



# Using GEL data for publications and reports

**Emily Perry**

10<sup>th</sup> September 2024



# Data security

- This training session will include data from the GEL Research Environment
- As part of your IG training you have agreed to not distribute these data in any way
- You are not allowed to:
  - Invite colleagues to watch this training with you
  - Take any screenshots or videos of the training
  - Share your webinar link (we will remove anyone who is here twice)
- We will record this training and distribute the censored video afterwards

# Questions



All your  
microphones  
are muted



Use the Zoom  
Q&A to ask  
questions



Upvote your  
favourite  
questions: if  
we are short  
on time we  
will prioritise  
those with the  
most votes

# Questions



**Mimoza Hoti**  
Clinical  
Research  
Interface  
Lead



**Joanne Yang**  
Clinical  
Research  
Interface  
Coordinator



**Pete  
O'Donovan**  
Airlock Lead



**Mitra Sato**  
Airlock  
Coordinator



**Zainab  
Mustafa**  
Airlock  
Coordinator



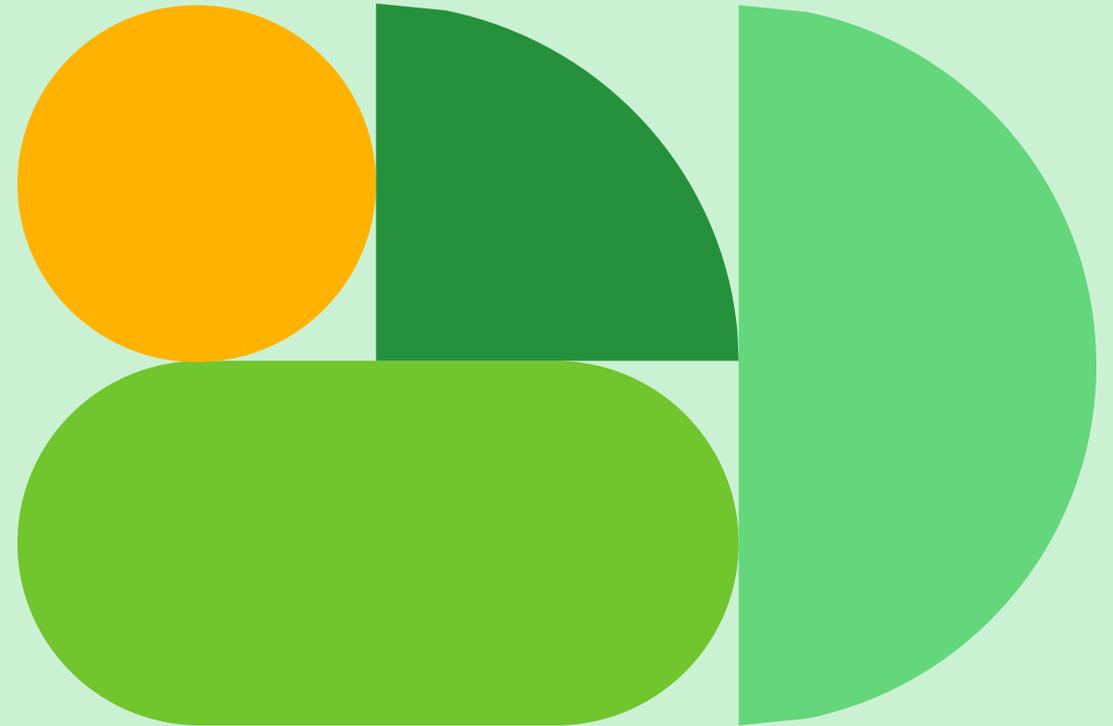
**Alexander  
Renziehausen**  
Head of the  
Science and  
Clinical Office



**Susan Walker**  
Director of  
Translational  
Genomics

# Agenda

- 1 Introduction and admin
- 2 Rules for data export from the RE
- 3 Working with LibreOffice in the RE
- 4 Using the Airlock
- 5 Before you publish
- 6 Sharing data with reviewers and examiners within the RE
- 7 Help and questions



## 2. Rules for data export from the RE

# Our contract with participants

"...although researchers can look at your data and ask questions about it, they can only take away the answers to their questions (their results). They can't copy or take away any of your individual data."



# A Trusted Research Environment

## Researchers can...

- Work with the data in the RE
- Copy/paste in
- Bring in Containers
- Access whitelisted websites
- Request to export the results of their analysis

## Researchers cannot...

- Share folders between their computer and the VM
- Copy/paste out
- Export files
- Access most of the internet

# The Airlock

Data in the RE



Outside world

# Airlock rules

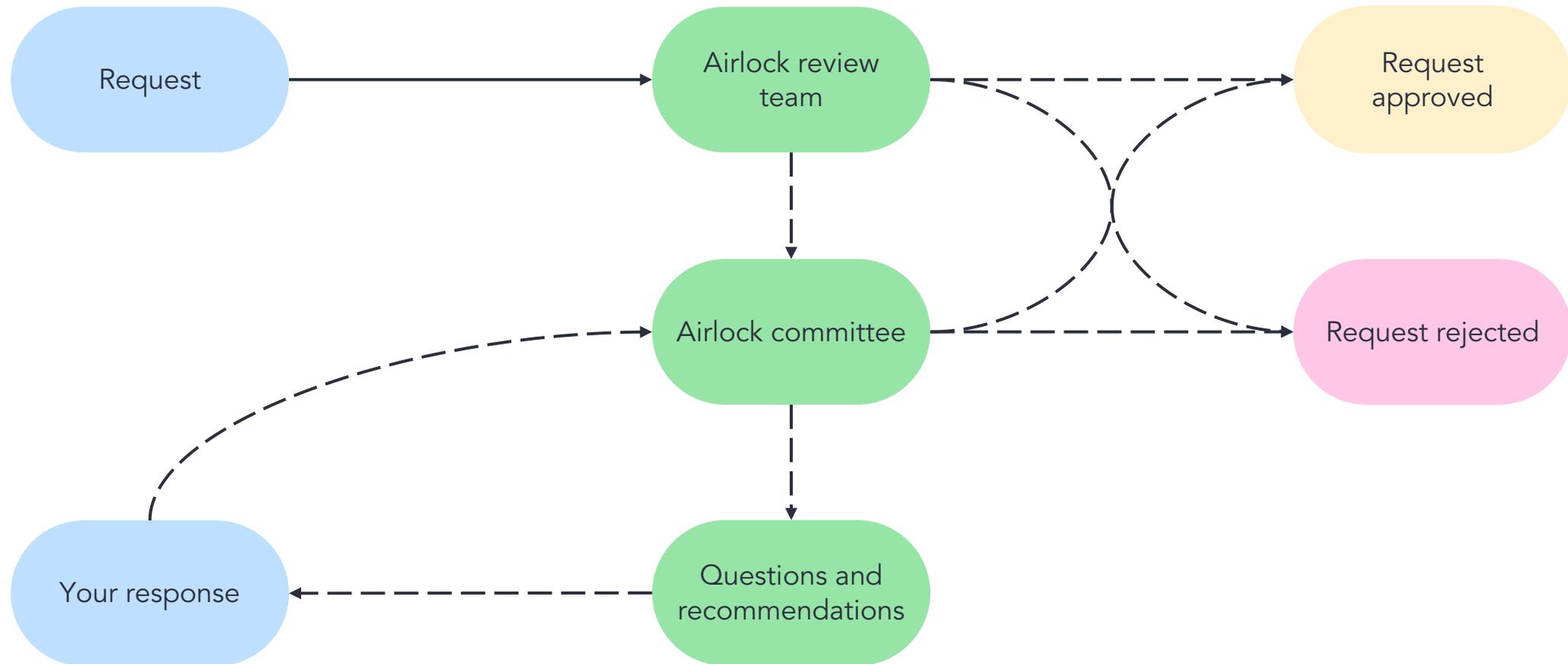


Approved research  
project

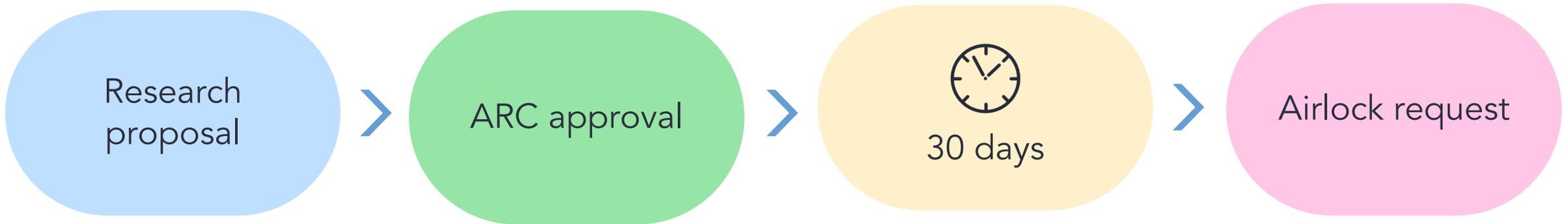


Participants cannot  
be identified

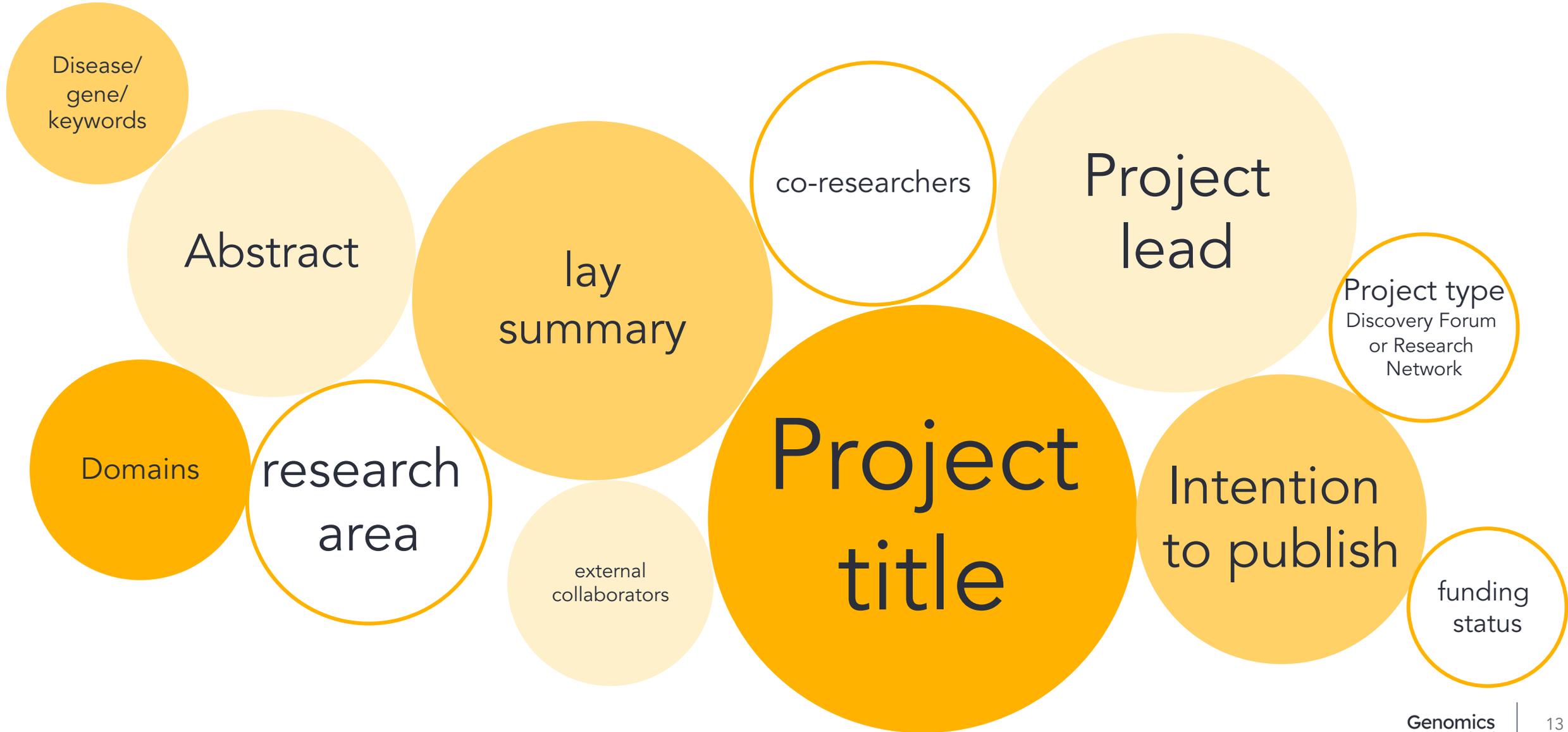
# What happens to my request?



# Approved research project



# Make your research proposal



# Make your research proposal

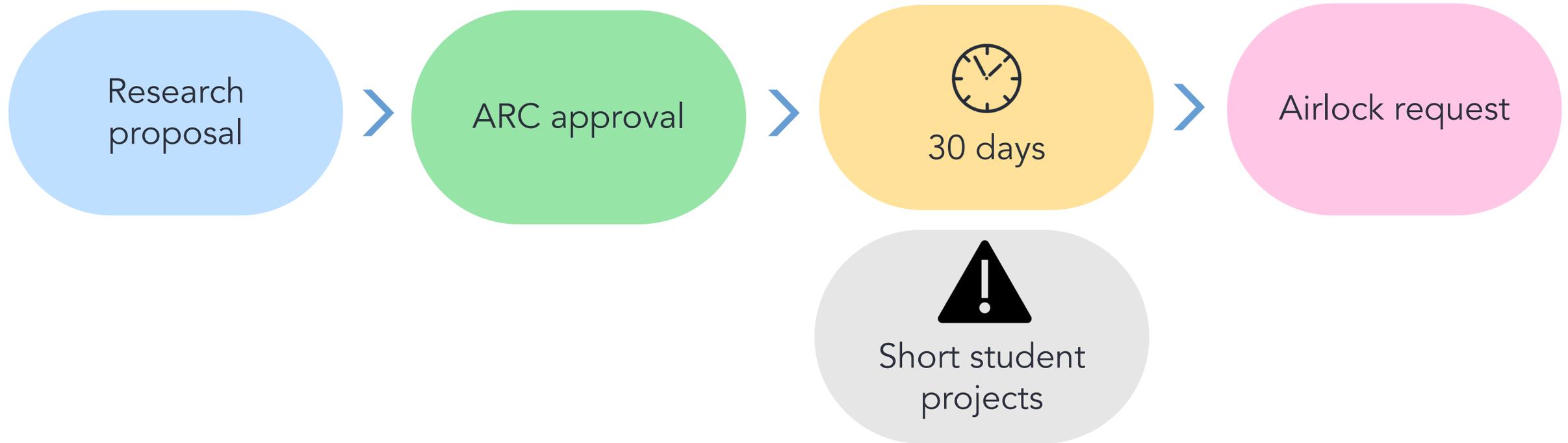


<https://research.genomicsengland.co.uk/research-registry/submit-research-project>

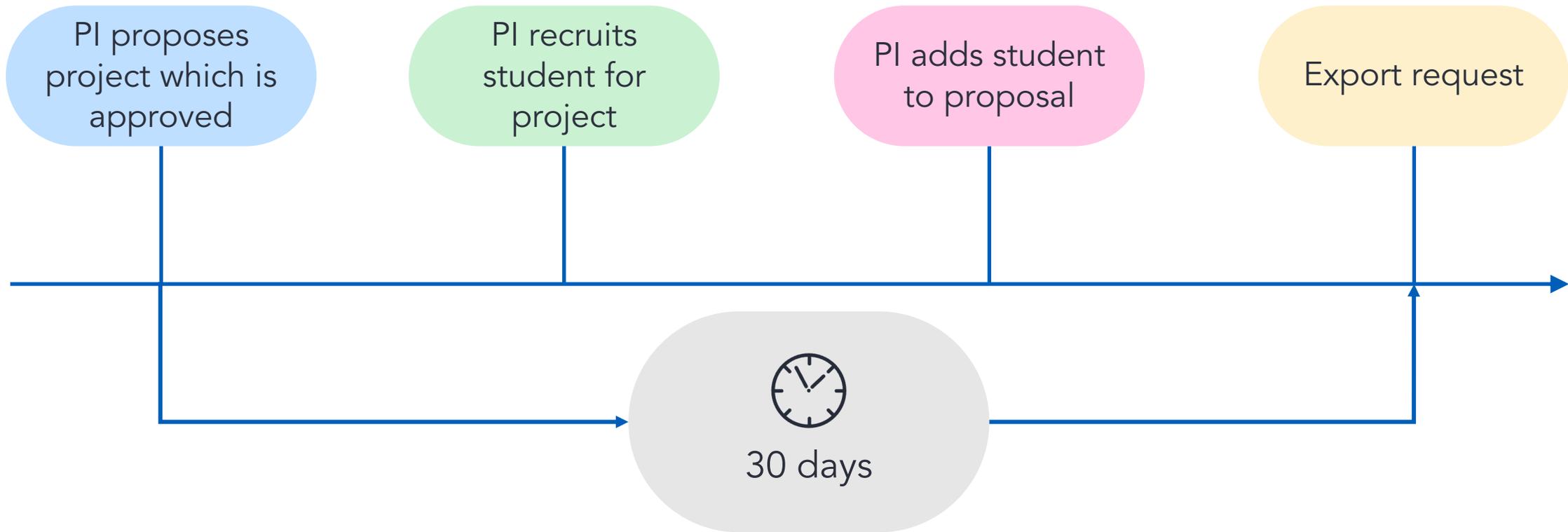


Discuss with your partnerships director

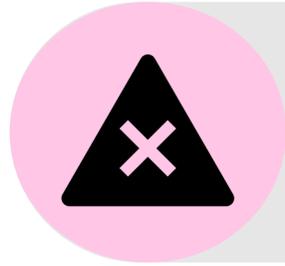
# Approved research project



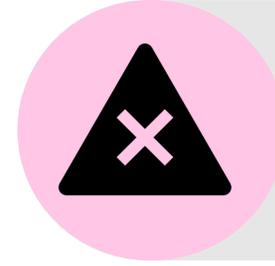
# Approved projects for students



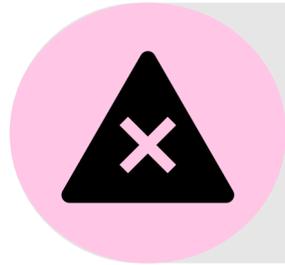
# Participants cannot be identified - to whom?



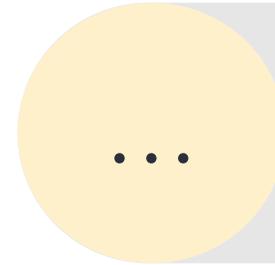
Friends  
and family



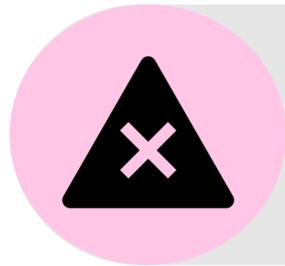
Themselves



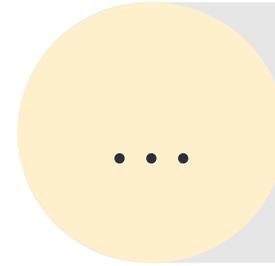
Colleagues/  
classmates



Their clinician

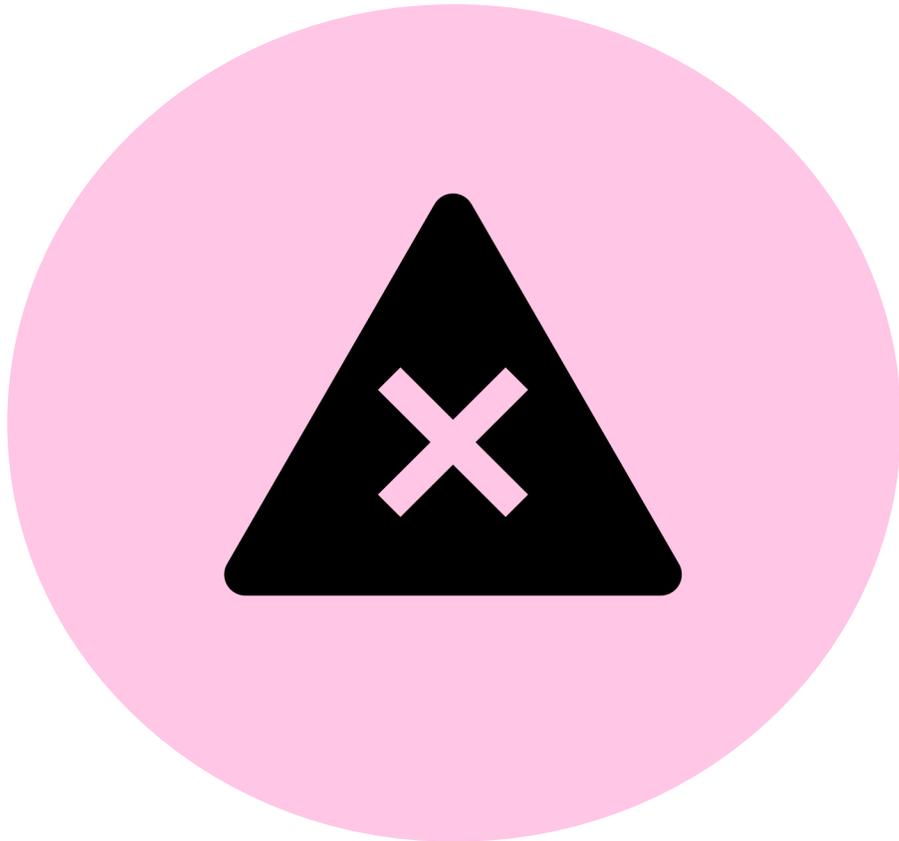


Social media



Scientists  
on their case

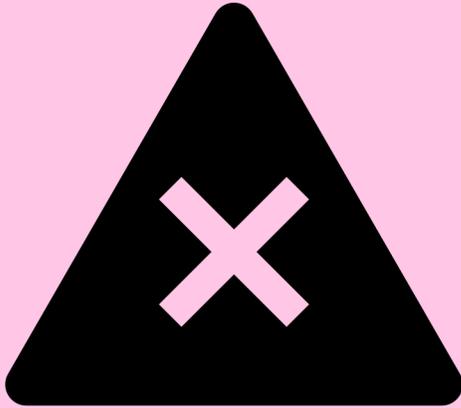
# Participant and sample IDs



Public sample IDs can be used with non-identifiable data (eg tumour mutational signatures). These allow meta-analyses between studies. These always require discussion with the Airlock committee.

# Nucleotides

CCTCAGCCTCCCAACTGTTGGGATTACAG  
GCGTGAGCC...GAATTTTA  
TCGTGG...TAGTT  
TTTGA...AAAA  
TAAA...CTA  
ATA...AAA  
TGCC...TGT  
TACAC...AATG  
GATTAA...ATGCCT  
TAACAAAAC...AGATCTTA  
AGCATT...TTTCC...ATGATCTTTAACTGT  
TCTGGGTCACAAATTTGTCTGTCACTGGTT



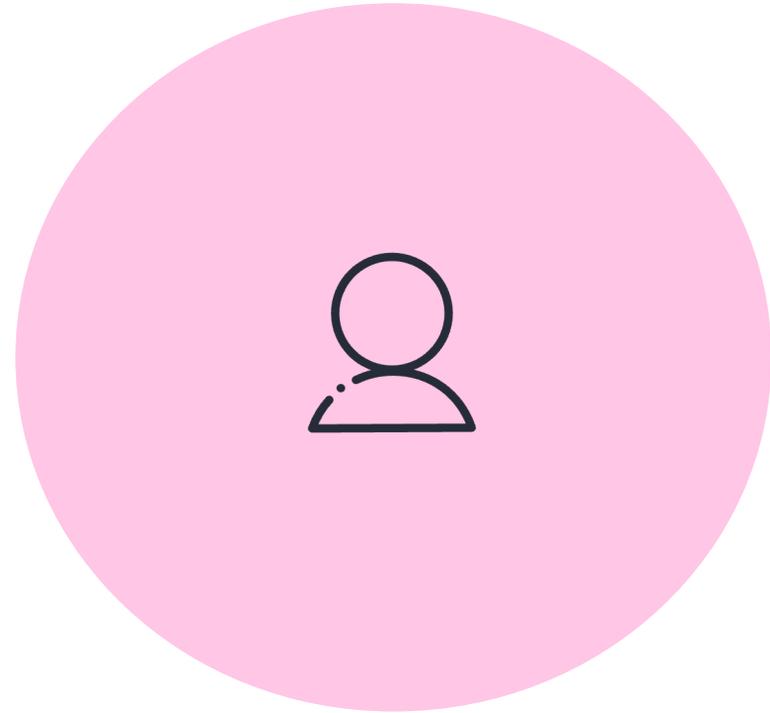
chr1 12345678 A T  
chr... G  
chr11... G C  
chr12 34567890 C A



# Rule of five



Five or more participants



Fewer than five participants

# Rule of five – frequency tables

	Male	Female
Phen 1	128	135
Phen 2	38	56
Phen 3	17	2
Total	185	193



	Male	Female
Phen 1	128	>132
Phen 2	38	56
Phen 3	17	<5
Total	185	193



# Rule of five – frequency tables

	Male	Female
Phen 1	128	135
Phen 2	38	56
Phen 3	17	2
Total	185	193



	Male	Female
Phen 1	128	135
Phen 2	38	56
Phen 3	17	<5



# Rule of five – frequency tables

	Male	Female
Phen 1	128	135
Phen 2	38	56
Phen 3	17	2
Total	185	193



	Male	Female
Phen 1	69%	70%
Phen 2	21%	29%
Phen 3	9%	1%



# Maxima and minima

minimum = 1

?

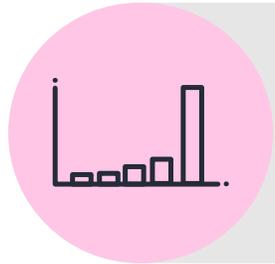
is this something where  
0 should be possible?

min/max is also  
the mode

?

does this correspond  
to a known group?

# Percentiles and medians



skewed distribution



normal distribution



binomial distribution



in any percentile



in any percentile

# Means, indices, ratios, indicators



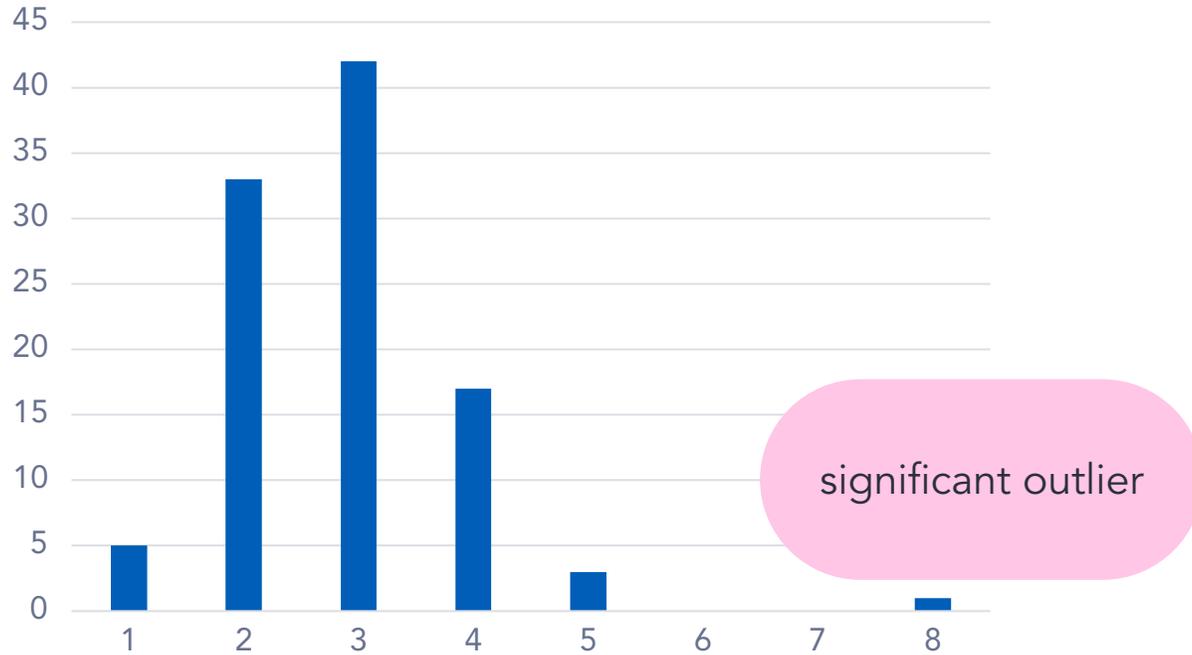
total = 10  
only possible values = 1  
or 0  
mean = 0.7

Value of 0 given away for  
three participants

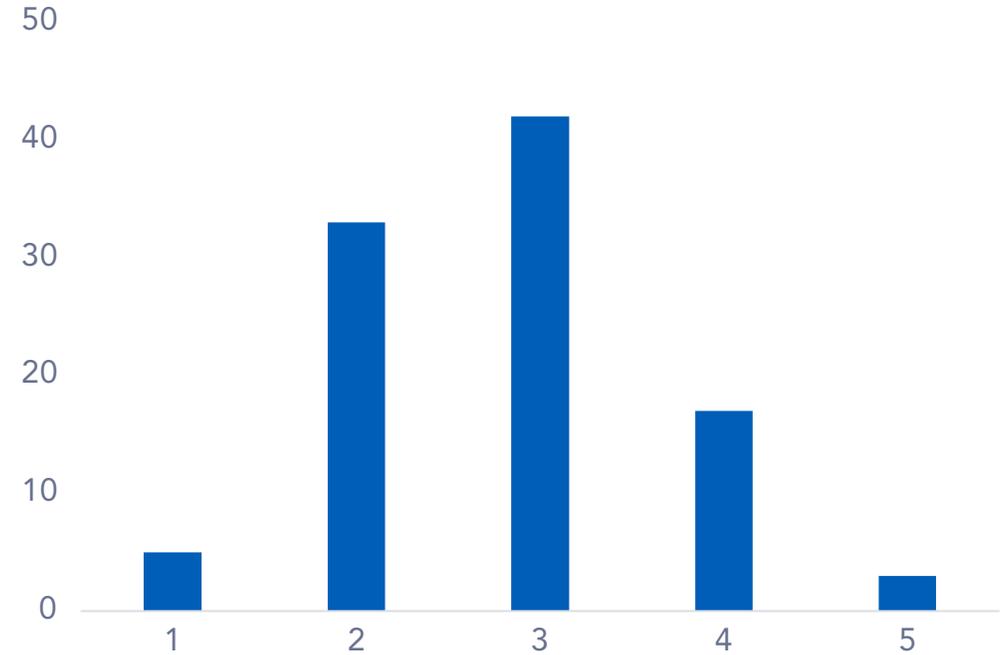


Most cases

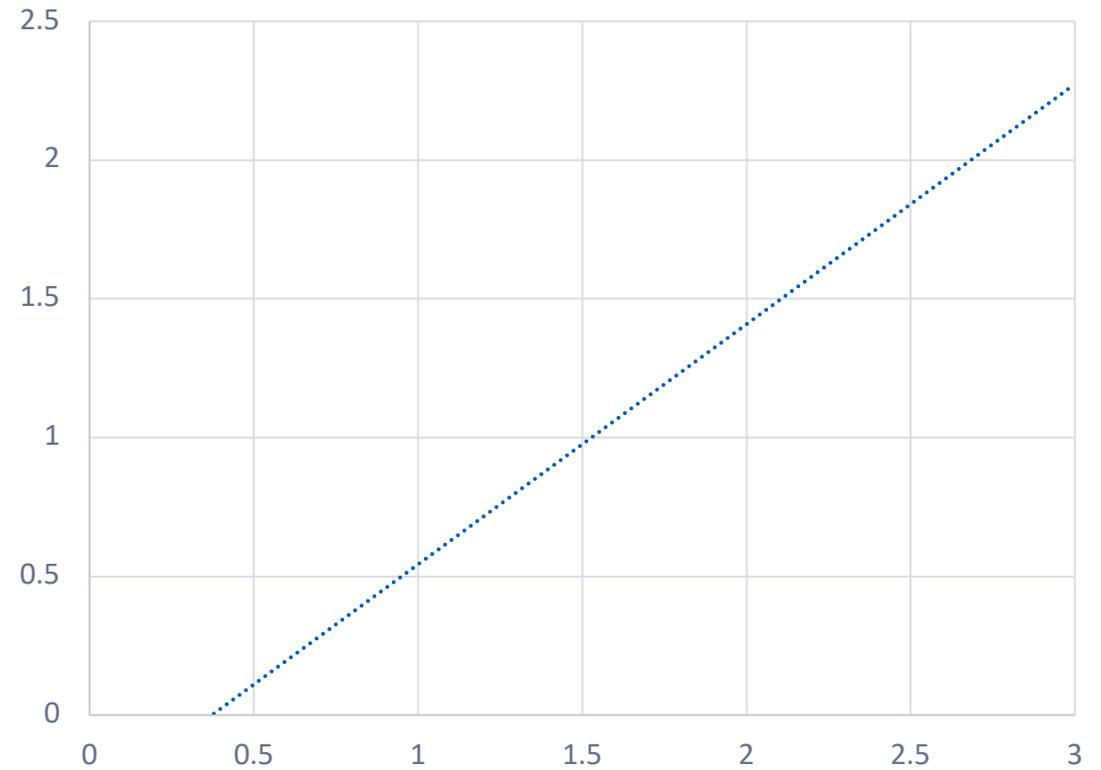
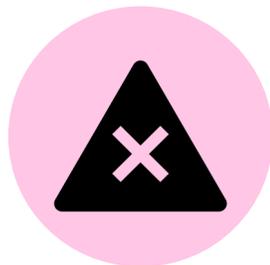
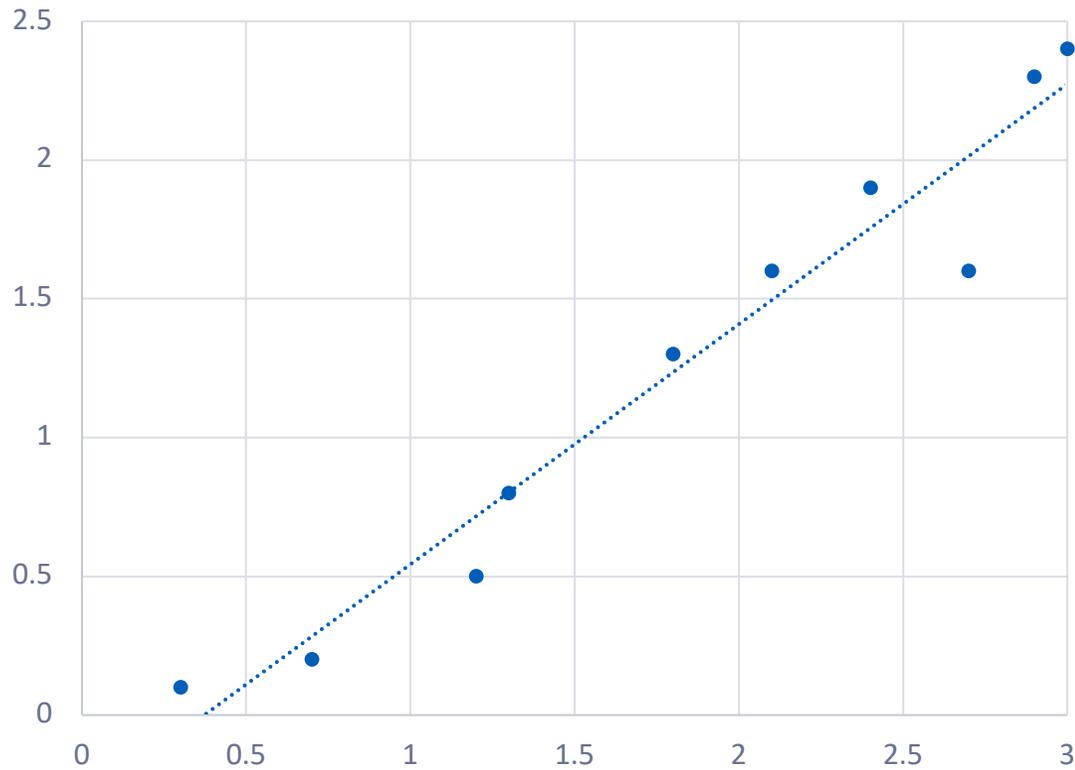
# Graphs



exact values easy to determine



# Regression models and residuals



# Summary and test statistics



# Explanation

Why do you want to export this data?

Where do you intend to publish further?

Why do you need to continue your analysis outside of the RE?

What data are you exporting?

Explain unlabelled columns and non-standard formats



# What should go through the Airlock?



Figures



Statistics or numbers for your text



Tables



Notes on the data

### 3. Working with LibreOffice in the RE

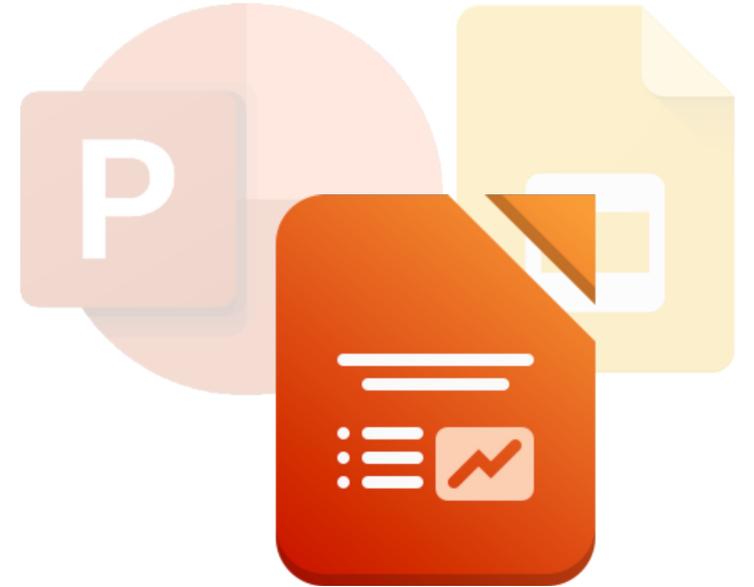
# Office tools



LibreOffice  
Writer



LibreOffice  
Calc



LibreOffice  
Impress

# Drawing tool

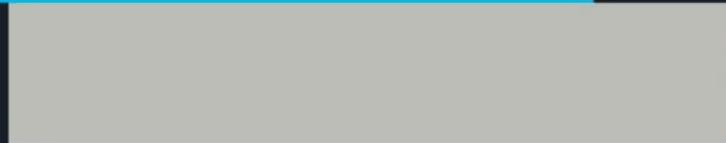
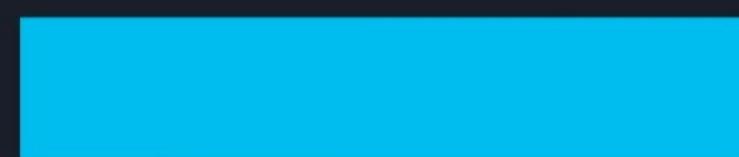


- Draw figures and illustrations
- Can embed images in Writer and Impress

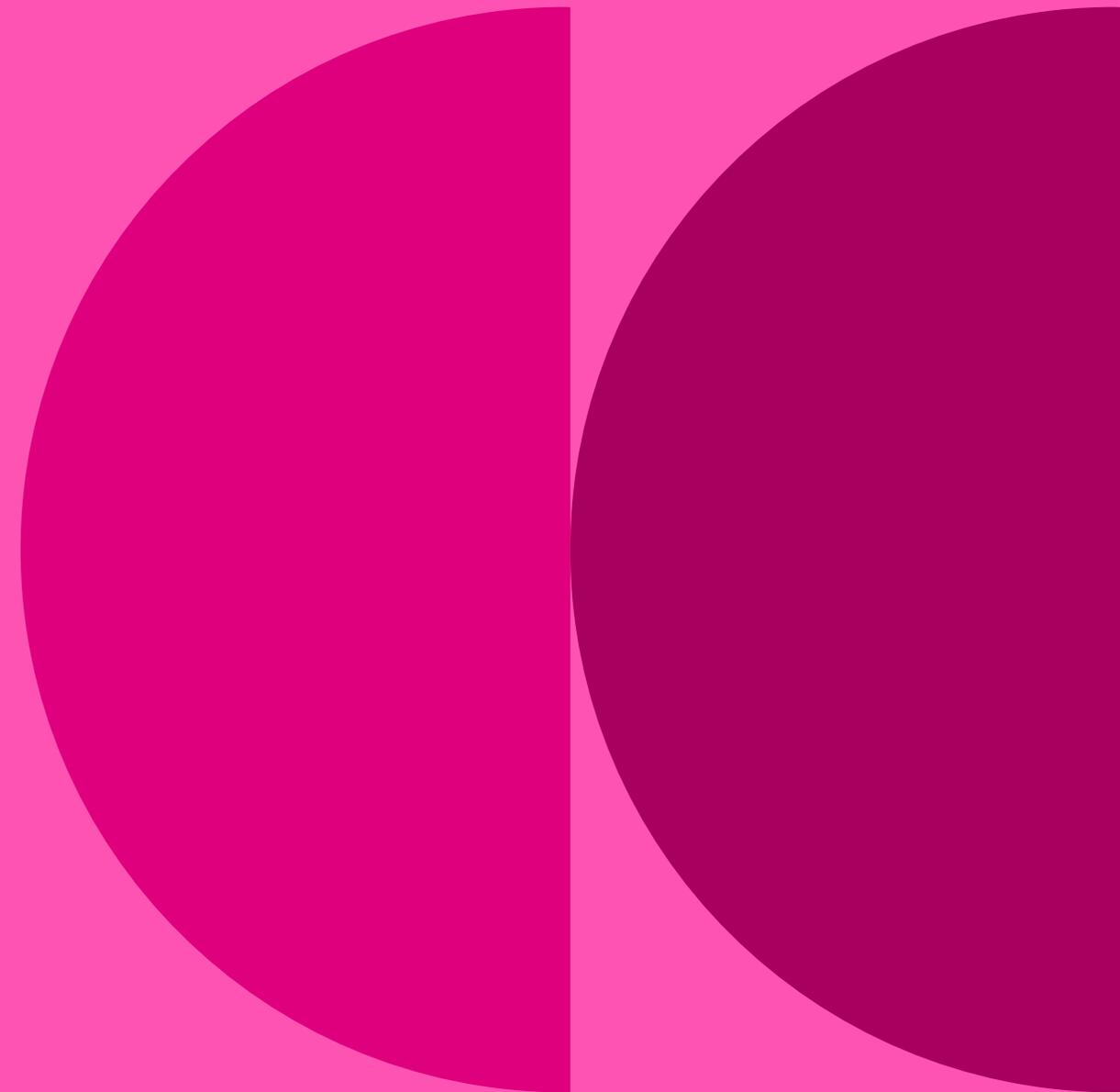
# Using LibreOffice demo



- Computer
- eperry's Home
- Trash
- Text Editor
- IVA 2.0
- Rocket Chat
- Gvim
- Document Viewer
- Airlock
- Data Discovery
- Open Targets
- IGV Browser
- RE Messages
- LibreOffice
- Labkey
- R
- Research Environment Documentation
- Participant Explorer
- Terminal Emulator
- Visual Studio Code
- Welcome Pack
- Panel App
- Git GUI
- RStudio
- Emacs



# 4. Using the Airlock



# Forms in Airlock



Export findings



Export analysis scripts and software



Contact clinical team and/or report potential diagnosis



# Export findings

User group

File(s) to export

List of participant IDs

Description of file content

Research registry ID

Cancer/rare disease

Aim of export

Transfer request justification

Research Network or Discovery Forum

Up to 20 files, up to 5GB

To check for consent for research

Describe everything, particularly unlabelled and non-standard

To check it matches your registered project

Publication, abstract, scientific talk, external analysis

Where you plan to publish and share

# Export analysis scripts and software

User group

File(s) to export

Description of file content

Research registry ID

Transfer request justification

Programming language

Software license

Research Network or Discovery Forum

Up to 20 files, up to 5GB, check to remove any hard-coded data

What the software does and any data included in the scripts

To check it matches your registered project

If you plan to publish and share

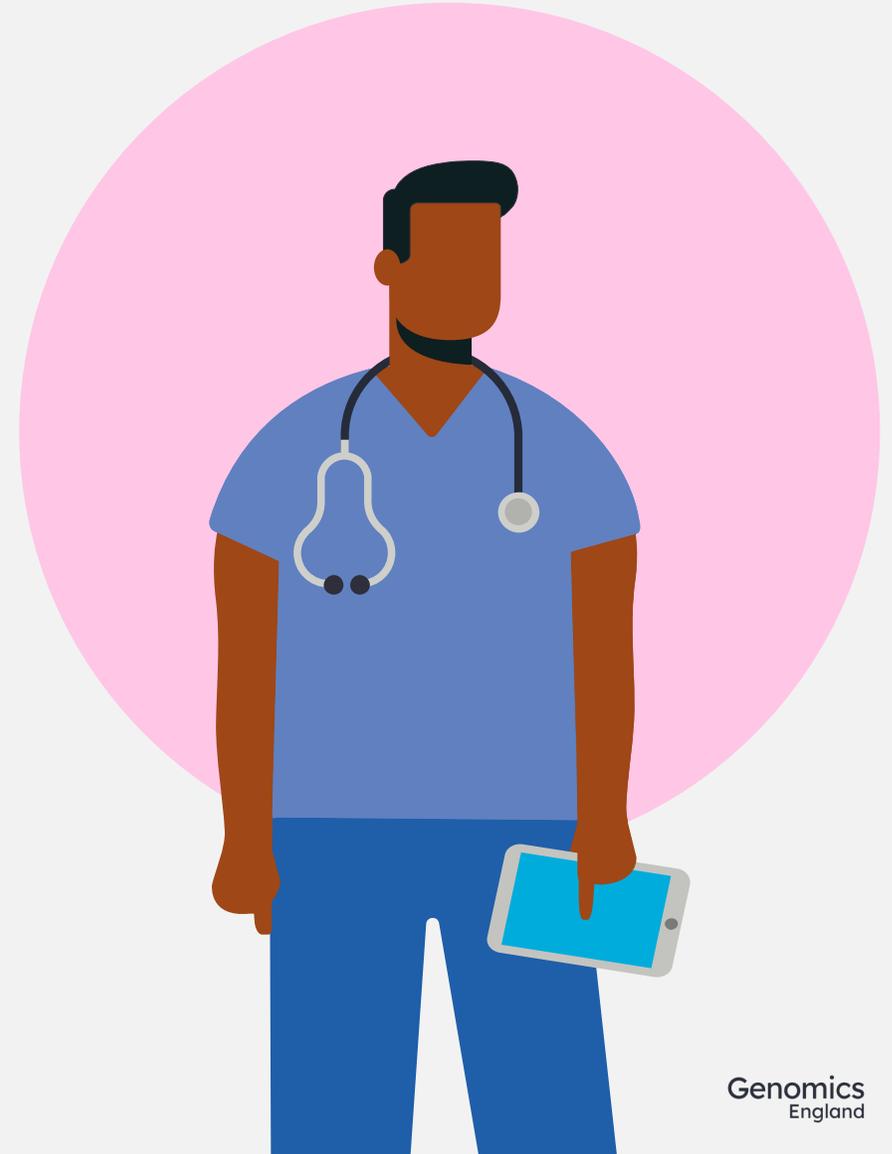
If multiple, select the most common

Who owns the software, if not you

# Clinical collaboration

Examples of collaboration include:

- Request for patient consent to publish paper
- Inviting a clinician as a co-author on a paper
- Request for further health information or clinical tests
- To discuss with the clinician a potential diagnostic variant
- To offer laboratory tests to investigate in more detail whether a particular variant is likely to be diagnostic or not



# Clinical collaboration

User group

Research registry ID

What is your expertise in this area?

Do you wish to be put in touch with the clinical team?

Why?

Research Network or Discovery Forum

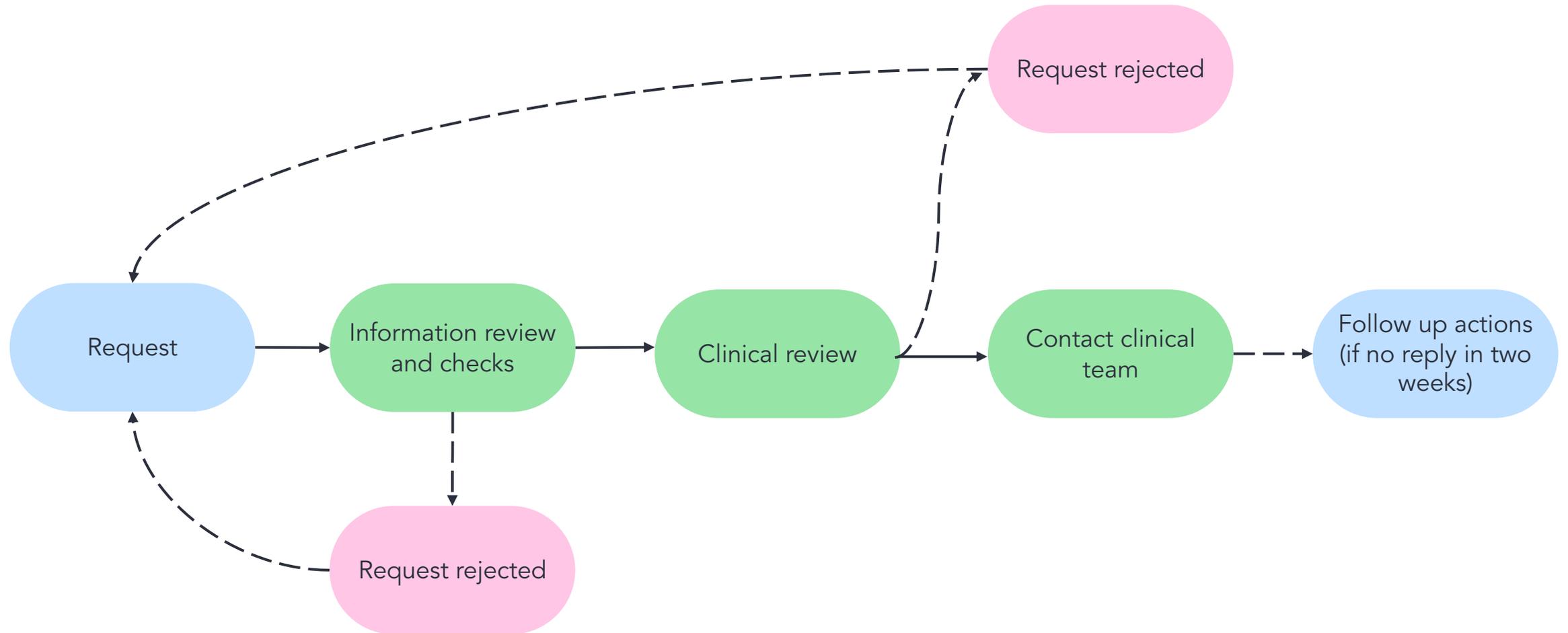
To check it matches your registered project

What research has led to this finding? Who is your PI?

Yes or no, you can also use this form for submitting diagnostic discoveries

What is your reason for contact? What do you want from the clinician, and what are you offering them?

# What happens to my request?



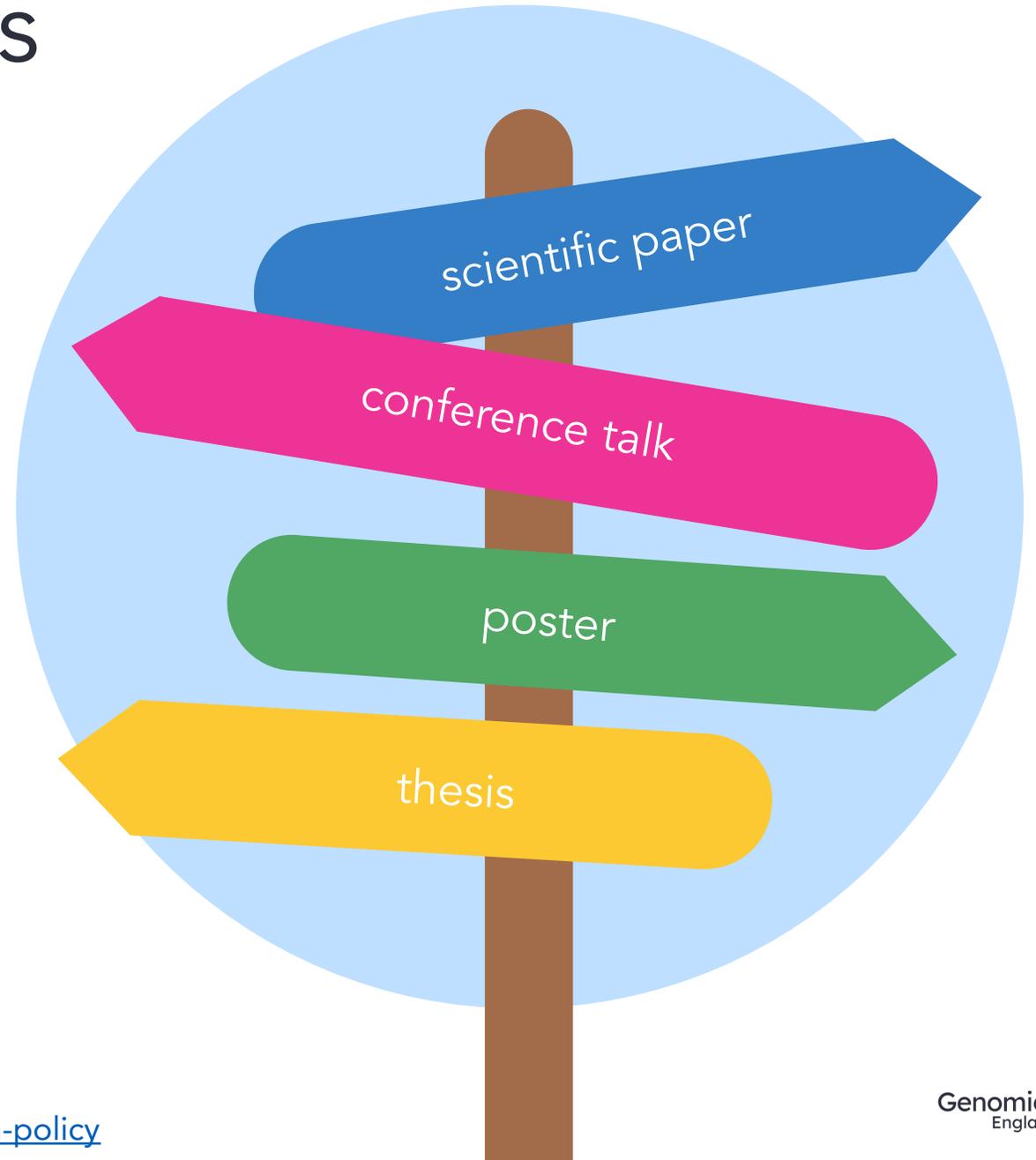
# Using Airlock demo

The desktop environment features a grid of application icons on the left side, including: Computer, Firefox, Panel App, RStudio, eperry's Home, Git GUI, Participant Explorer, Terminal Emulator, Old Firefox Data, GVim, R, Text Editor, Airlock, IGV Browser, RE Messages, Visual Studio Code, CloudOS, IVA 2.0, Research Environment Documentation, Welcome Pack, Document Viewer, Labkey, Research Registry, and Trash. In the center, there is a large grey circle. On the right side, the Genomics England logo is displayed, consisting of the text "Genomics" and "england" next to a stylized map of the United Kingdom made of dots. A mouse cursor is positioned over the logo area. At the bottom right, there are two overlapping rectangular shapes, one cyan and one grey.

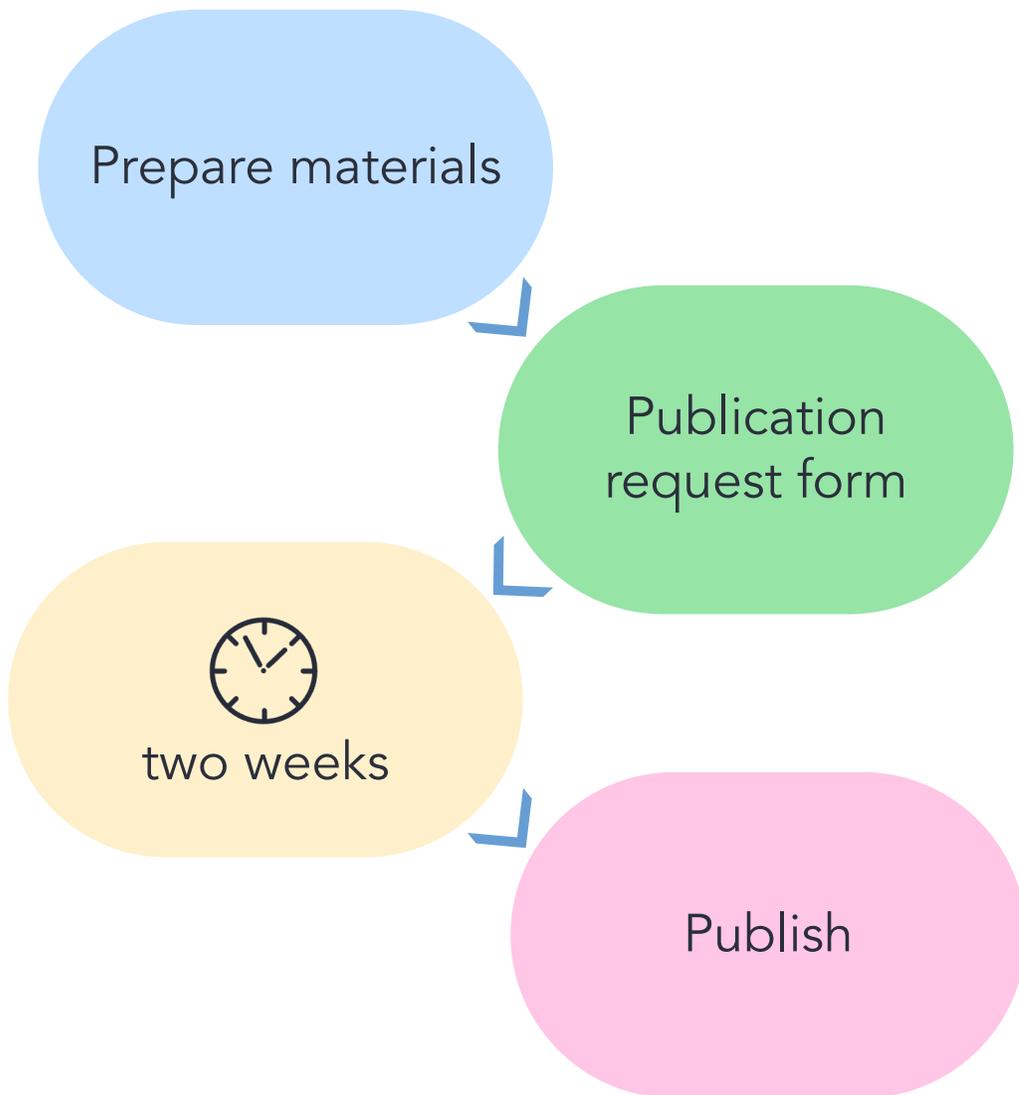


# Publication guidelines

- Must follow Airlock rules...
- ...or have a documented clinical collaboration with consent to publish
- Individual Genomics England help in the acknowledgments
- Credit in acknowledgement or methods (copy set text)
- Request permission for publication



# Approval for publication



# Publication request demo



# Research Portal

Home > Research Registry > Browse the Research Registry

## Browse the Research Registry

### In this section

[Browse the Research Registry](#)

[Search the Research Registry](#)

All Research Registry projects are visible to logged in Research Network members and Discovery Forum members. Other users, including anyone who isn't logged in, will be able to see limited details of Research Registry projects (title, profile lead, communities and lay summary), and will only be able to see approved projects.

To view more details on a project, click the *Project ID* to go to the details page.

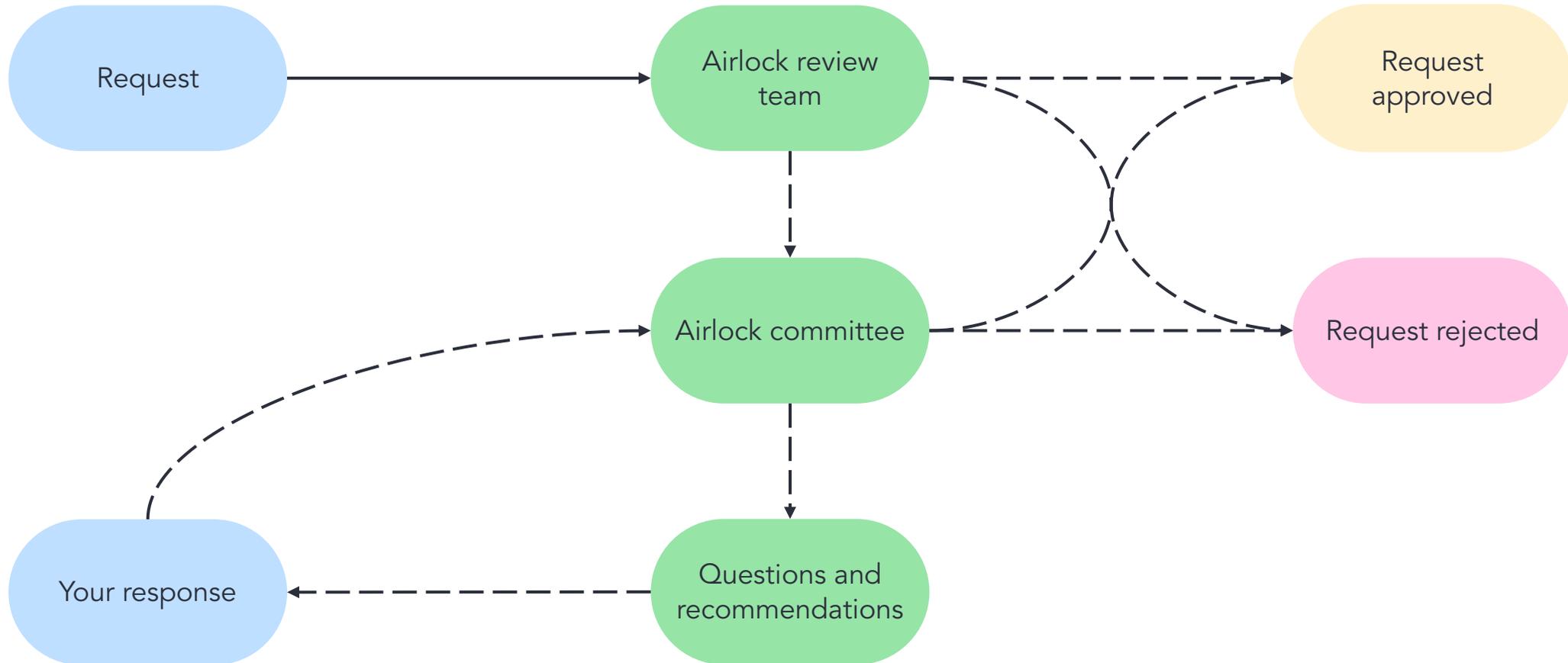
**Are you writing a paper on your work in the Research Environment?** We have some requirements on acknowledgement text and authorship and ask that you send it to us for review before submission. Please visit our [publications page](#) for more information.

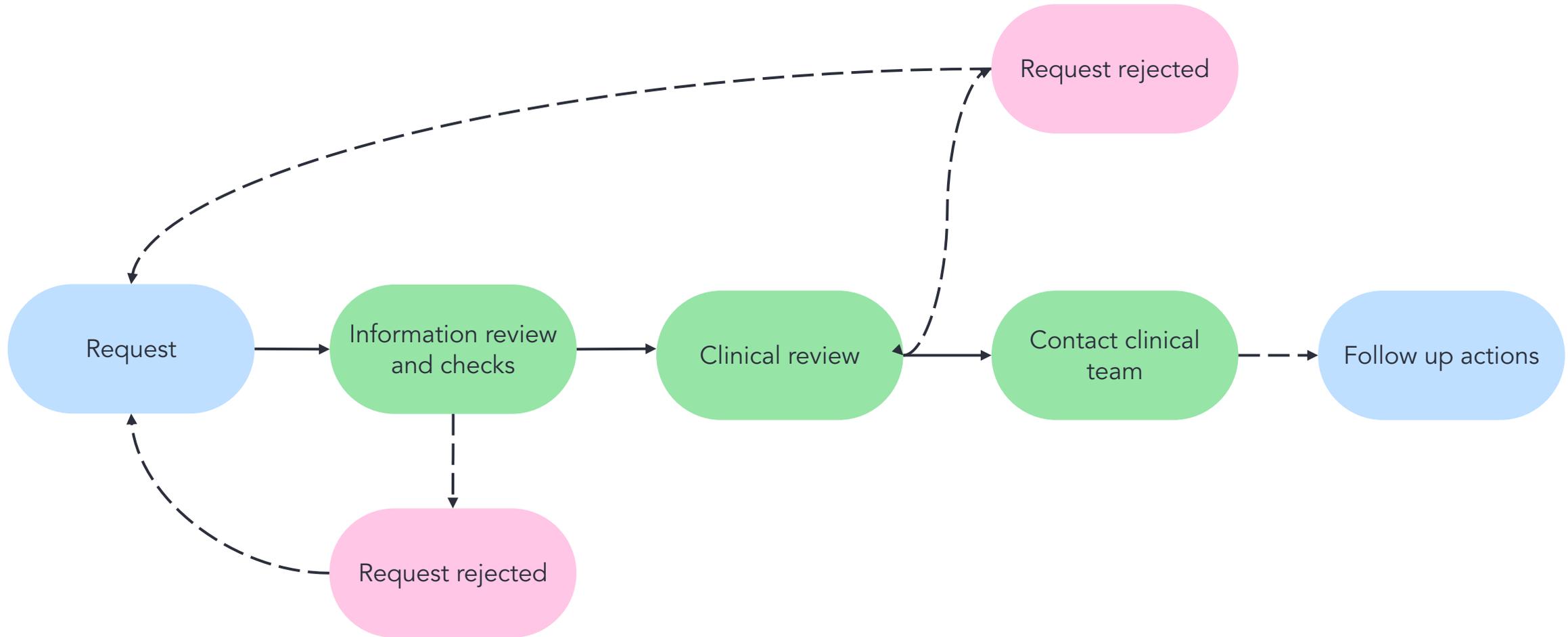
Project ID ↑	Project date	Title	Lead	Primary Communities	Status
4	02/05/2018	Functional Genomics of CAKUT	Melanie Chan		Approved

## 6. Sharing data with reviewers and examiners in the RE

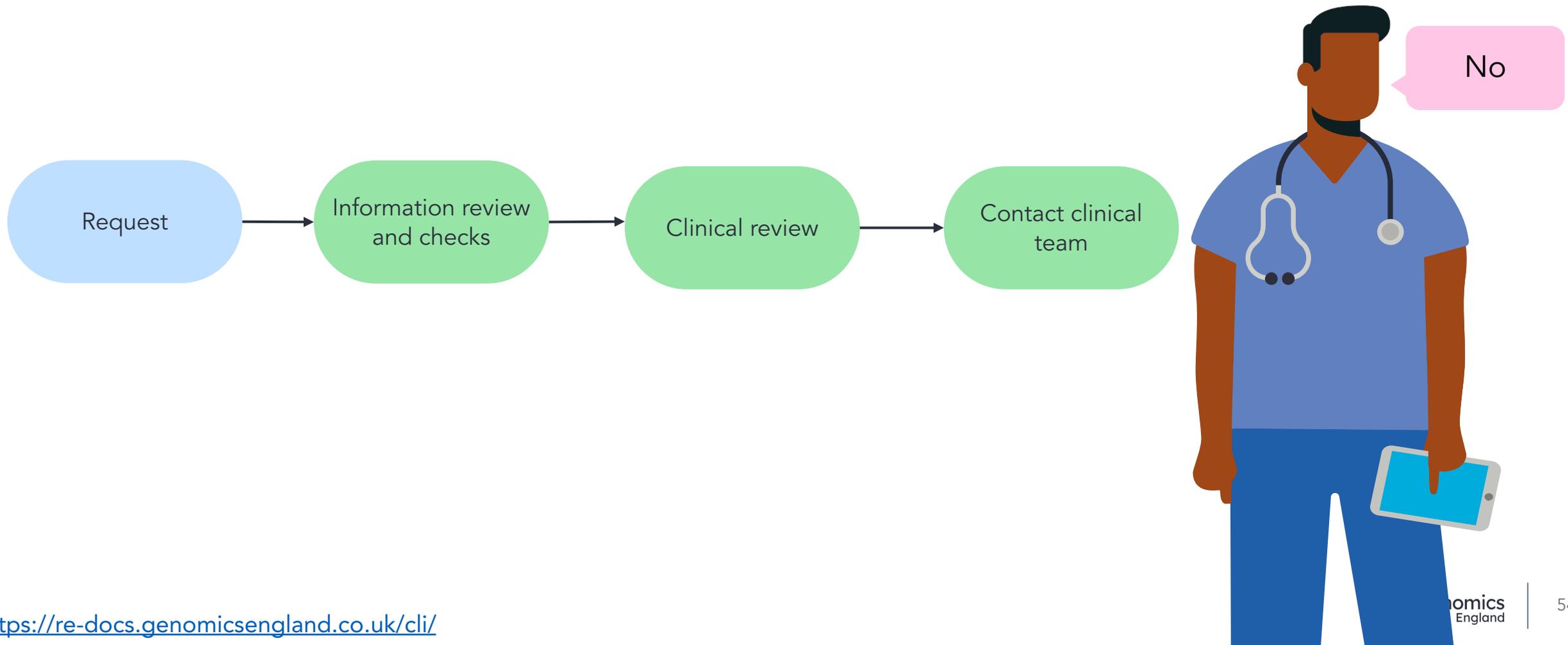
# Airlock request rejected



# Clinical collaboration request rejected



# Clinician chooses not to collaborate



# Show your examiners your results in the RE

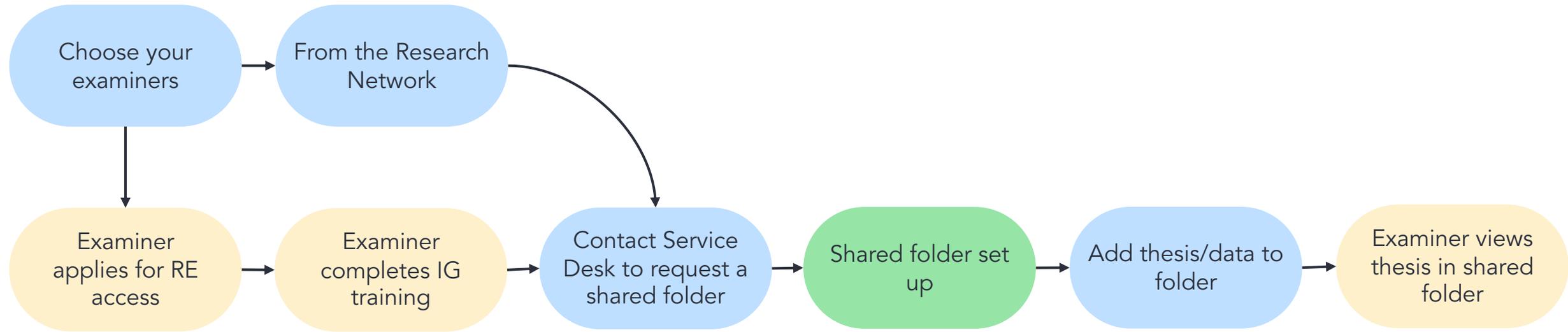


Redacted thesis sent out

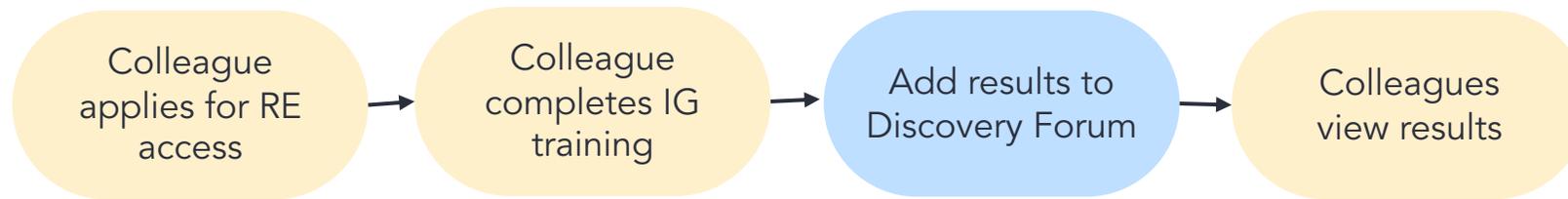


Full thesis inside RE

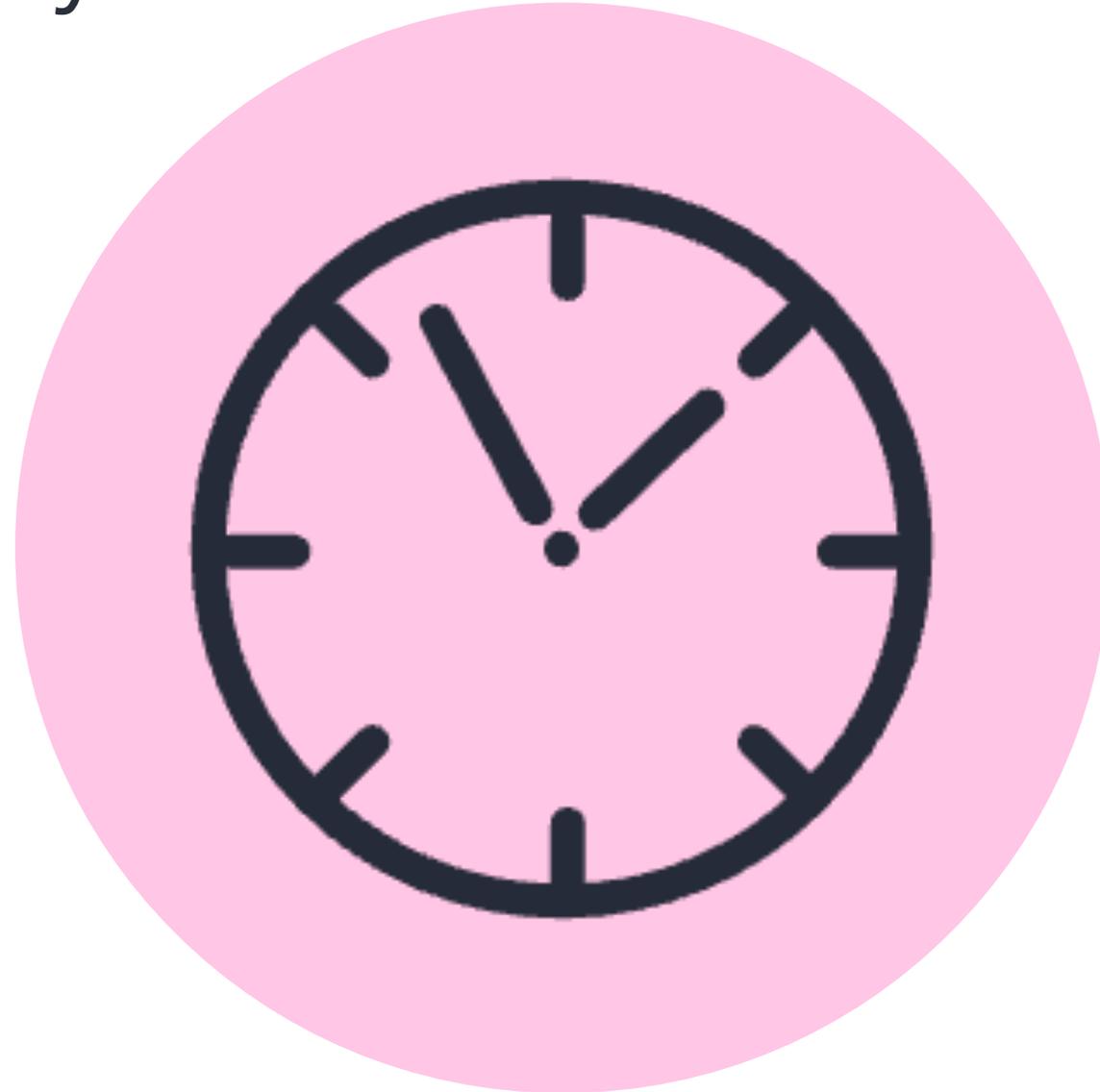
# Show your examiners your results in the RE



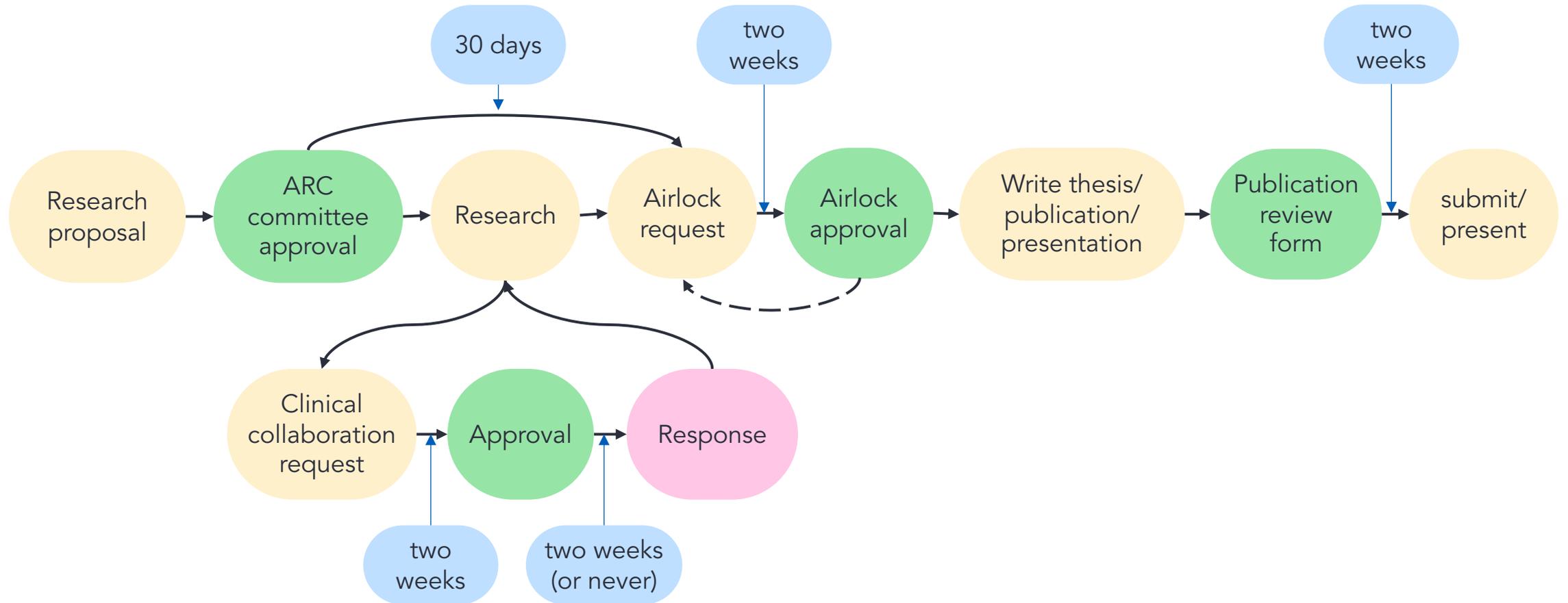
# Share your results with your company in the RE



# Key takeaway



# Full timeline

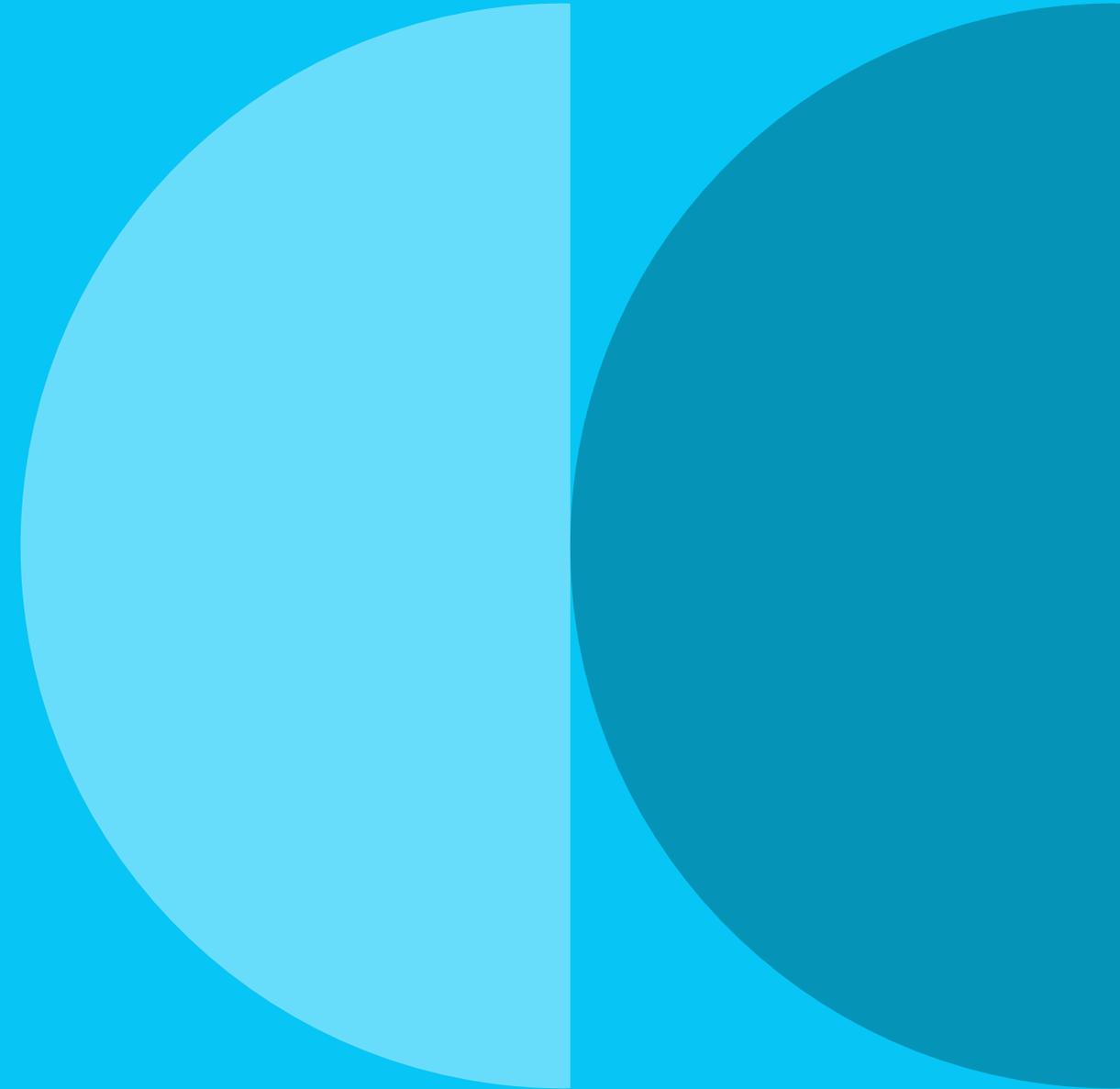


# Top tips

- Register your research project ASAP
- Get your Airlock/publication requests approved first time
  - Follow the rules
  - Clear labels and descriptions
- Work with collaborators/lab-mates/supervisors inside the RE rather than post-Airlock
- Include delays in your project plan



# 7. Help and questions



# Getting help



Check our documentation:  
<https://re-docs.genomicsengland.co.uk/>  
Click on the documentation icon in the environment



Contact our Service Desk:  
<https://jiraservicedesk.extge.co.uk/plugins/servlet/desk>

# Questions



All your  
microphones  
are muted



Use the Zoom  
Q&A to ask  
questions



Upvote your  
favourite  
questions: if  
we are short  
on time we  
will prioritise  
those with the  
most votes

# Training sessions 2024

8/10

What tools and workflows should I use to fulfil an overall goal?

12/11

Running workflows on the HPC and Cloud

10/12

Introduction to the RE



Materials  
from past  
training all  
online

# Training sessions 2025

3<sup>rd</sup> Tuesday every  
month

Introduction to the RE

21/1

18/2

18/3

15/4

20/5

17/6



Materials  
from past  
training all  
online

# Training sessions 2025

14/1

Using the Research Environment for  
clinical diagnostic discovery

11/2

Importing data and tools to use in  
the RE

11/3

Working with R in the RE

8/4

Working with python in the RE

13/5

Building cancer cohorts and survival  
analysis

10/6

Building rare disease cohorts with  
matching controls

<https://re-docs.genomicsengland.co.uk/upcoming/>



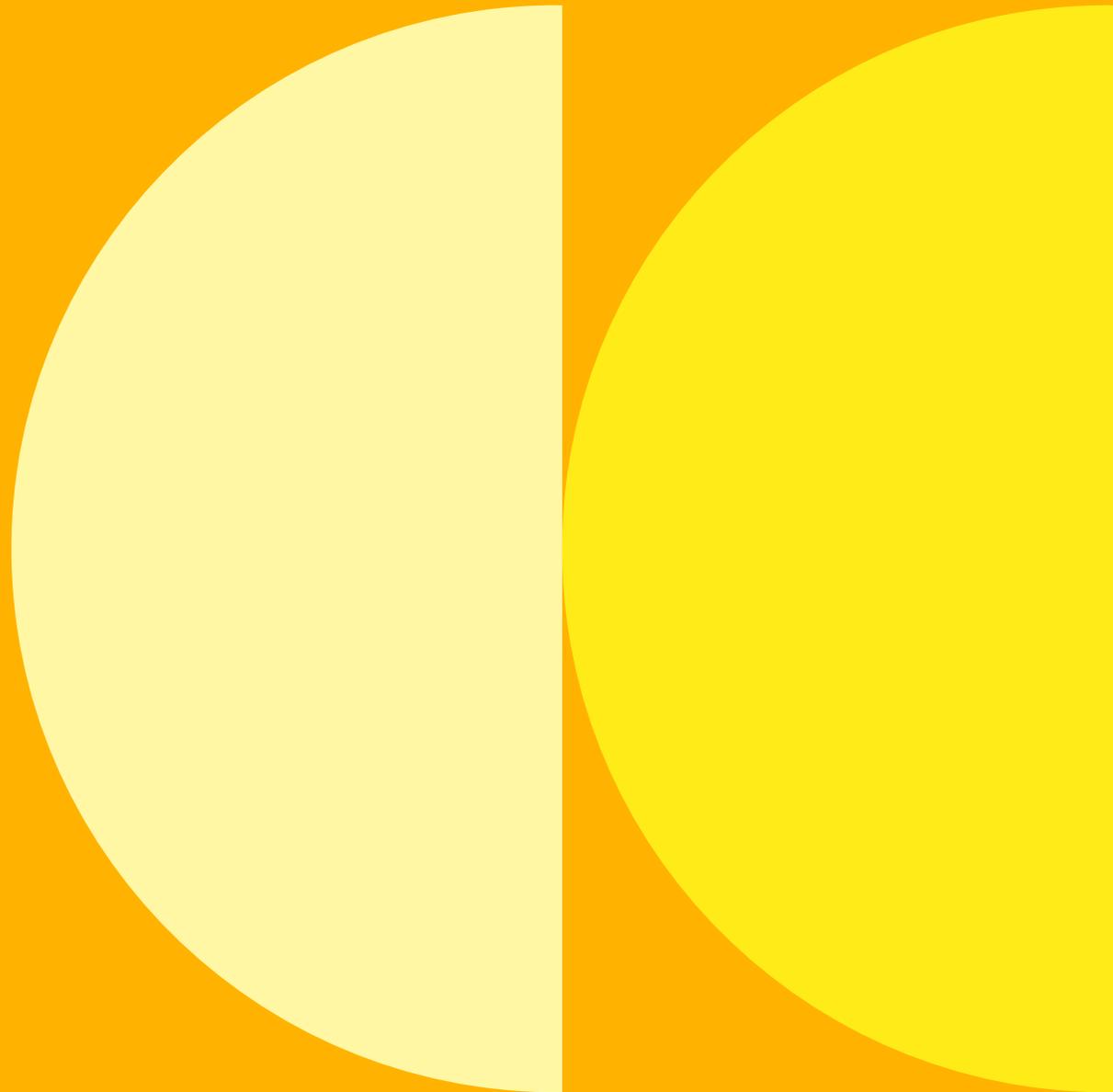
Materials  
from past  
training all  
online

# In-person training day

- For any registered RE users
- 20<sup>th</sup> November 2024
- Held in our offices at Canary Wharf
- Hybrid attendance
- Hands-on practicals and exercises



# Feedback



# Thank you

Visit: <https://re-docs.genomicsengland.co.uk/>