

Guidelines for IMMUcan data upload to external database

Version	Date	Summary of changes
1.0	13 January 2025	1 st release

EORTC mandates the sharing of data, generated within EORTC clinical trials and research projects (see EORTC Policy on [data sharing](#)). Therefore, raw and processed data generated into IMMUcan, should be uploaded to public repositories with controlled access (IMMUcan specific policy available on the website: [IMMUcan Policy Results and Human Samples v1.pdf](#))

The list of currently approved repositories can be found in the sections below (2.1 for genomics data and 3.1 for imaging data). Uploaded metadata should be anonymized or pseudonymized prior to the upload to ensure patients' privacy and comply with EORTC Data sharing policies and GDPR (POL008 and POL002).

The upload of data should be done by the party generating the data (e.g. consortium partners or external collaborators). In this document, data were divided in 3 categories: clinical data, OMICs data, Imaging data.

1 Upload of Clinical data

All data related to the medical history of the patient (e.g. demographics, disease, treatments..) entered in (electronic) Case Report Forms are defined as clinical data.

Individual and pseudonymized clinical data sets as available on the KMS (knowledge management system, IMMUcan database) should NOT be uploaded to any public data repository by any consortium members or any external collaborators other than the EORTC.

Such data are indeed subject to strict regulatory and data privacy rules and laws. The EORTC acts as a controller of such data collected for the purpose of the IMMUcan project and will therefore control all access requests received from 3rd parties, following the EORTC data sharing policy and the IMMUcan specific policy (links above).

2 Upload of -OMICs data

2.1 OMICs data

OMICs data refer to datasets generated from human biological material and related to the identification, characterization and/or quantification of various sets of biological molecules (i.e. genomics, transcriptomics, proteomics, metabolomics, lipidomics,...).

2.2 Current EORTC approved repositories

All repositories must be compliant with all regulations for personal data protection and must provide control over the data access. For all EORTC studies, omics data access must be regulated by EORTC DAC committee.

Current list of IMMUcan approved repositories for omics data:

Name	Link	Request new account
European Genome-Phenome Archive (EGA)	https://ega-archive.org/	https://ega-archive.org/submission-form.php

If the chosen repository is not yet in the approved list (see above), a separate request by the repository chooser should be submitted to EORTC by sending an email to EORTC TR group (tr@eortc.org) and the IMMUCan general email address 1828@eortc.org.

EORTC must review the proposed repository and, if it complies with data protection standards, the repository will be added to the list prior to data upload.

The requestor must provide all necessary support to facilitate EORTC's review of the proposed repository. This includes, but is not limited to, providing EORTC with access to the repository assessment and assisting in any further investigations EORTC may require regarding the repository.

2.3 Information requested prior the upload of data

To upload data on an external repository, on behalf of IMMUCan, please follow the following steps.

Prior to upload, please contact 1828@eortc.org, with:

- "1828 IMMUCan" as reference
- WP number and deliverable (*if applicable*)
- Name of the lab and head of the lab that generated the data
- Type of data (technique used and file format)
- Repository name and link
- Confirmation that you have read the following policies and that you will comply with them:
[IMMUCan Policy Results and Human Samples v1.pdf](#)
[IMMUCan-Publication-Policy v1.pdf](#)

EORTC will send the authorization for the upload of the data by the collaborators as reply to the email containing all the information and confirmation listed above.

2.4 Steps to upload/submit data on a repository

The following steps should be fulfilled prior or during data upload/submission:

- All omics metadata should be anonymized or pseudonymized prior to the upload to ensure patients' privacy and comply with GDPR. This includes the sample names appearing in BAM or VCF headers.
- EORTC encourages upload of all types of files (FASTQ, BAM, VCF) under the same submission but as separate datasets. Similarly, the recommendation is to submit the different cohorts under the same submission but as separate datasets. This will allow requests for only one type of files, e.g. BAM only.
- All files (fastq, bam and vcf) should be named with the IMMUCan sample identifiers (e.g.: IMMUN-NSCLC-0075-FIXT-01-DNA-01) to facilitate their association with other data types (imaging and clinical, for example).

When the upload is finalized and the dataset is released, following information should be provided to 1828@eortc.org:

- “1828 IMMUcan” as reference Datasets ID(s) and link(s) (required) If uploaded BAM files have been generated with a different pipeline than the standard IMMUcan-Vegan or IMMUcan-STAR pipeline,
- please specify following: (optional but encouraged)
 - o Which aligner was used and its version
 - o If duplicated reads are marked/removed
 - o If file is sorted
- If uploaded VCF files have been generated with different pipeline than the standard IMMUcan-Vegan,
- please specify following: (optional but encouraged)
 - o Variant caller and version
 - o If annotated and using which databases
 - o If variant filtering was performed, specify most important steps.
- If the data analysis is published, please provide the title and doi.

2.5 Use of EGA as repository

IF EGA is used as repository, the following steps should be followed:

1. If collaborator doesn't have **an account on EGA** (EGA box), they should apply for a new account/box by filling in this EGA [submission form](#).
2. **Following DUOs should be added** when releasing study datasets:

DUO:0000007	disease-specific research and clinical care	This primary category consent code indicates that use is allowed provided it is related to the specified disease.
MONDO:0045054	cancer related condition	DUO:0000007 requires MONDO ID
DUO:0000019	publication required	This requirement indicates that requestor agrees to make results of studies using the data available to the larger scientific community.
DUO:0000021	ethics approval required	This requirement indicates that the requestor must provide documentation of local IRB/ERB approval.
DUO:0000026	user-specific restriction	This requirement indicates that use is limited to use by approved users.
DUO:0000027	project-specific restriction	This requirement indicates that use is limited to use within an approved project.
DUO:0000028	institution-specific restriction	This requirement indicates that use is limited to use within an approved institution.

3. A DAC (Data Access Committee) must be defined

For each upload into EGA, a DAC (data access committee) and associated sharing policy(ies) should be appointed.

The role of the DAC is to assess who can have access to the uploaded data.

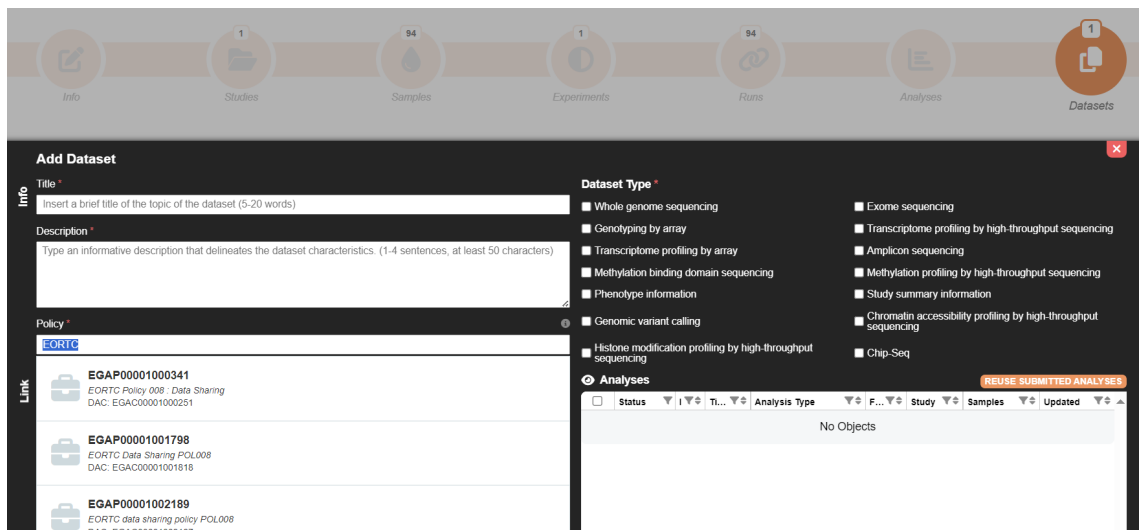
For IMMUcan related uploads, the DAC that will be used, will be an IMMUcan specific DAC constituted by IMMUcan consortium representatives. Representatives of this DAC are the following:

Name	Email	Telephone	Organization	Main contact
IMMUcan representatives	1828@eortc.org pct@immucan.eu	+32 2 7741570	EORTC HQ	Yes

The policies associated with the IMMUcan DAC are:

- EORTC Policy on [data sharing](#)
- [IMMUcan Policy Results and Human Samples v1.pdf](#)

During the submission, the DAC will have to be inserted in the dataset folder of the submission by selecting the associated policies (see example in picture below).



3 Upload of images data

3.1 Images data

The term “images” refers to tissue derived slides which have been stained and further scanned, as described in the IMMUcan project.

3.2 Current EORTC approved repositories

All repositories must be compliant with all regulations for personal data protection and must provide control over the data access.

Current list of EORTC approved repositories for imaging data:

Name	Link	Request new account
zenodo	https://zenodo.org/	https://zenodo.org/signup/

If the chosen repository is not yet in the approved list (see above), a separate request by the repository chooser should be submitted to EORTC by sending an email to EORTC TR Unit (tr@eortc.org) and the IMMUcan general email address 1828@eortc.org.

EORTC must review the proposed repository and, if it complies with data protection standards, the repository will be added to the list prior to data upload.

The requestor must provide all necessary support to facilitate EORTC's review of the proposed repository. This includes, but is not limited to, providing EORTC with access to the repository assessment and assisting in any further investigations EORTC may require regarding the repository.

3.3 Information requested prior the upload of data

EORTC mandates the sharing of images data generated in the scope of EORTC clinical trials and research projects. To facilitate this, EORTC recommends the use of public database with controlled access described above.

To upload data on an external repository, on behalf of EORTC, please follow the following steps.

If data is produced, temporary stored or analysed by an investigator or a collaborator, they should upload the data to a public database with controlled access approved by EORTC and described in section 1.

After selecting the sharing platform, please contact 1828@eortc.org with:

- "1828 IMMUcan" as reference
- WP number and deliverable (*if applicable*)
- Name of the lab and head of the lab that generated the data
- Type of data (technique used and file format)
- Repository name and link
- Confirmation that you have read the following policies and that you will comply with them:
[IMMUcan Policy Results and Human Samples v1.pdf](#)
[IMMUcan-Publication-Policy v1.pdf](#)

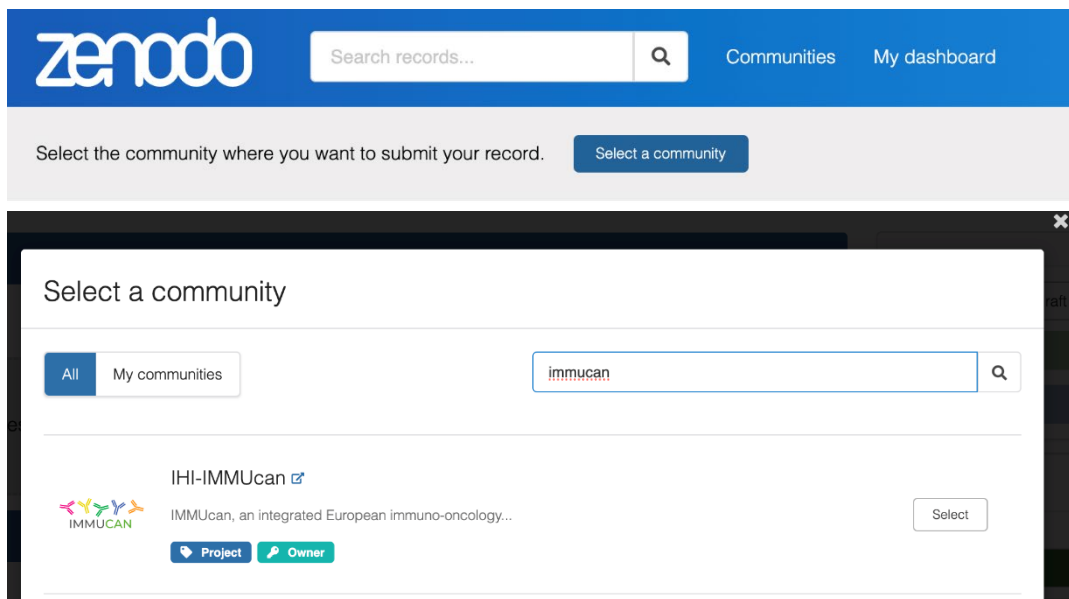
EORTC will send the authorization for the upload of the data by the collaborators as reply to the email containing all the information and confirmation listed above.

3.4 Images Data upload/submission on Zenodo

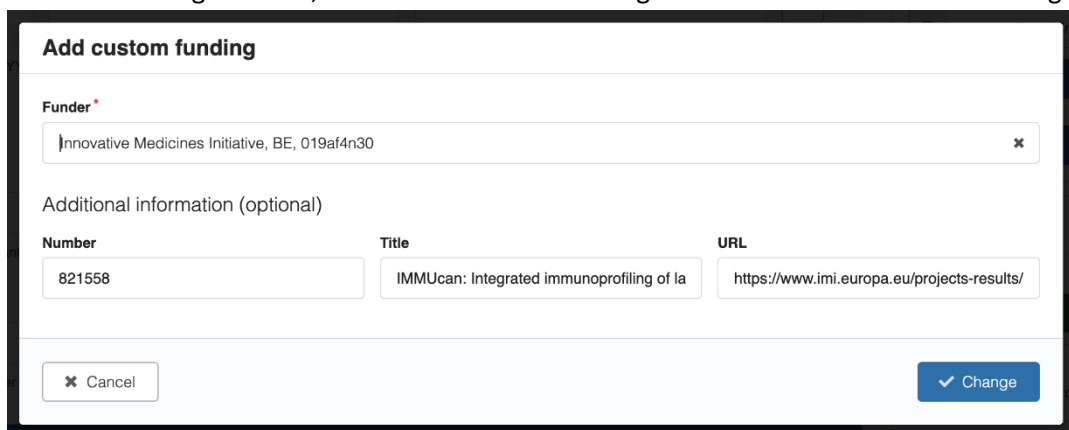
The following should be fulfilled prior or during data upload/submission:

- All metadata should be anonymized or pseudonymized prior to the upload to ensure patients' privacy and comply with GDPR.

- In case a repository not mentioned in 3.2 is used, EORTC DAC and data sharing policy (or similar) should be created and used in all submissions and datasets that contain EORTC data
- All files (tiff, ndpi, ..) should be named with the IMMUcan sample identifiers (e.g.: IMMU-NSCLC-0075-FIXT-01-HES-01, IMMU-NSCLC-0075-FIXT-01-IF1-02) to facilitate their association with other data types (molecular and clinical for example).
- Zenodo can be used to publish a ‘small’ set of images (<50GB).
- An IMMUcan community has been created in Zenodo. Ensure to associate your dataset to this community:



- In the Funding section, add a “custom funding” to mention the IMMUcan grant:



When the upload is finalized and the dataset is released, the following information should be provided to EORTC using 1828@eortc.org :

- “1828 IMMUcan” as reference (required)
- Datasets ID(s) and link(s) (required)

- If the data analysis is published, please provide the title and doi

3.5 Images Data upload/submission on The Image Data Resource (IDR)

Instructions on how to prepare an upload to IDR can be found at:
<https://idr.openmicroscopy.org/about/experiments.html>

4 Contact

For any further information, please contact the IMMUcan team by sending an email to:
1828@eortc.org