# Workshop Series
# Strategies to overcome your challenges in multi-omics data integration

**Anna Niehues**
*Radboud University Medical Center, Nijmegen*

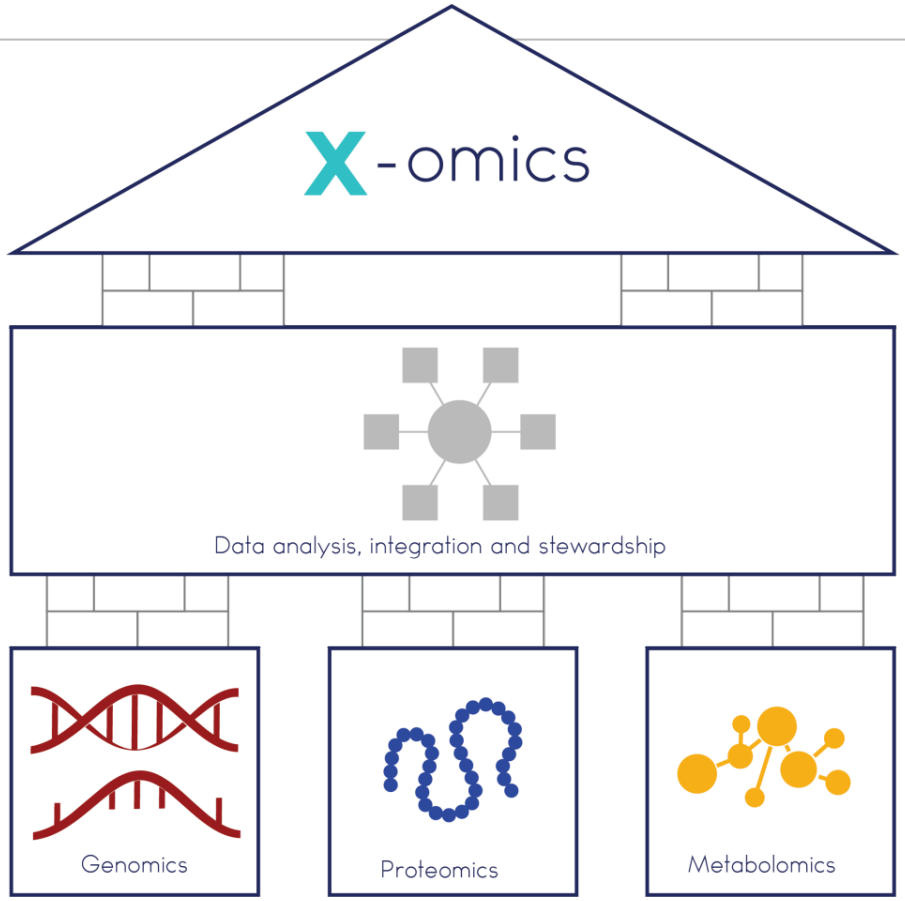
## Workshop series organizing committee

Jasmin Böhmer *(UMC Utrecht, CMM)*
Jenny van Dongen (*Vrije Universiteit Amsterdam*)
Victor Guryev *(UMCG, Groningen)*
Yanick Paco Hagemeijer *(University of Groningen)*
Peter-Bram 't Hoen *(Radboudumc, Nijmegen)*
Peter Horvatovich *(University of Groningen)*
Purva Kulkarni *(Radboudumc, Nijmegen)*
Anna Niehues *(Radboudumc, Nijmegen)*
Gurnoor Singh *(Radboudumc, Nijmegen)*

Daniella Kasteel
Jessie Smits
Jeroen Ewals
(*X-omics project management team*)

# Workshop Series

# Strategies to overcome your challenges in multi-omics data integration




**X**-omics research infrastructure

**Outreach**

- Helpdesk
- **Training**
- Community
- Demonstrators
  - Cell
  - Individual
  - Population

***X-omics approach***

- Data analysis
- Data integration
- Data stewardship
- Study design
- Sample handling

***Pushing omics technologies***

- Resolution, sensitivity, coverage
- Harmonisation, standardisation, data FAIR-at-source

# Workshop Series
# **Strategies to overcome your challenges in multi-omics data integration**

18th June - Data standards and multi-omics data integration

22nd June - Linked data in practice

25th June - Showcases of multi-omics data integration

30th June - Pitch your own multi-omics project

Visit and register: https://www.x-omics.nl/training-outreach/see-all-events

# Workshop Series

# Strategies to overcome your challenges in multi-omics data integration

**> 90 registered attendees**

**Expertise**

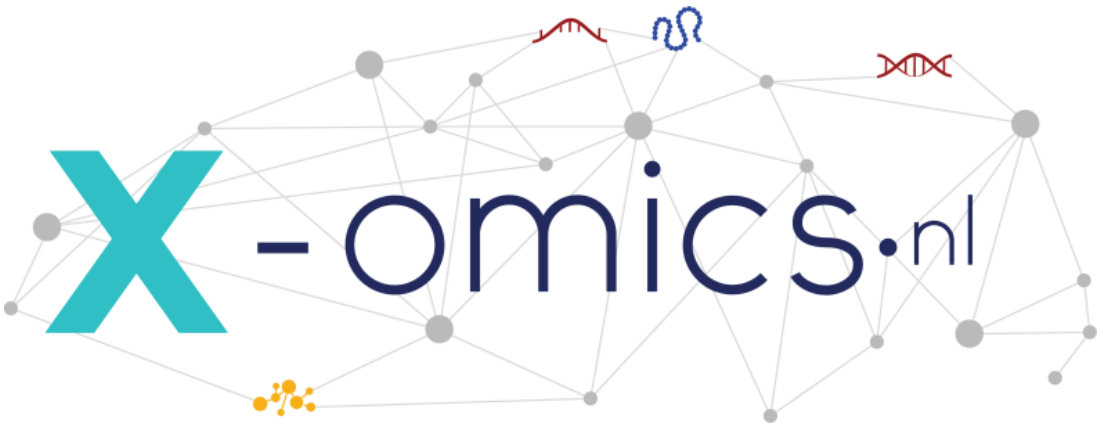
- Genomics
- Proteomics
- Metabolomics
- Data analysis, integration and stewardship
- Other

Workshop Series

# Strategies to overcome your challenges in multi-omics data integration

Thursday 18th June 2020

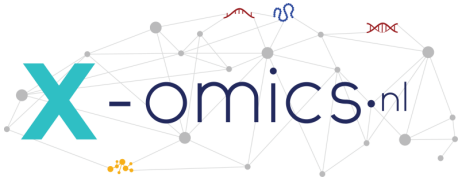
**Data standards and multi-omics data integration**



# Data standards and multi-omics data integration

| | |
|---|---|
| **10.00** | Opening and intro by Anna Niehues, Radboud UMC |
| **10.05** | Keynote by Juan Antonio Vizcaino, EMBL-EBI |
| **11.00** | FAIR genomes showcase by Joeri van der Velde, UMCU Groningen |
| **11.15** | EGA data standard showcase by Jasmin Böhmer, UMCU Utrecht |
| **11.30** | Interactive Quiz |
| **11.45** | Q&A |

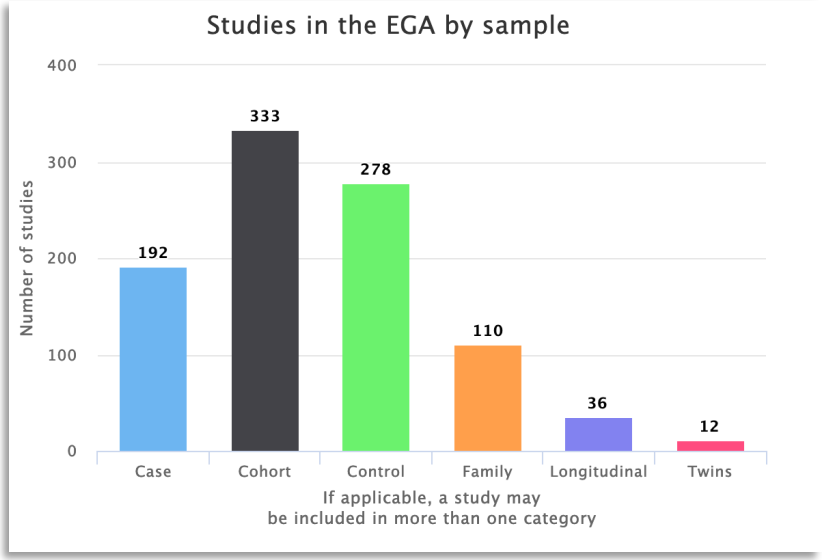
X-omics.nl

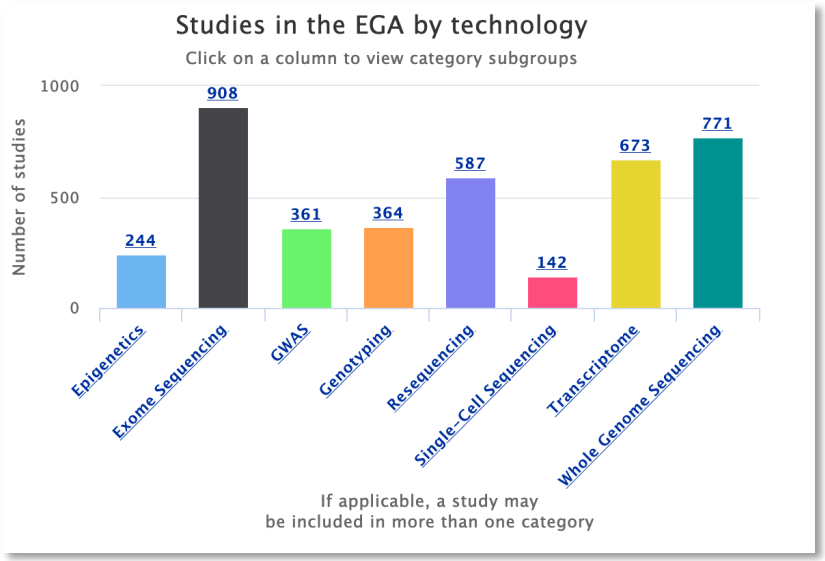

X-omics.nl

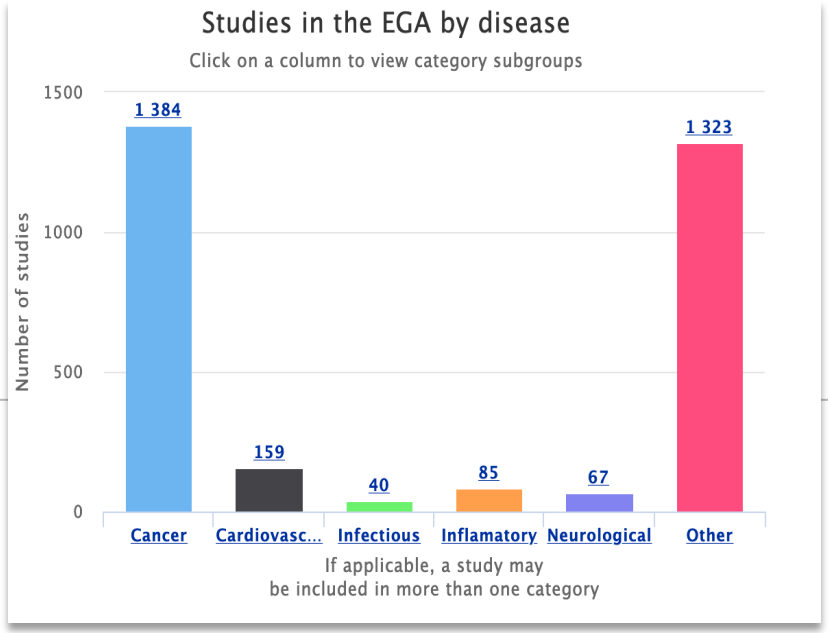
# EGA Data Standard

- EGA intro
- Submitting and Requesting to/from the EGA
- Data Use Ontology (DUO)

# The European Genome Phenome Archive




https://ega-archive.org/

# EGA and Data Access Committee

EUROPEAN
GENOME-PHENOME
ARCHIVE

**Data Access Committee (DAC)**

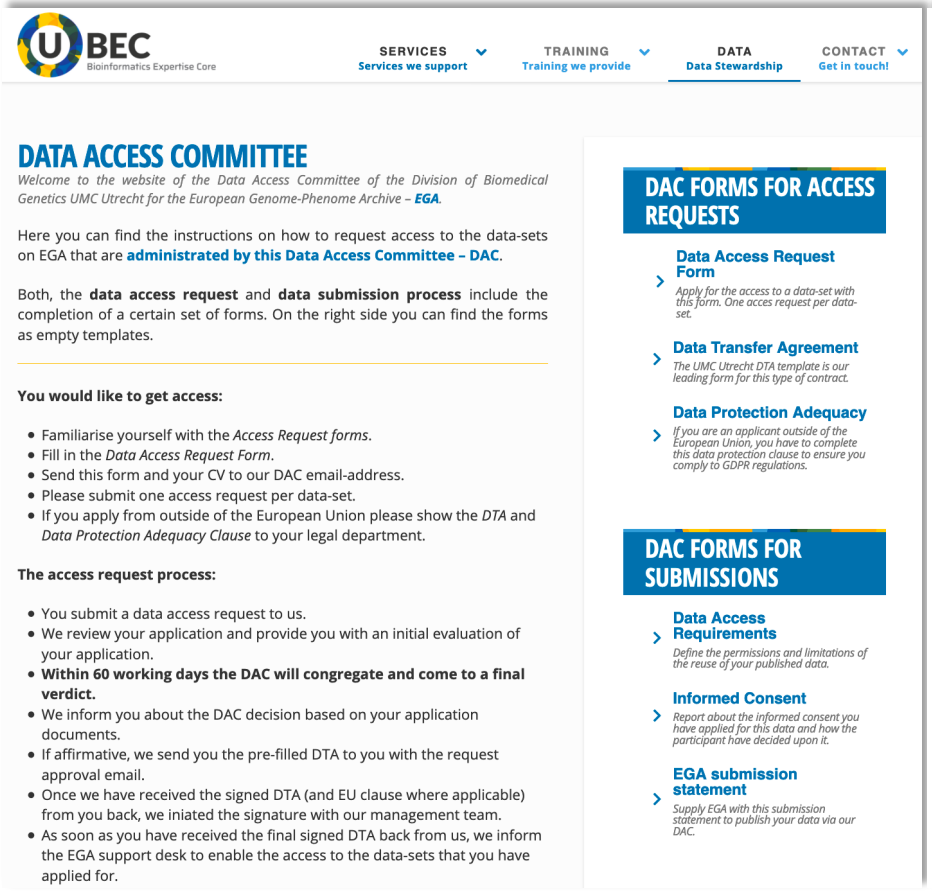
**SUBMISSION**

Data Access Requirements
Informed Consent
Data Transfer Agreement – DTA
EU GDPR Contract

**REQUEST**

...

# DAC example



UBEC Bioinformatics Expertise Core

SERVICES Services we support | TRAINING Training we provide | DATA Data Stewardship | CONTACT Get in touch!

## DATA ACCESS COMMITTEE

*Welcome to the website of the Data Access Committee of the Division of Biomedical Genetics UMC Utrecht for the European Genome-Phenome Archive – **EGA**.*

Here you can find the instructions on how to request access to the data-sets on EGA that are **administrated by this Data Access Committee – DAC**.

Both, the **data access request** and **data submission process** include the completion of a certain set of forms. On the right side you can find the forms as empty templates.

**You would like to get access:**

- Familiarise yourself with the *Access Request forms*.
- Fill in the *Data Access Request Form*.
- Send this form and your CV to our DAC email-address.
- Please submit one access request per data-set.
- If you apply from outside of the European Union please show the *DTA* and *Data Protection Adequacy Clause* to your legal department.

**The access request process:**

- You submit a data access request to us.
- We review your application and provide you with an initial evaluation of your application.
- **Within 60 working days the DAC will congregate and come to a final verdict.**
- We inform you about the DAC decision based on your application documents.
- If affirmative, we send you the pre-filled DTA to you with the request approval email.
- Once we have received the signed DTA (and EU clause where applicable) from you back, we iniated the signature with our management team.
- As soon as you have received the final signed DTA back from us, we inform the EGA support desk to enable the access to the data-sets that you have applied for.

### DAC FORMS FOR ACCESS REQUESTS

**Data Access Request Form**
*Apply for the access to a data-set with this form. One acces request per data-set.*

**Data Transfer Agreement**
*The UMC Utrecht DTA template is our leading form for this type of contract.*

**Data Protection Adequacy**
*If you are an applicant outside of the European Union, you have to complete this data protection clause to ensure you comply to GDPR regulations.*

### DAC FORMS FOR SUBMISSIONS

**Data Access Requirements**
*Define the permissions and limitations of the reuse of your published data.*

**Informed Consent**
*Report about the informed consent you have applied for this data and how the participant have decided upon it.*

**EGA submission statement**
*Supply EGA with this submission statement to publish your data via our DAC.*

## DAC

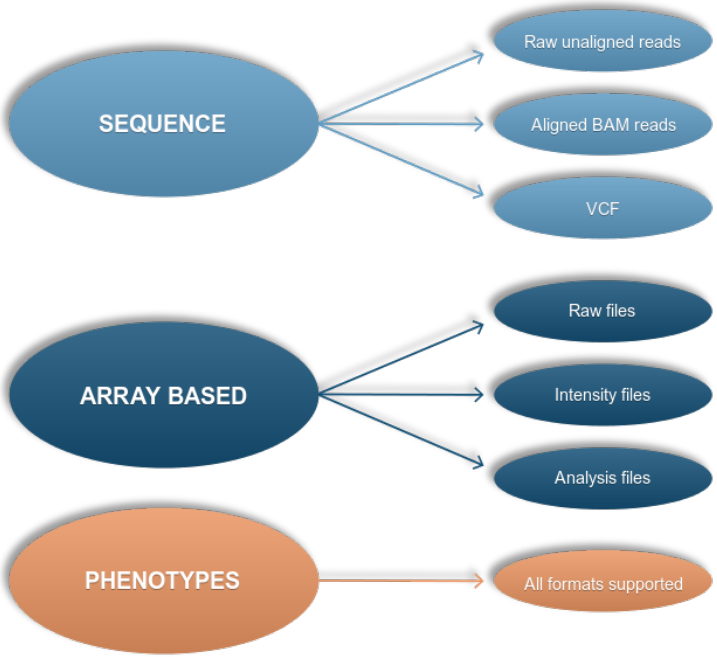
### Division of Biomedical Genetics UMC Utrecht

| Dac ID | Contact Person | Email | Access Information |
|---|---|---|---|
| EGAC00001000432 | DAC DBG | dacdbg [at] umcutrecht [dot] nl | http://ubec.nl/data-access-committee/ |

**This DAC controls 35 datasets:**

| Dataset ID | Description | Technology | Samples |
|---|---|---|---|
| EGAD00001001900 | DNA sequencing reads of human adult stem cell cultures from liver, colon and small intestine. Including biopsy or blood samples of the donors. | HiSeq X Ten,Illumina HiSeq 2500,NextSeq 500 | 61 |
| EGAD00001002242 | This dataset contains RNA-seq and Hi-C data files of induced pluripotent stem (iPS) cells and iPS cell-derived neural progenitors (NPCs) derived from a germline chromothripsis patient and both parents. iPS cells of the patient (cell lines 14 and 15), the father (lines 23 (with two replicates) and 32) and mother (line 30) were differentiated to NPCs and RNA was collected on day 0, day 7 and day 10 of differentiation. In addition, Hi-C data for two iPS cell-derived NPC lines from the patient (14 and 15) and two lines from the father (23 and 32) was generated. | AB 5500xl Genetic Analyzer,Illumina HiSeq 2500,NextSeq 500 | 22 |
| EGAD00001002719 | This dataset contains whole-genome sequencing data files from colon organoid cultures, which were mutated using CRISPR-Cas9 for specific genes (APC, KRAS, TP53 and SMAD4) to generate in vitro transformed cancer cells. After introducing each mutation, the resulting cultures were subjected to whole-genome sequencing. In addition, some cultures were xenotransplanted in recipient mice. The resulting primary tumors and corresponding metastases were subjected to whole-genome sequencing. | HiSeq X Ten | 30 |
| EGAD00001003291 | This dataset represents RNA-sequencing data from 278 primary colon cancers obtained from fresh-frozen tumor sections. RNA-sequencing was performed using TruSeq library preparation and samples were sequenced on Illumina NextSeq and HiSeq. The data are available as Illumina NextSeq and HiSeq fastq files (_R1.fastq and _R2.fastq for each tumor sample, 556 files in total). | Illumina HiSeq 2500,NextSeq 500 | 278 |

# Submitting to EGA




Accepted Data Types by EGA

```
########
# STUDY #
########
STUDY :
  title : ""                    #Full title for your study
  type :                        #Choose from STUDY_TYPES in ENUMS below
  short_name : ""               #Short title for your study
  abstract : |
  Please start typing your abstract here. It can contain
  multiple lines and indentations.

    It doesn't have to be quoted, line ending and special characters will be preserved.
    Please do not remove the | after 'abstract : '.

##########
# SAMPLES #
##########

SAMPLES:                        #Each - represents a new sample
  - alias : ""                  #Short name for sample 1
    title : ""                  #Full name for sample 1
    decription : ""             #Optional decription for sample 1
    subject_id : ""             #Name of whatever sample 1 is derived from (e.g. a patient ID or cellline id)
    case/control :              #Choose from CASE_CONTROL_TYPES in ENUMS below
    gender :                    #Choose from GENDER_TYPES in ENUMS below
    organism_part :             #Optional organism part for sample 1
    cell_line :                 #Optional cell line name for sample 1
    region :                    #Optional region for sample 1
    phenotype/disease :         #Choose a phenotype/disease for sample 1
  - alias : ""                  #Short name for sample 2
    title : ""                  #Full name for sample 2
    decription : ""             #Optional decription for sample 2
    subject_id : ""             #Name of whatever sample 2 is derived from (e.g. a patient ID or cellline id)
    case/control :              #Choose from CASE_CONTROL_TYPES in ENUMS below
    gender :                    #Choose from GENDER_TYPES in ENUMS below
    organism_part :             #Optional organism part for sample 2
    cell_line :                 #Optional cell line name for sample 2
    region :                    #Optional region for sample 2
    phenotype/disease :         #Choose a phenotype/disease for sample 2

###################
# SEQUENCING RUNS #
###################
```

Requested Metadata by EGA

# Submitting to EGA via the DAC

67 Files

5.16 TB

Version 1.1

UMC Utrecht

**DATA ACCESS REQUIREMENTS**

***DAC DBG UMC Utrecht***

This document contains the details of the requirements to access a research dataset submitted to a repository and is completed by the researchers involved in the project. The competed 'Data Access Requirements' will be the only information that is used by the Data Access Committee DBG of the UMC Utrecht to judge data access requests that are received, and thus completeness of this form is crucial.

**Submission date:**

**Dataset details:**
**Dataset reference (Study ID and Dataset Details)**

**Name of the project that created the dataset**

**Names of other data producers/collaborators**

**Data owner**
UMC Utrecht

**Specific limitations on areas of research**[1]

Consent type?

Consent is available for the individuals that participated in this project: ☐ Yes / ☐ No.
If 'Yes', the information and consent forms is located at:

**Other limitations for use**

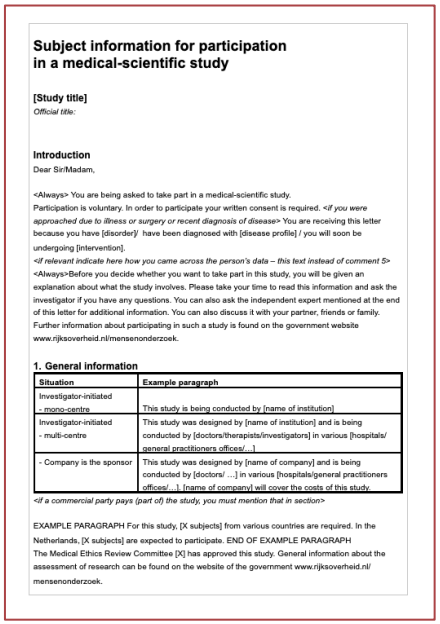
Citing and crediting

License that is applicable on this dataset

[1] For example, has to be excluded due to consent type/contents or disallowing commercial parties to gain access.
[2] Please also specify if a tiered consent if used and how this affects the way the data could be used.

*DAC DBG | Center for Molecular Medicine | Department of Genetics | UMC Utrecht | HP STR2.203 | Heidelberglaan 100 | 3584 CX | Utrecht | The Netherlands | DAC[DBG@umcutrecht.nl](mailto:DBG@umcutrecht.nl)*

Access Requirements



**Subject information for participation in a medical-scientific study**

**[Study title]**
*Official title:*

**Introduction**

Dear Sir/Madam,

<Always> You are being asked to take part in a medical-scientific study.
Participation is voluntary. In order to participate your written consent is required. *<if you were approached due to illness or surgery or recent diagnosis of disease>* You are receiving this letter because you have [disorder]/ have been diagnosed with [disease profile] / you will soon be undergoing [intervention].
*<if relevant indicate here how you came across the person's data – this text instead of comment 5>*
<Always>Before you decide whether you want to take part in this study, you will be given an explanation about what the study involves. Please take your time to read this information and ask the investigator if you have any questions. You can also ask the independent expert mentioned at the end of this letter for additional information. You can also discuss it with your partner, friends or family.
Further information about participating in such a study is found on the government website www.rijksoverheid.nl/mensenonderzoek.

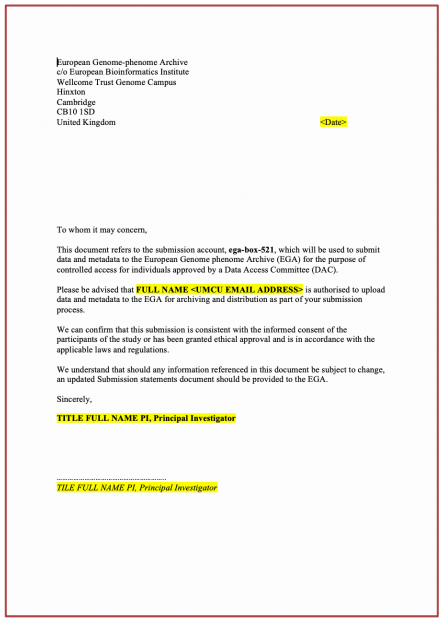
**1. General information**

| Situation | Example paragraph |
|---|---|
| Investigator-initiated - mono-centre | This study is being conducted by [name of institution] |
| Investigator-initiated - multi-centre | This study was designed by [name of institution] and is being conducted by [doctors/therapists/investigators] in various [hospitals/ general practitioners offices/…] |
| - Company is the sponsor | This study was designed by [name of company] and is being conducted by [doctors/ …] in various [hospitals/general practitioners offices/…]. [name of company] will cover the costs of this study. |

*<if a commercial party pays (part of) the study, you must mention that in section>*

EXAMPLE PARAGRAPH For this study, [X subjects] from various countries are required. In the Netherlands, [X subjects] are expected to participate. END OF EXAMPLE PARAGRAPH
The Medical Ethics Review Committee [X] has approved this study. General information about the assessment of research can be found on the website of the government www.rijksoverheid.nl/mensenonderzoek.

Informed Consent



European Genome-phenome Archive
c/o European Bioinformatics Institute
Wellcome Trust Genome Campus
Hinxton
Cambridge
CB10 1SD
United Kingdom

<Date>

To whom it may concern,

This document refers to the submission account, **ega-box-521**, which will be used to submit data and metadata to the European Genome phenome Archive (EGA) for the purpose of controlled access for individuals approved by a Data Access Committee (DAC).

Please be advised that **FULL NAME <UMCU EMAIL ADDRESS>** is authorised to upload data and metadata to the EGA for archiving and distribution as part of your submission process.

We can confirm that this submission is consistent with the informed consent of the participants of the study or has been granted ethical approval and is in accordance with the applicable laws and regulations.

We understand that should any information referenced in this document be subject to change, an updated Submission statements document should be provided to the EGA.

Sincerely,

**TITLE FULL NAME PI, Principal Investigator**

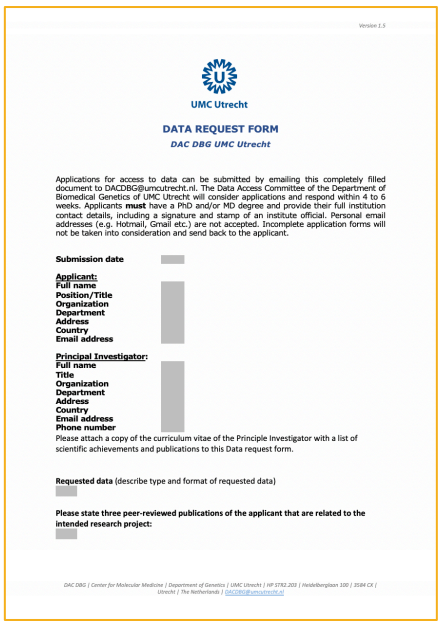
.....

*TILE FULL NAME PI, Principal Investigator*

EGA Submission Statement

# Requesting Access from EGA via the DAC

67 Files  5.16 TB



UMC Utrecht

DATA REQUEST FORM

*DAC DBG UMC Utrecht*

Applications for access to data can be submitted by emailing this completely filled document to DACDBG@umcutrecht.nl. The Data Access Committee of the Department of Biomedical Genetics of UMC Utrecht will consider applications and respond within 4 to 6 weeks. Applicants **must** have a PhD and/or MD degree and provide their full institution contact details, including a signature and stamp of an institute official. Personal email addresses (e.g. Hotmail, Gmail etc.) are not accepted. Incomplete application forms will not be taken into consideration and send back to the applicant.

**Submission date**

**Applicant:**
**Full name**
**Position/Title**
**Organization**
**Department**
**Address**
**Country**
**Email address**

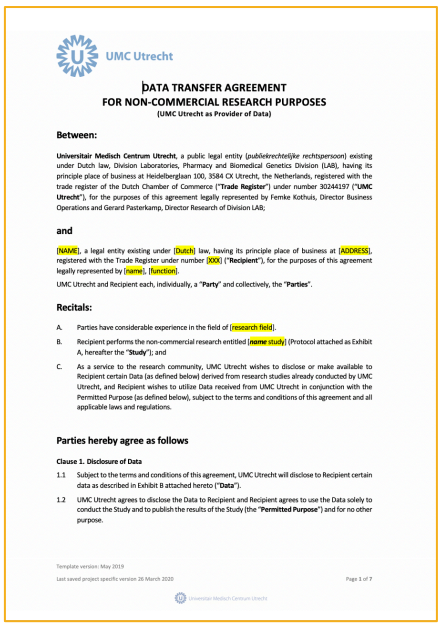
**Principal Investigator:**
**Full name**
**Title**
**Organization**
**Department**
**Address**
**Country**
**Email address**
**Phone number**

Please attach a copy of the curriculum vitae of the Principle Investigator with a list of scientific achievements and publications to this Data request form.

**Requested data** (describe type and format of requested data)

**Please state three peer-reviewed publications of the applicant that are related to the intended research project:**

## Access Request and Application



UMC Utrecht

DATA TRANSFER AGREEMENT
FOR NON-COMMERCIAL RESEARCH PURPOSES
(UMC Utrecht as Provider of Data)

**Between:**

**Universitair Medisch Centrum Utrecht**, a public legal entity (*publiekrechtelijke rechtspersoon*) existing under Dutch law, Division Laboratories, Pharmacy and Biomedical Genetics Division (LAB), having its principle place of business at Heidelberglaan 100, 3584 CX Utrecht, the Netherlands, registered with the trade register of the Dutch Chamber of Commerce (**"Trade Register"**) under number 30244197 (**"UMC Utrecht"**), for the purposes of this agreement legally represented by Femke Kothuis, Director Business Operations and Gerard Pasterkamp, Director Research of Division LAB;

**and**

[NAME], a legal entity existing under [Dutch] law, having its principle place of business at [ADDRESS], registered with the Trade Register under number [XXX] (**"Recipient"**), for the purposes of this agreement legally represented by [name], [function].

UMC Utrecht and Recipient each, individually, a **"Party"** and collectively, the **"Parties"**.

**Recitals:**

A. Parties have considerable experience in the field of [research field].

B. Recipient performs the non-commercial research entitled [***name study***] (Protocol attached as Exhibit A, hereafter the **"Study"**); and

C. As a service to the research community, UMC Utrecht wishes to disclose or make available to Recipient certain Data (as defined below) derived from research studies already conducted by UMC Utrecht, and Recipient wishes to utilize Data received from UMC Utrecht in conjunction with the Permitted Purpose (as defined below), subject to the terms and conditions of this agreement and all applicable laws and regulations.

**Parties hereby agree as follows**

**Clause 1. Disclosure of Data**

1.1 Subject to the terms and conditions of this agreement, UMC Utrecht will disclose to Recipient certain data as described in Exhibit B attached hereto (**"Data"**).

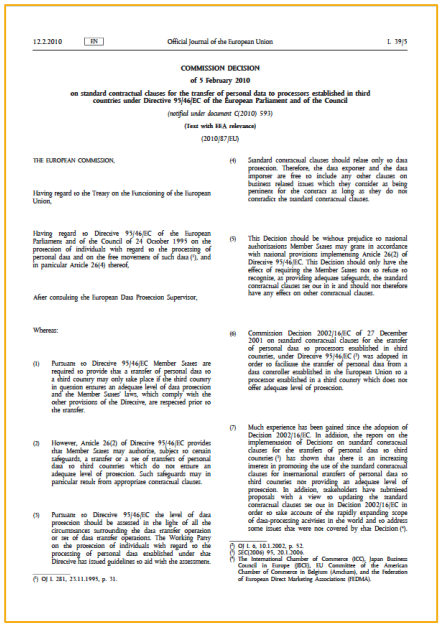
1.2 UMC Utrecht agrees to disclose the Data to Recipient and Recipient agrees to use the Data solely to conduct the Study and to publish the results of the Study (the **"Permitted Purpose"**) and for no other purpose.

Template version: May 2019

Last saved project specific version 26 March 2020

Page 1 of 7

## Data Transfer Agreement



12.2.2010 EN Official Journal of the European Union L 39/5

**COMMISSION DECISION**
**of 5 February 2010**
**on standard contractual clauses for the transfer of personal data to processors established in third countries under Directive 95/46/EC of the European Parliament and of the Council**
*(notified under document C(2010) 593)*
**(Text with EEA relevance)**
**(2010/87/EU)**

THE EUROPEAN COMMISSION,

Having regard to the Treaty on the Functioning of the European Union,

Having regard to Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (1), and in particular Article 26(4) thereof,

After consulting the European Data Protection Supervisor,

Whereas:

(1) Pursuant to Directive 95/46/EC Member States are required to provide that a transfer of personal data to a third country may only take place if the third country in question ensures an adequate level of data protection and the Member States' laws, which comply with the other provisions of the Directive, are respected prior to the transfer.

(2) However, Article 26(2) of Directive 95/46/EC provides that Member States may authorise, subject to certain safeguards, a transfer or a set of transfers of personal data to third countries which do not ensure an adequate level of protection. Such safeguards may in particular result from appropriate contractual clauses.

(3) Pursuant to Directive 95/46/EC the level of data protection should be assessed in the light of all the circumstances surrounding the data transfer operation or set of data transfer operations. The Working Party on the protection of individuals with regard to the processing of personal data established under that Directive has issued guidelines to aid with the assessment.

(4) Standard contractual clauses should relate only to data protection. Therefore, the data exporter and the data importer are free to include any other clauses on business related issues which they consider as being pertinent for the contract as long as they do not contradict the standard contractual clauses.

(5) This Decision should be without prejudice to national authorisations Member States may grant in accordance with national provisions implementing Article 26(2) of Directive 95/46/EC. This Decision should only have the effect of requiring the Member States not to refuse to recognise, as providing adequate safeguards, the standard contractual clauses set out in it and should not therefore have any effect on other contractual clauses.

(6) Commission Decision 2002/16/EC of 27 December 2001 on standard contractual clauses for the transfer of personal data to processors established in third countries, under Directive 95/46/EC (2) was adopted in order to facilitate the transfer of personal data from a data controller established in the European Union to a processor established in a third country which does not offer adequate level of protection.

(7) Much experience has been gained since the adoption of Decision 2002/16/EC. In addition, the reports on the implementation of Decisions on standard contractual clauses for the transfers of personal data to third countries (3) has shown that there is an increasing interest in promoting the use of the standard contractual clauses for international transfers of personal data to third countries now providing an adequate level of protection. In addition, stakeholders have submitted proposals with a view to updating the standard contractual clauses set out in Decision 2002/16/EC in order to take account of the rapidly expanding scope of data-processing activities in the world and to address some issues that were not covered by this Decision (4).

(1) OJ L 281, 23.11.1995, p. 31.

(2) OJ L 6, 10.1.2002, p. 52.
(3) SEC(2006) 95, 20.1.2006.
(4) The International Chamber of Commerce (ICC), Japan Business Council in Europe (JBCE), EU Committee of the American Chamber of Commerce in Belgium (Amcham), and the Federation of European Direct Marketing Associations (FEDMA).

## EC GDPR Clause outside of EU

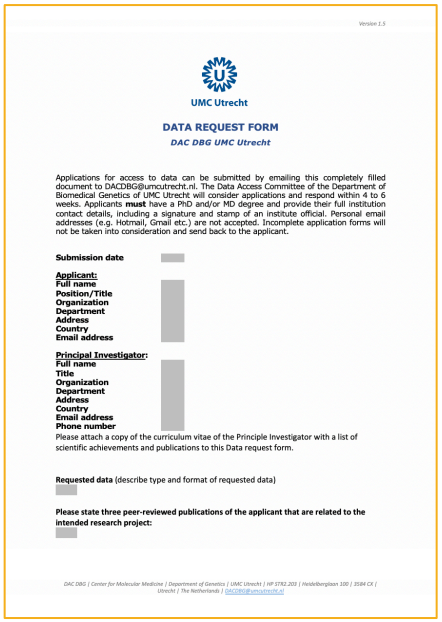


X-omics.nl

# Assessing data access applications

## ACCESS REQUEST:

**Affiliation**
**Collaboration**
**Research Question**



UMC Utrecht

DATA REQUEST FORM

*DAC DBG UMC Utrecht*

Applications for access to data can be submitted by emailing this completely filled document to DACDBG@umcutrecht.nl. The Data Access Committee of the Department of Biomedical Genetics of UMC Utrecht will consider applications and respond within 4 to 6 weeks. Applicants **must** have a PhD and/or MD degree and provide their full institution contact details, including a signature and stamp of an institute official. Personal email addresses (e.g. Hotmail, Gmail etc.) are not accepted. Incomplete application forms will not be taken into consideration and send back to the applicant.

**Submission date**

**Applicant:**
**Full name**
**Position/Title**
**Organization**
**Department**
**Address**
**Country**
**Email address**

**Principal Investigator:**
**Full name**
**Title**
**Organization**
**Department**
**Address**
**Country**
**Email address**
**Phone number**

Please attach a copy of the curriculum vitae of the Principle Investigator with a list of scientific achievements and publications to this Data request form.

**Requested data** (describe type and format of requested data)

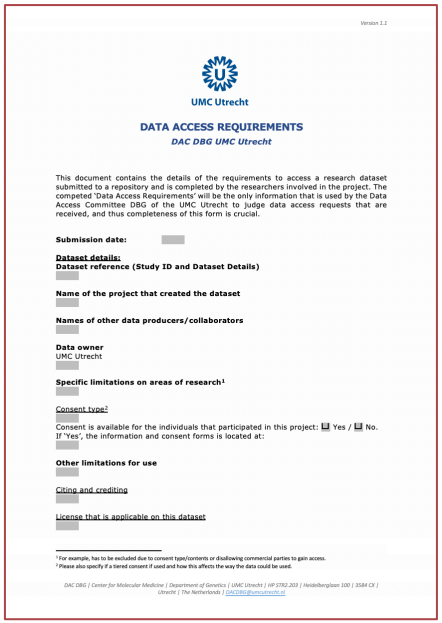
**Please state three peer-reviewed publications of the applicant that are related to the intended research project:**

DAC DBG | Center for Molecular Medicine | Department of Genetics | UMC Utrecht | HP STR2.203 | Heidelberglaan 100 | 3584 CX | Utrecht | The Netherlands | DACDBG@umcutrecht.nl

VS.

## ACCESS RESTRICTIONS:

**Re-use consent**
**Further research limitations**
**Other restrictions**



UMC Utrecht

DATA ACCESS REQUIREMENTS

*DAC DBG UMC Utrecht*

This document contains the details of the requirements to access a research dataset submitted to a repository and is completed by the researchers involved in the project. The competed 'Data Access Requirements' will be the only information that is used by the Data Access Committee DBG of the UMC Utrecht to judge data access requests that are received, and thus completeness of this form is crucial.

**Submission date:**

**Dataset details:**
**Dataset reference (Study ID and Dataset Details)**

**Name of the project that created the dataset**

**Names of other data producers/collaborators**

**Data owner**
UMC Utrecht

**Specific limitations on areas of research**[1]

Consent type[2]

Consent is available for the individuals that participated in this project: ☐ Yes / ☐ No.
If 'Yes', the information and consent forms is located at:

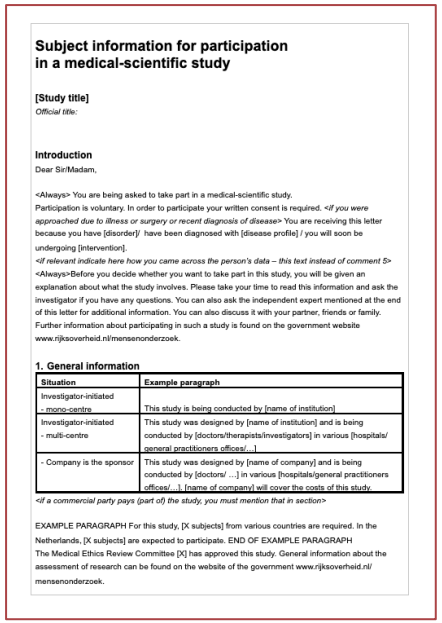
**Other limitations for use**

Citing and crediting

License that is applicable on this dataset

[1] For example, has to be excluded due to consent type/contents or disallowing commercial parties to gain access.
[2] Please also specify if a tiered consent if used and how this affects the way the data could be used.

DAC DBG | Center for Molecular Medicine | Department of Genetics | UMC Utrecht | HP STR2.203 | Heidelberglaan 100 | 3584 CX | Utrecht | The Netherlands | DACDBG@umcutrecht.nl



**Subject information for participation in a medical-scientific study**

**[Study title]**
*Official title:*

**Introduction**
Dear Sir/Madam,

<Always> You are being asked to take part in a medical-scientific study.
Participation is voluntary. In order to participate your written consent is required. *<If you were approached due to illness or surgery or recent diagnosis of disease>* You are receiving this letter because you have [disorder]/ have been diagnosed with [disease profile] / you will soon be undergoing [intervention].
*<If relevant indicate here how you came across the person's data – this text instead of comment 5>*
<Always>Before you decide whether you want to take part in this study, you will be given an explanation about what the study involves. Please take your time to read this information and ask the investigator if you have any questions. You can also ask the independent expert mentioned at the end of this letter for additional information. You can also discuss it with your partner, friends or family. Further information about participating in such a study is found on the government website www.rijksoverheid.nl/mensenonderzoek.

**1. General information**

| Situation | Example paragraph |
|---|---|
| Investigator-initiated - mono-centre | This study is being conducted by [name of institution] |
| Investigator-initiated - multi-centre | This study was designed by [name of institution] and is being conducted by [doctors/therapists/investigators] in various [hospitals/ general practitioners offices/…] |
| - Company is the sponsor | This study was designed by [name of company] and is being conducted by [doctors/ …] in various [hospitals/general practitioners offices/ …]. [name of company] will cover the costs of this study. |

*<If a commercial party pays (part of) the study, you must mention that in section>*

EXAMPLE PARAGRAPH For this study, [X subjects] from various countries are required. In the Netherlands, [X subjects] are expected to participate. END OF EXAMPLE PARAGRAPH
The Medical Ethics Review Committee [X] has approved this study. General information about the assessment of research can be found on the website of the government www.rijksoverheid.nl/ mensenonderzoek.

# Standardising data use conditions

## Data Use Categories and Requirements (Consent Codes)

Global Alliance for Genomics & Health

| Consent Codes | | |
|---|---|---|
| **Name** | **Abbreviation** | **Description** |
| **Primary Categories** ($I^{ry}$) | | |
| no restrictions | NRES | No restrictions on data use. |
| general research use and clinical care | GRU(CC) | For health/medical/biomedical purposes and other biological research, including the study of population origins or ancestry. |
| health/medical/biomedical research and clinical care | HMB(CC) | Use of the data is limited to health/medical/biomedical purposes, does not include the study of population origins or ancestry. |
| disease-specific research and clinical care | DS-[XX](CC) | Use of the data must be related to [disease]. |
| population origins/ancestry research | POA | Use of the data is limited to the study of population origins or ancestry. |
| **Secondary Categories** ($II^{ry}$) (can be one or more extra conditions, in addition to $I^{ry}$ category) | | |
| other research-specific restrictions | RS-[XX] | Use of the data is limited to studies of [research type] (e.g., pediatric research). |
| research use only | RUO | Use of data is limited to research purposes (e.g., does not include its use in clinical care). |
| no "general methods" research | NMDS | Use of the data includes methods development research (e.g., development of software or algorithms) ONLY within the bounds of other data use limitations. |
| genetic studies only | GSO | Use of the data is limited to genetic studies only (i.e., no research using only the phenotype data). |
| **Requirements** | | |
| not-for-profit use only | NPU | Use of the data is limited to not-for-profit organizations. |
| publication required | PUB | Requestor agrees to make results of studies using the data available to the larger scientific community. |
| collaboration required | COL-[XX] | Requestor must agree to collaboration with the primary study investigator(s). |
| return data to database/resource | RTN | Requestor must return derived/enriched data to the database/resource. |
| ethics approval required | IRB | Requestor must provide documentation of local IRB/REC approval. |
| geographical restrictions | GS-[XX] | Use of the data is limited to within [geographic region]. |
| publication moratorium/embargo | MOR-[XX] | Requestor agrees not to publish results of studies until [date]. |
| time limits on use | TS-[XX] | Use of data is approved for [x months]. |
| user-specific restrictions | US | Use of data is limited to use by approved users. |
| project-specific restrictions | PS | Use of data is limited to use within an approved project. |
| institution-specific restrictions | IS | Use of data is limited to use within an approved institution. |

SOM Dyke, *et al*. Consent Codes: Upholding Standard Data Use Conditions. *PLoS Genetics* 12(1): e1005772. http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1005772

Contact: Dr. Stephanie Dyke (stephanie.dyke@mcgill.ca)

https://www.ga4gh.org/genomic-data-toolkit/regulatory-ethics-toolkit/

## Data Use Ontology at EGA

The EGA is committed to its involvement in the work of GA4GH. In an effort to enhance data discoverability & streamline data access, EGA have implemented the use of the Data Use Ontology (**DUO**), based on consent codes as described in **Dyke et al. 2017**. The Data Use Ontology codes will be displayed on the live dataset page of your submission to advise any would be requestor on how the data can be used and also to enhance data discoverability as users will be able to search on these codes to find applicable datasets.

Detailed in the table below are the current DUO codes that should be added into the policy section of your submission in **webin** or used in your XML where submitting programmatically. These terms are verified against the current version **here**.

For each policy please select a maximum of one primary code and any number of secondary category codes (if appropriate), which are given in the table below.

| Term | Label | Description |
|---|---|---|
| **Primary Terms** | | |
| DUO:0000004 | no restriction | This consent code primary category indicates there is no restriction on use. |
| DUO:0000005 | general research use and clinical care | This primary category consent code indicates that use is allowed for health/medical/biomedical purposes and other biological research, including the study of population origins or ancestry. |
| DUO:0000006 | health/medical/biomedical research and clinical care | This primary category consent code indicates that use is allowed for health/medical/biomedical purposes; does not include the study of population origins or ancestry. |
| 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. |
| DUO:0000011 | population origins or ancestry research | This primary category consent code indicates that use of the data is limited to the study of population origins or ancestry. |
| **Secondary Terms** (can be one or more extra conditions, in addition to one primary term) | | |

https://ega-archive.org/data-use-conditions

X-omics.nl

# Data Use Ontology (DUO) tags on EGA

## GWG Dataset

| Dataset ID | Technology | Samples |
|---|---|---|
| EGAD00010001874 | Infinium HD Super Microarray | 576 |

### Dataset Description

Patients with T1DM genotyped on Illumina HiScan using Illumina Infinium OmniExpress Exome-8 v1.4 arrays

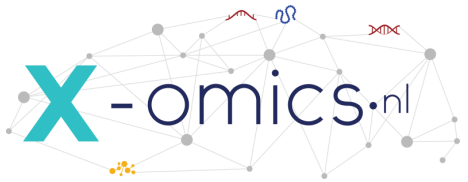
### Data Use Conditions

HMB RUO IRB US

See further information on **Data Use Conditions**

| Label | Code | Version | Modifier |
|---|---|---|---|
| health or medical or biomedical research | DUO:0000006 | 2019-01-07 | |
| research use only | DUO:0000014 | 2019-01-07 | |
| ethics approval required | DUO:0000021 | 2019-01-07 | |
| user specific restriction | DUO:0000026 | 2019-01-07 | |

https://ega-archive.org/datasets/EGAD00010001874

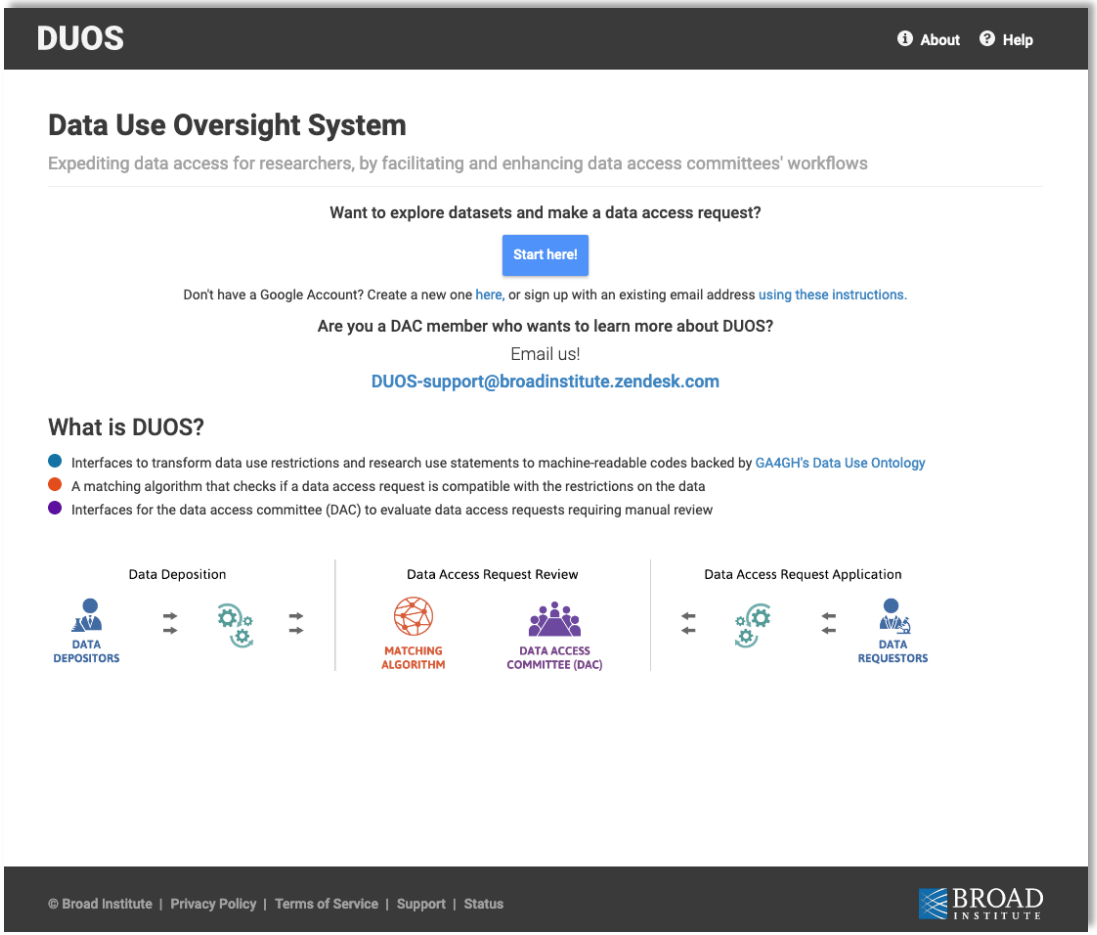


# Benefits of DUO tags

- Standardised and harmonised use conditions
- Better indexing and querying
- Enables future automation
- Enables better interoperability across platforms

- Better informed consent options for study/project participants in the future

# The future:
# Data Use Oversight System (DUOS)




https://duos.broadinstitute.org/

# DAC's, DUO-tags, and DUOS in X-Omics?

# Summary

- Publishing and archiving data via the EGA requires a DAC
- Informed consent conditions are crucial to enable and define future re-use by others

- Data Use Conditions help standardise
- The future is to enable automatised application reviews

- X-Omics will carefully consider to implement DUO and DUOS standards