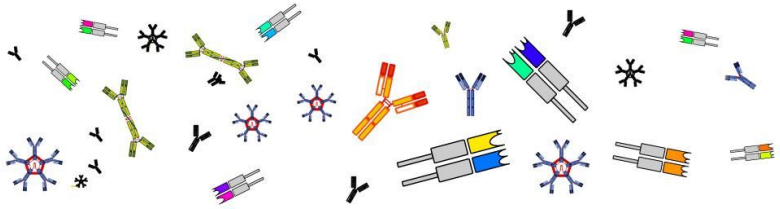
Use case overview

The table below provides a list of the use cases, divided by category, that have been acquired thus far. The purpose statement (ROLE performs ACTION for PURPOSE) for each use case is provided.

Table 1: Use case purposes

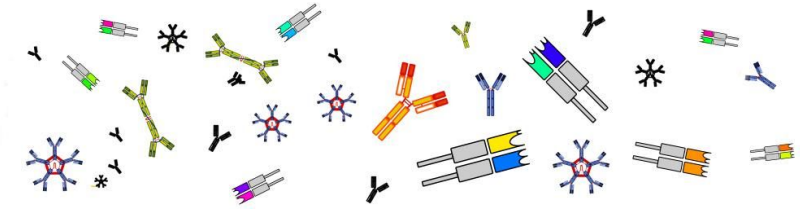
Discovery Use Cases
DI-01: As a researcher/clinician/etc who is unaware of the blessings of AIRR-seq I want to use a meta-search engine to discover AIRR-seq data sets within iR+.

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



DI-02: As a researcher I want to find data sets with similar subject-sample-processing profiles (i.e. metadata) within the iR+ framework to perform a case-control study against datasets of my own.
Statistics Use Cases
SI-01: As a researcher studying the antigen specificity of B cells* in the context of an infectious disease** I would like to query for paired chains, retrieve paired chain data and perform basic statistics on chain pairing (e.g. enrichment or depletion of specific combinations vs. a neutral model). Alternatives: * T cells; ** malignancy / autoimmune condition
Analysis Use Cases
AN-01: For a given filtration of the data according to the associated metadata (e.g., all samples coming from PBMCs of MS patients using UMI library preparation protocol), a researcher would like to perform a series of exploratory data analyses (EDA). The idea is to have quick analyses that will return the user graphs and tables and not force him/her to download all the data and run these analyses locally.
AN-02: Researchers want to perform a non-exact (fuzzy) search for a given sequence feature (e.g. CDR3) and find all of the AIRR-seq data (and its associated MiAIRR metadata) that meets that search criteria.
AN-03: As a data steward responsible for the content of an institutional iR+ repository, I want to clone public data sets from other iR+ repositories, to avoid bottlenecks in computational resources, achieve load balancing and keep queries that contain confidential information "in house".
AN-04: As a researcher I would like to have a simple way to perform allele inference on datasets present in iR+ to be able to create aggregate statistics about allele frequencies.
AN-05: As a researcher developing machine learning algorithms, I want to find AIRR-seq data sets with specific characteristics, so that I can train/test/validate my machine learning algorithm with the AIRR-seq data sets.
AN-06: As a computational immunologist, I would like to leverage the data stored in the ireceptor database to perform large-scale machine learning driven pattern detection in the effort to identify patterns that predict immune status.
Comparison Use Cases
CO-01: Researchers want to find the overlap between two data sets (typically partitioned at the “sample” metadata level). The overlap is determined by searching AIRR-seq features by either comparing identical sequence features (e.g. CDR3) at the nucleotide or amino acid level or by fuzzy matches that allow some variations in the sequence feature comparison
CO-02: As a researcher studying autoimmune and inflammatory diseases, I want to find AIRR-seq data sets from other similar studies with a focus on TCR data, so that I can compare features in my data to features from other AIDs AIRR-seq data sets.

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



CO-03: We would like to understand how unique these BCRs are to this gut disorder, i.e. whether they are found in other AIRR-seq data sets. This analysis will help the understanding of the pathogenesis of celiac disease, and ultimately this endeavor has the potential to develop diagnostic tests based on BCR repertoire interrogation.
CO-04: As a researcher studying the TCR repertoire in different vaccination settings I would like to compare the “bystander” (i.e. not antigen-specific) T helper cell response of immunized mice in the context of immunizations to understand the effect of various adjuvant systems.
CO-05: Given a researcher or company possesses an Ig sequencing dataset from an immunization trial, it might be useful to know which antibody possesses features (in this case mutations) that are not common in other datasets
CO-06: As a researcher, I would like to find the same (or similar) antibodies in different datasets to compare antibody reactivity.
CO-07: As a researcher using VDJServer to analyze my data, I want to find AIRR-seq data sets from other similar studies, so that I can perform comparative analysis between my data and the queried AIRR-seq data.
CO-08: As a clinical stage drug development company (pharma/biotech) will use the TCR repertoire diversity to know the T cell exhaustion of a given patient or cohort to then use this TCR diversity level as a surrogate marker for cancer immunotherapy enabling patient stratification, monitor drug efficacy or patient prognosis.
CO-09: A researcher would like to compare their own TCR data with iReceptor+ data from similar studies. The purpose could be to analyze clinical prognosis, diagnosis, and treatment of various cancers.
CO-10: As a cancer research center or drug discovery company looking for drug candidates will use comparisons of the TCR repertoire diversity to identify T cell exhaustion and reinvigorate immune responses.
Tool Integration Use Cases
IN-01: As a bioinformatician developing new analysis tools, I want to acquire AIRR-seq data sets, so that I can implement my method, test it on a variety of different data sets, and deploy it to the iReceptor+ platform so other users can use the analysis method.

Use Case Category Descriptions

Building on the use cases above we can begin to understand several key workflows that users want to perform, and it is these workflows that are captured in our use case categorizations.

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

Discovery use case category

The most simple use case category is that of data discovery. Data discovery in this instance goes beyond finding a data set that is associated with a specific paper, which can be done today. Data discovery in the case of the iReceptor Plus platform implies being able to find data sets by performing searches over the rich metadata at the study, subject, disease, and sample processing levels. In addition, it may be of interest for the user to search the sequence rearrangement level of AIRR-seq data to further refine this search. This use case category is captured succinctly in use case D1-02.

DI-02: As a researcher I want to find data sets with similar subject-sample-processing profiles (i.e. metadata) within the iR+ framework to perform a case-control study against datasets of my own.

Because of the richness of the MiAIRR metadata available, this can be a complex and powerful search. The output of this interaction would typically be a report on the data that was found that met the user's search criteria. A conceptual interaction/sequence diagram is given below.

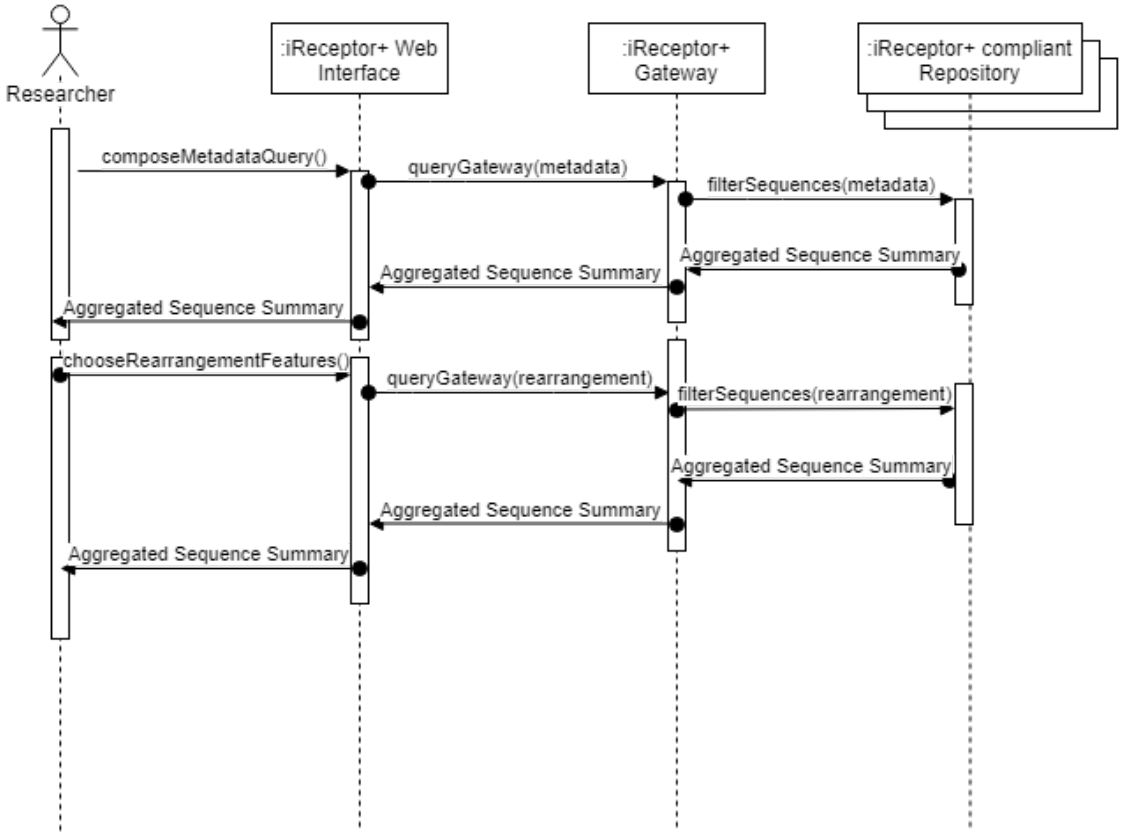


Figure 3: Discovery use case interaction/sequence diagram

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

Note that in the interaction/sequence diagram in this section, the blue filled boxes can be considered entities that the user can reproduce, store, or select for later interaction. For example, in the diagram above, the “Rearrangement Set” is the result of a query and is something that the user can either store in its entirety or store as a computational description such that the entity can be utilized or reproduced at a later date.

Statistics use case category

In terms of complexity, the next use case category moves from finding AIRR-seq data of interest to performing basic statistical analysis of features within that data. This is the Statistical use case category. This use case category is an extension of the Discovery use case category in that a user would typically go through a discovery process to find data of interest before performing a statistical analysis of features within that data set.

SI-01: As a researcher studying the antigen specificity of B cells in the context of an infectious disease** I would like to query for paired chains, retrieve paired chain data and perform basic statistics on chain pairing (e.g. enrichment or depletion of specific combinations vs. a neutral model). Alternatives: * T cells; ** malignancy / autoimmune condition*

In this case the researcher is looking for disease specific data sets and has a need to perform basic statistics on features (chain pairing) within that data. This use case category is captured in the interaction/sequence diagram below.

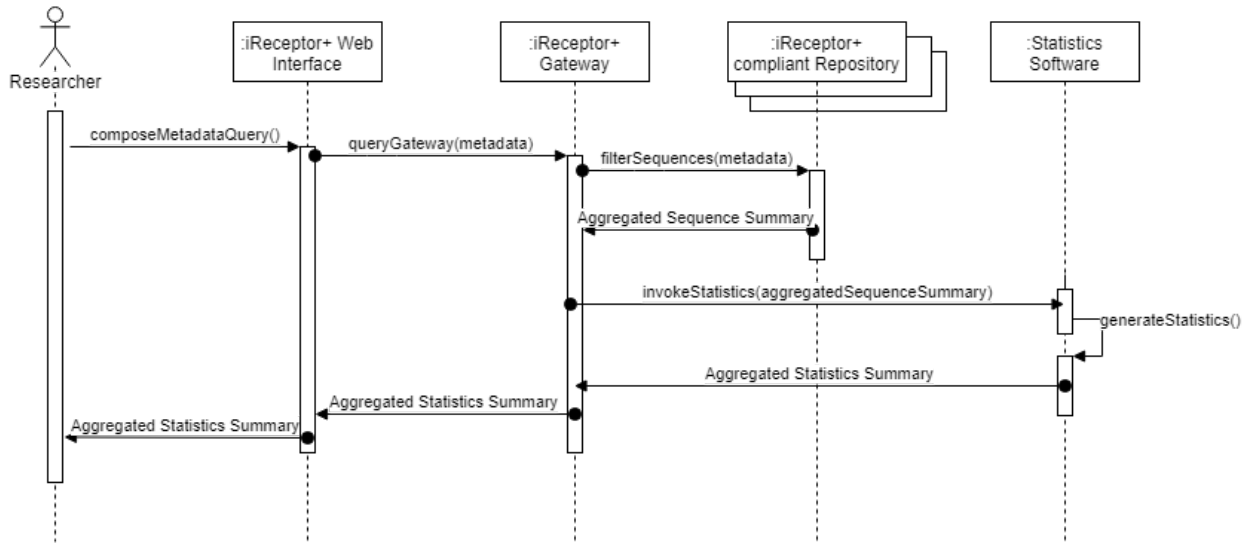
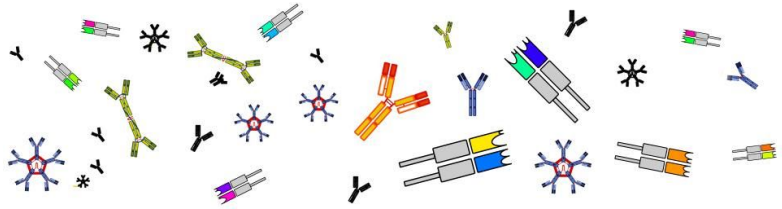


Figure 4: Statistics use case interaction/sequence diagram

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



Analysis use case category

With data as rich and complex as AIRR-seq data and its accompanying metadata, simple statistics rapidly move to performing complex analyses. Recall from our earlier description of the Statistics and Analysis categories, "*Analyses are differentiated from Statistics in that they are either computationally complex (and therefore take time to perform) or produce ancillary data that needs to be maintained and used in downstream analyses.*" As such, the workflow and the user/system interaction is conceptually similar, but the analysis, and in particular on which computational system that analysis takes place, may be quite different. This is the driver for separating these two use cases and creating an Analysis category.

AN-02: Researchers want to perform a non-exact (fuzzy) search for a given sequence feature (e.g. CDR3) and find all of the AIRR-seq data (and its associated MiAIRR metadata) that meets that search criteria.

In the use case above, performing a "fuzzy" search at the sequence rearrangement level, where there are billions of records, implies that the computational complexity of this search may be significant. In this case, it may not make sense for the repository to perform the analysis. In the diagram below, it may be necessary to move the data that was discovered through the discovery process (the Rearrangement Set) from the system on which the data discovery took place to a different computational system that can perform a more complex analysis. In addition, because the results of these analyses are "valuable" in that they are costly to compute, it might be desirable to have the platform store the results of these analyses (or at least how the analyses were performed). One of the key questions that we will attempt to answer within the project, through WP4: Analysis Pipelines, is which analyses can be performed by the repository and which need to be performed on an "Analysis Engine". Another key question to be answered by WP4 is how to integrate such analysis engines into the iReceptor Plus platform and if, where, and how to store the analysis results for further exploration. An interaction/sequence diagram that captures this is given below.

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

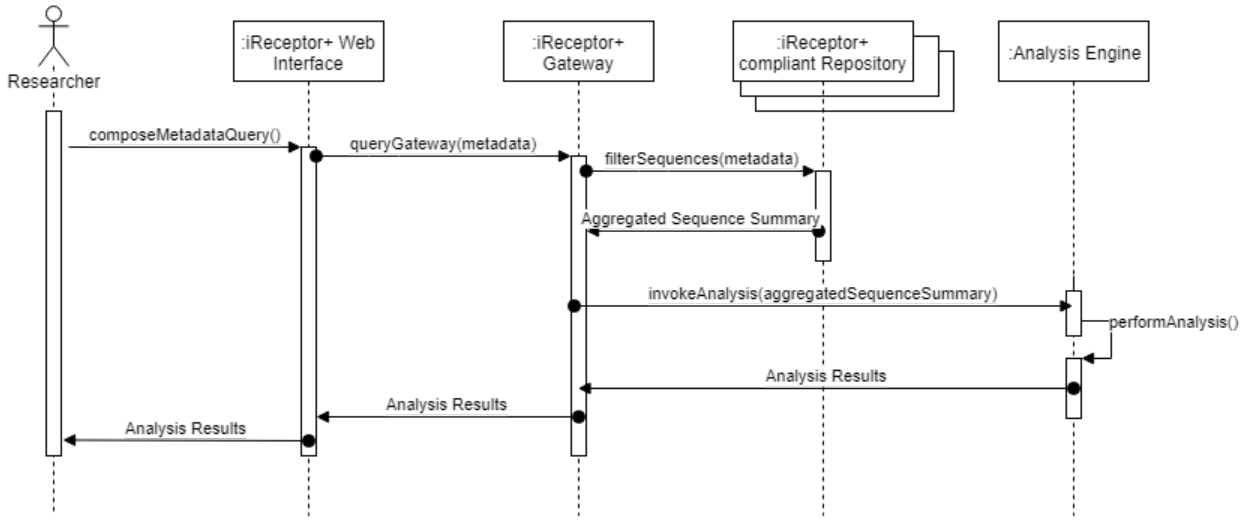


Figure 5: Analysis use case interaction/sequence diagram

Comparison use case category

The most complex discovery and analysis use category is that of the Comparison category. This use case category is similar to the Analysis category with the exception that the discovery process identifies two different “Rearrangement Sets” of interest and the analysis is a more complex comparative analysis across or between the two data sets of interest.

CO-01: Researchers want to find the overlap between two data sets (typically partitioned at the “sample” metadata level). The overlap is determined by searching AIRR-seq features by either comparing identical sequence features (e.g. CDR3) at the nucleotide or amino acid level or by fuzzy matches that allow some variations in the sequence feature comparison.

All of the questions that apply to the Analysis use case category also apply to this use case category, in terms of data movement between systems and the computational complexity of the algorithms being used. An interaction/sequence diagram for this use case category is given below.

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

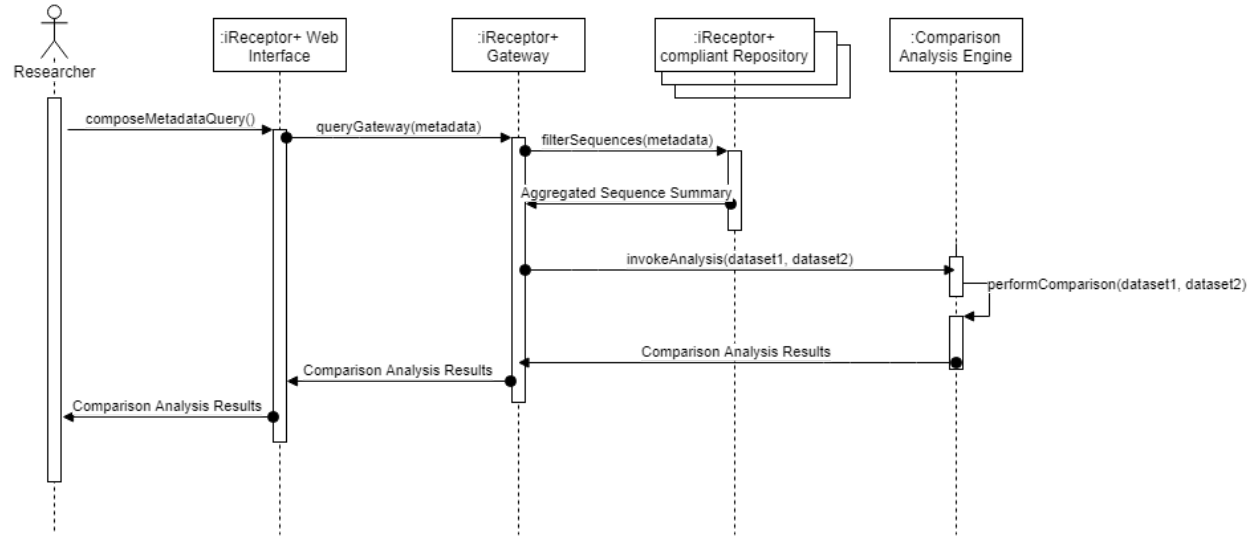


Figure 6: Comparison use case interaction/sequence diagram.

Tool Integration use case category

The final use case category, although very different from those discussed above, is critical to the success of the platform. Each of the above use cases either performs statistics or analyses on AIRR-seq data. Analysis tools in this domain are changing rapidly, and one of the challenges faced by the iReceptor Plus platform is to make the use of such advanced analysis tools easy for the end user. This is one of the key tasks that exist for both WP1 and WP4 and is captured in the Tool Integration use case category.

IN-01: As a bioinformatician developing new analysis tools, I want to acquire AIRR-seq data sets, so that I can implement my method, test it on a variety of different data sets, and deploy it to the iReceptor+ platform so other users can use the analysis method.

This use case captures both the need to use the platform to develop new tools, but more importantly once new tools are developed to make them accessible to the AIRR-seq community through integrating them into the platform so that they can be used. Within the scope of the iReceptor Plus platform, a general solution to this problem is not within the scope of the tasks and deliverables for WP4 (see WP4, D4.1 report). Instead, WP4 will identify a set of important analysis tools for integrating into the iReceptor Plus platform, with the goal of making this integration extensible to general tools possible in the future. It is anticipated that there will be a manual process for installing and using new analysis tools into the iReceptor Plus platform, as shown in the activity diagram below.

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

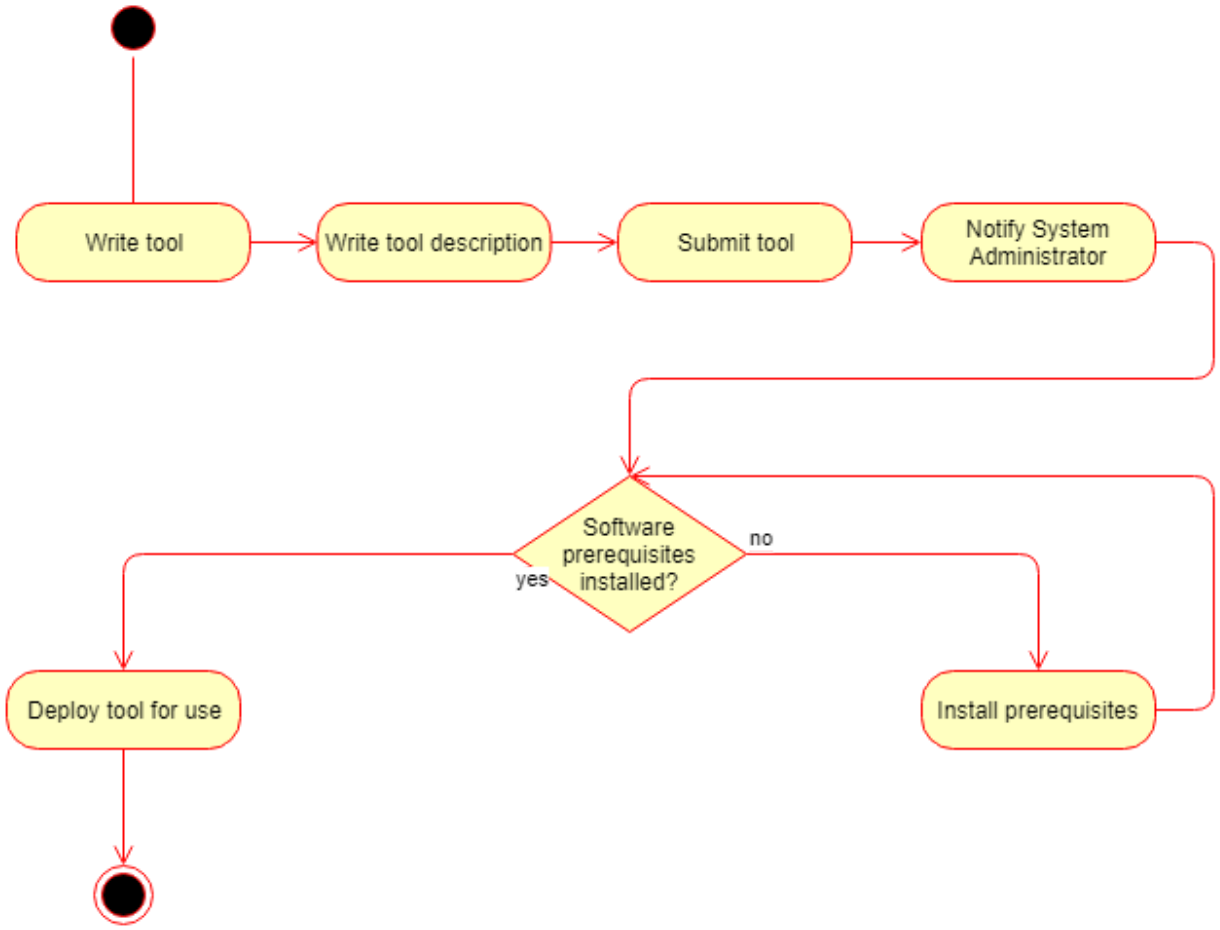


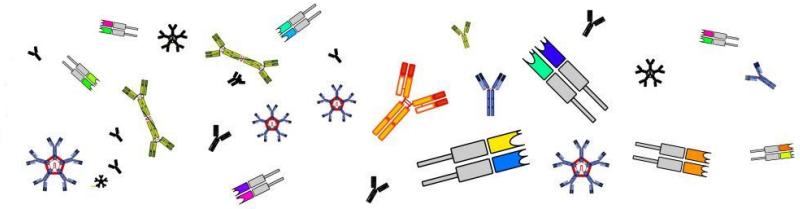
Figure 7: Tool Integration activity diagram

Conclusions

This document provides both an overview of the process used to elicit use cases from the iReceptor Plus partners as well as a description of the use cases that have been gathered to date. We provide a complete list of use cases and their purpose, as well as a more detailed description of these use cases based on a higher-level categorization of the specific use cases. These use cases, and in particular the use case categorizations, are fundamental to the development of the iReceptor Plus platform and will guide the research and development that is performed across all of the work packages in the project.

As stated in the Executive Summary, user centered design is a dynamic and evolving process. This continual evolution is captured in WP1, Task 1.2 - “Use case development and end-user feedback”. Although we have been able to gather a significant number of use cases from our partners, we expect these use cases to evolve. In addition, we expect to acquire more use cases as the project proceeds, in particular as we see new users using the platform. These use cases will be iteratively updated through user feedback as the iReceptor Plus platform develops and user needs change based on those

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



developments. The outputs of this task will be communicated to other WP Leaders to ensure that user needs are reflected across the entire platform. This task will be carried out throughout the project to ensure that the user community's requirements are met at all stages of the project and to facilitate the uptake of deliverables across the AIRR-seq research community.

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