

Using GEL data for publications and reports

Emily Perry

14th October 2025



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



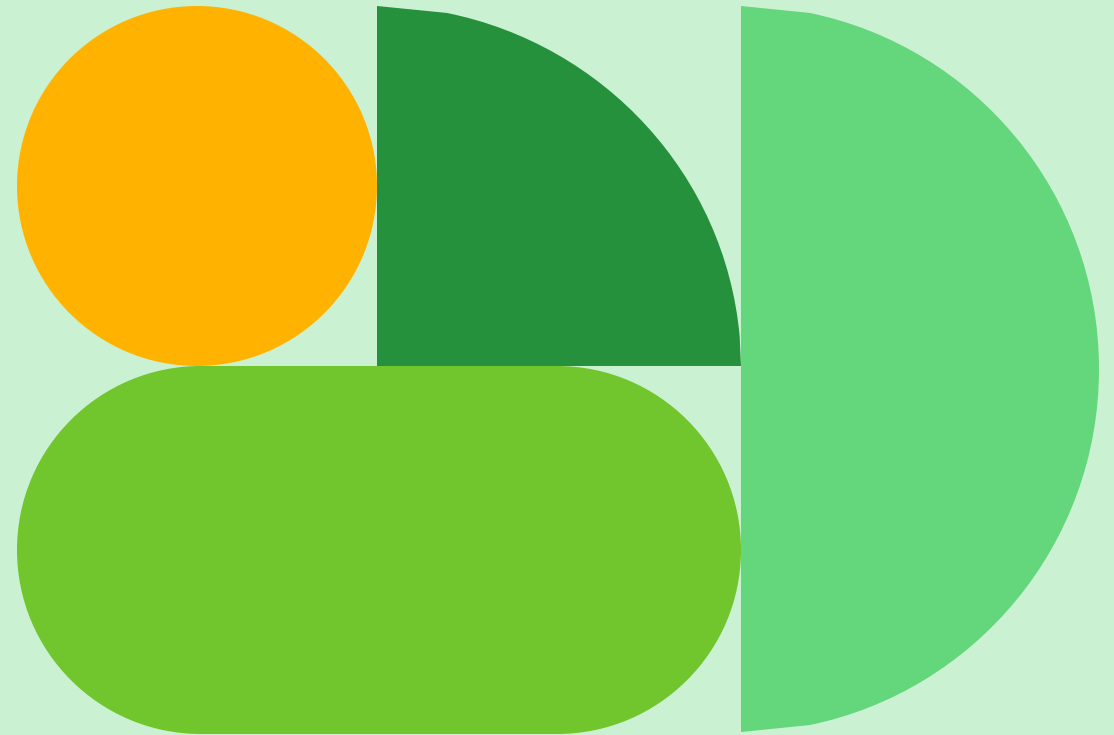
**Pete
O'Donovan**
Airlock Lead



Cerys Ballard
Interim
Research
Management
Lead

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



Research and generate
results inside RE

Fill in a form to request
export

Access results outside the
RE

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

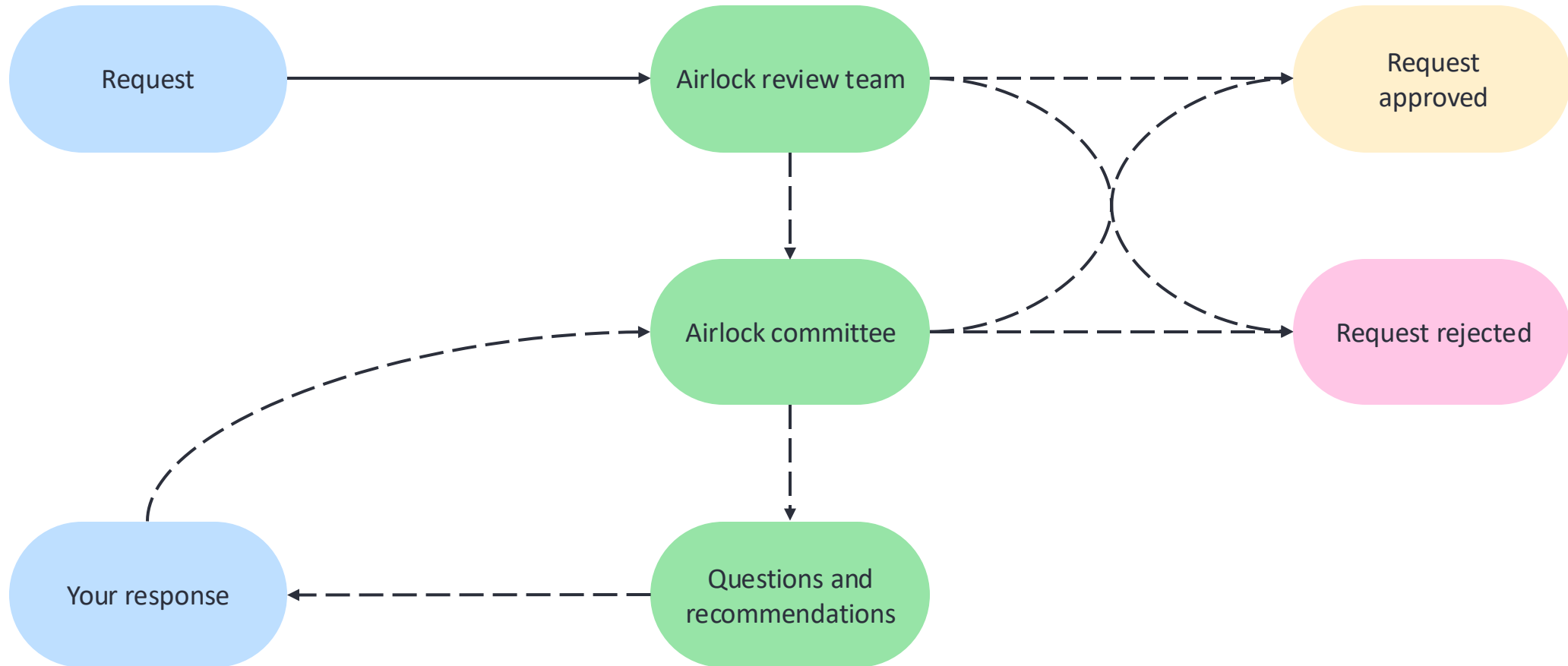
The Airlock

Data in the RE

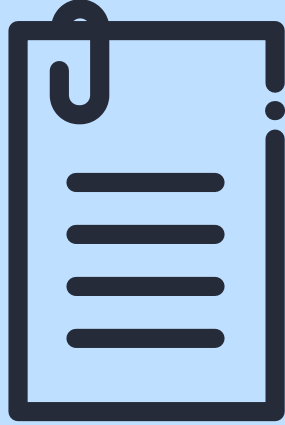


Outside world

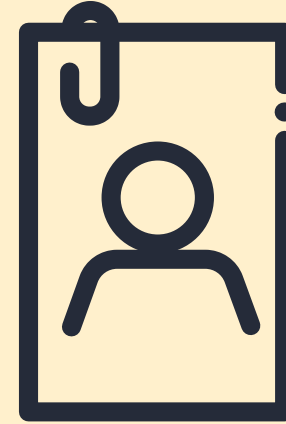
What happens to my request?



Airlock rules

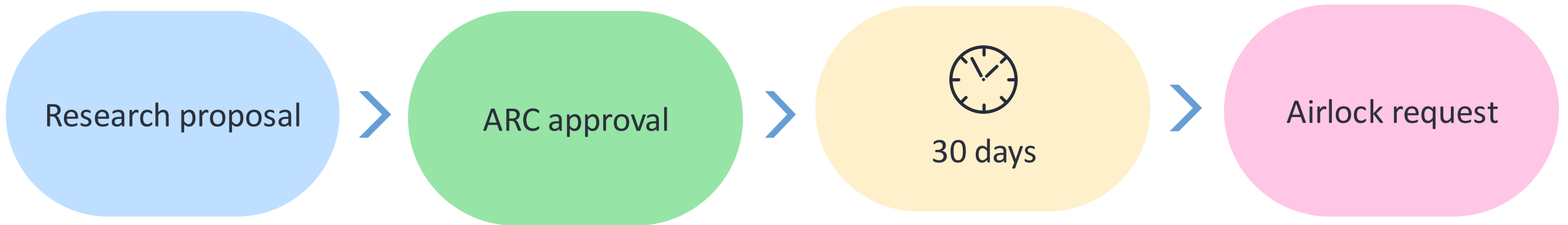


Approved research
project



Participants cannot be
identified

Approved research project



Make your research proposal



Make your research proposal

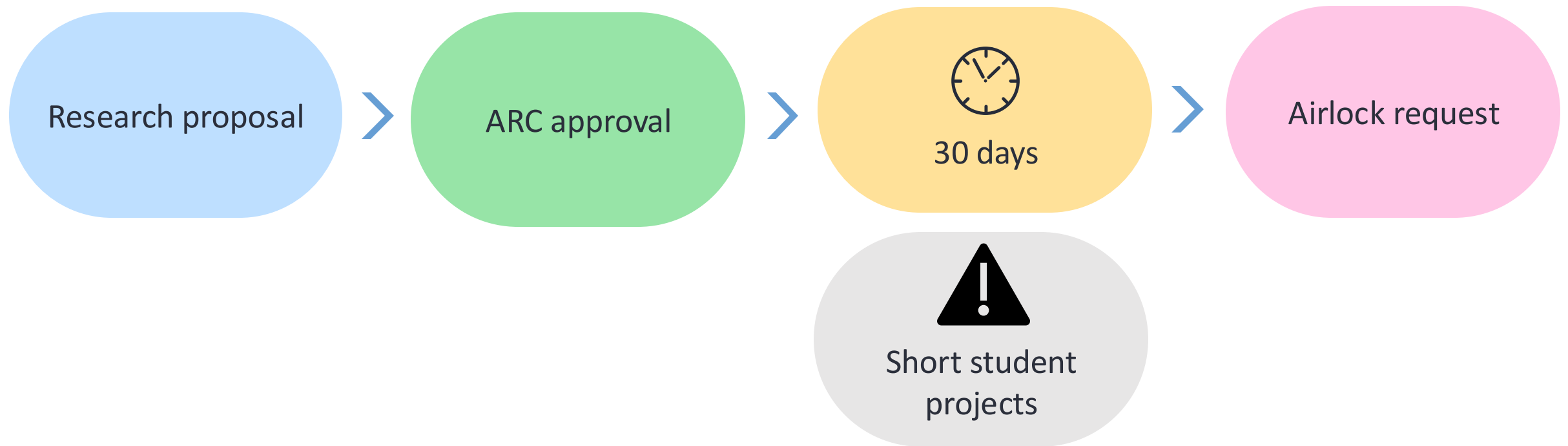


<https://research.genomicsengland.co.uk/research-registry/submit-research-project>

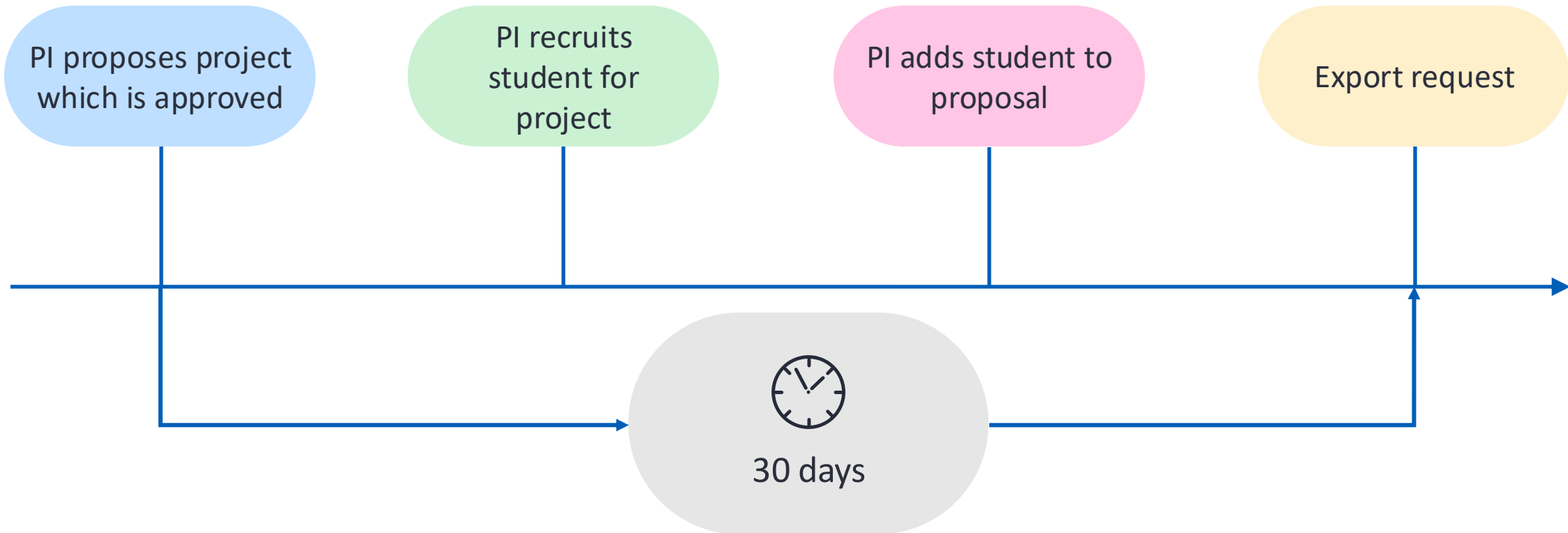


Discuss with your partnerships director

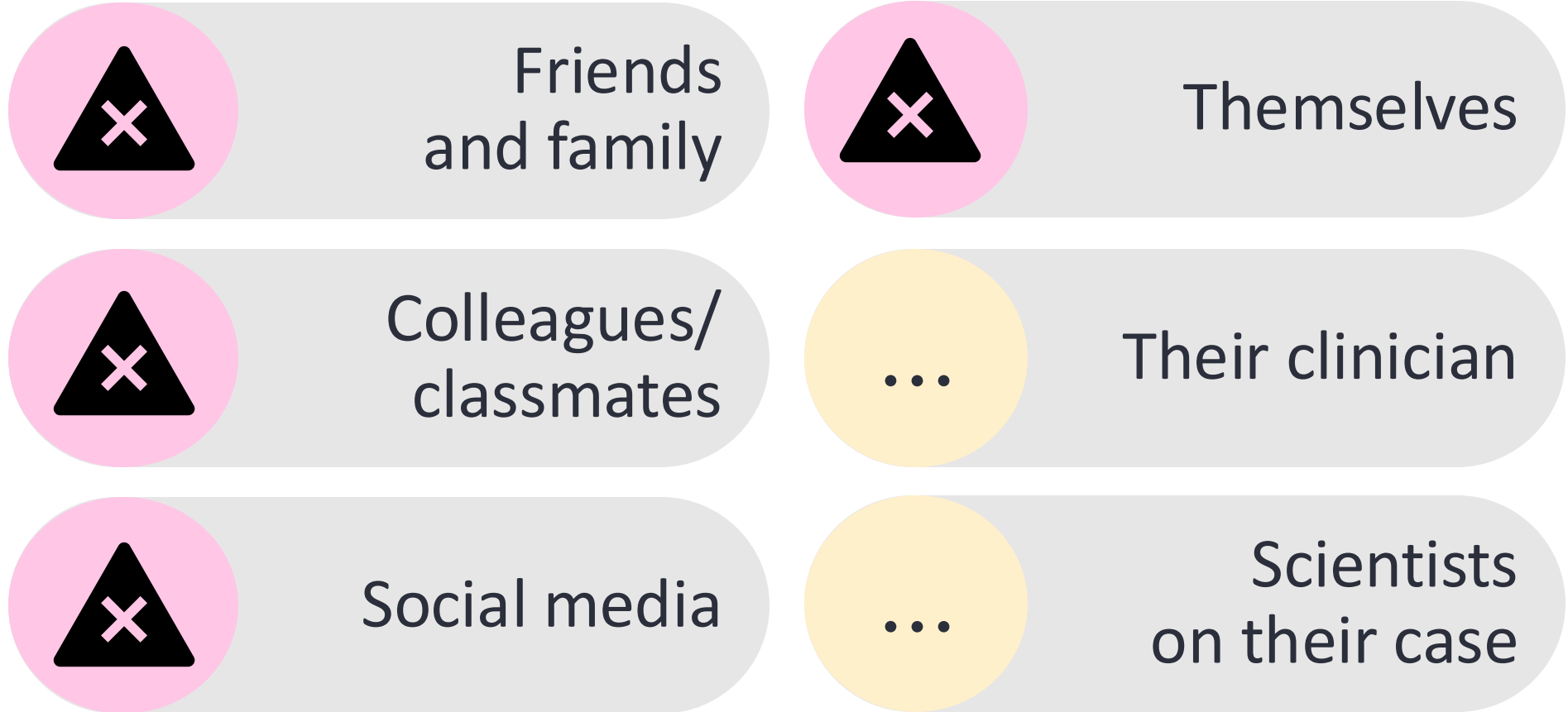
Approved research project



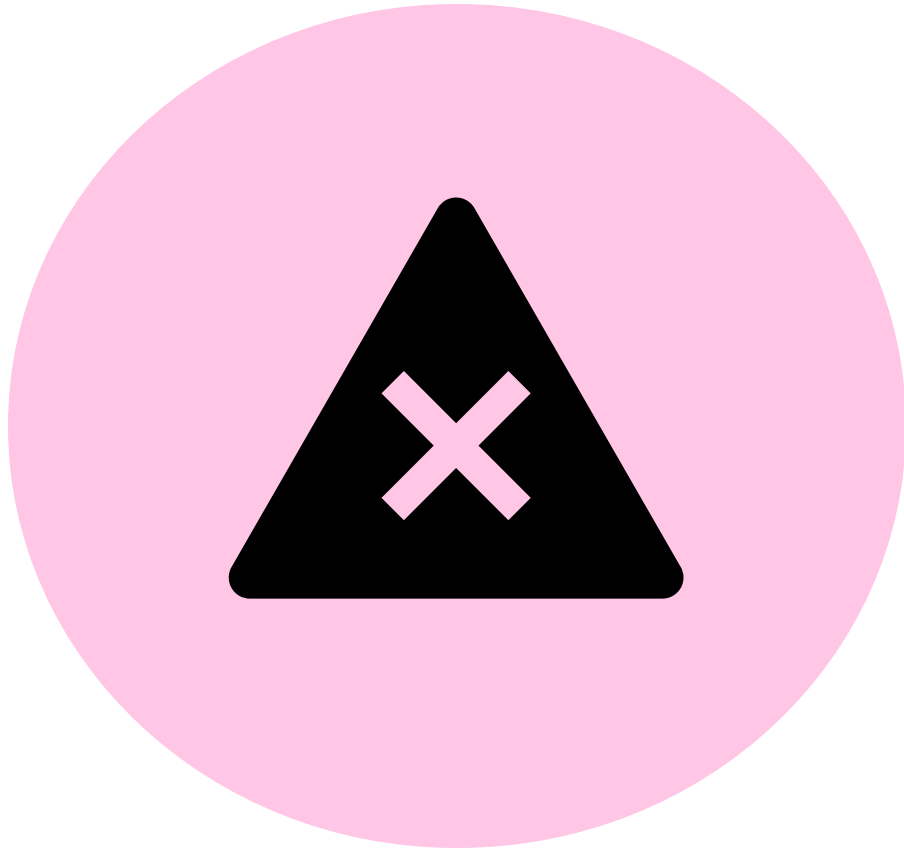
Approved projects for students



Participants cannot be identified - to whom?



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

CCTCAGCCTCCCAAACTGTTGGGATTACAG
GCGTGAGCCGGAATTTTA
TCGTGGGATAGTT
TTTGA
TAAAGCTA
ATAAAA
TGCCGTGT
TACAGAAATG
GATTAAATGCCT
TAACAAAACAGATCTTA
AGCATTTTTTTCCTATGATCTTTAACTGT
TCTGGGTCACAAATTTGTCTGTCACTGGTT



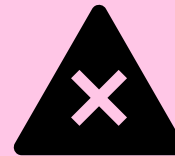
chr1 12345678 A T
chr
chr11
chr12 34567890 C A



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

?

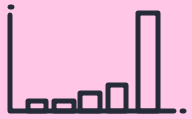
is this something where
0 should be possible?

min/max is also the
mode

?

does this correspond
to a known group?

Percentiles and medians



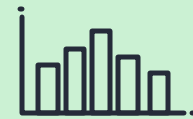
skewed distribution



binomial distribution

<5

in any percentile



normal distribution

≥ 5

in any percentile

Means, indices, ratios, indicators



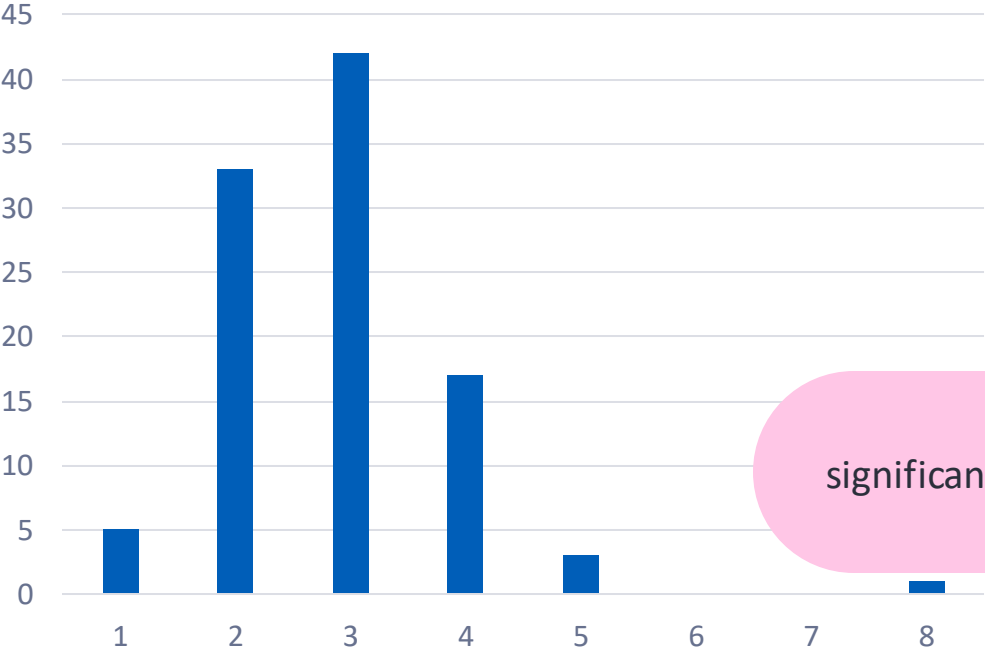
total = 10
only possible values = 1 or 0
mean = 0.7

Value of 0 given away for
three participants

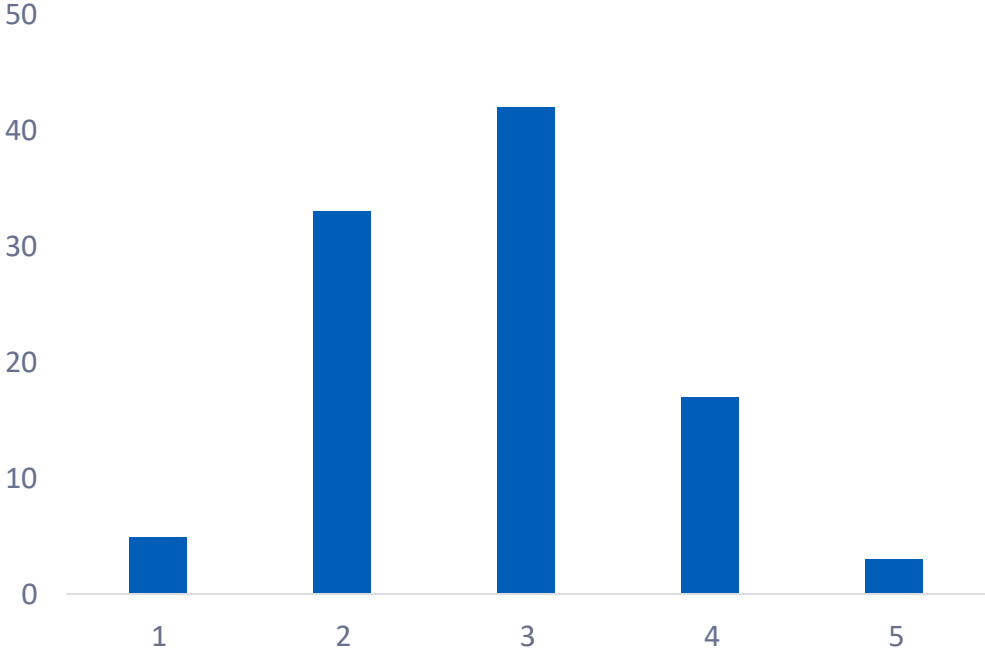


Most cases

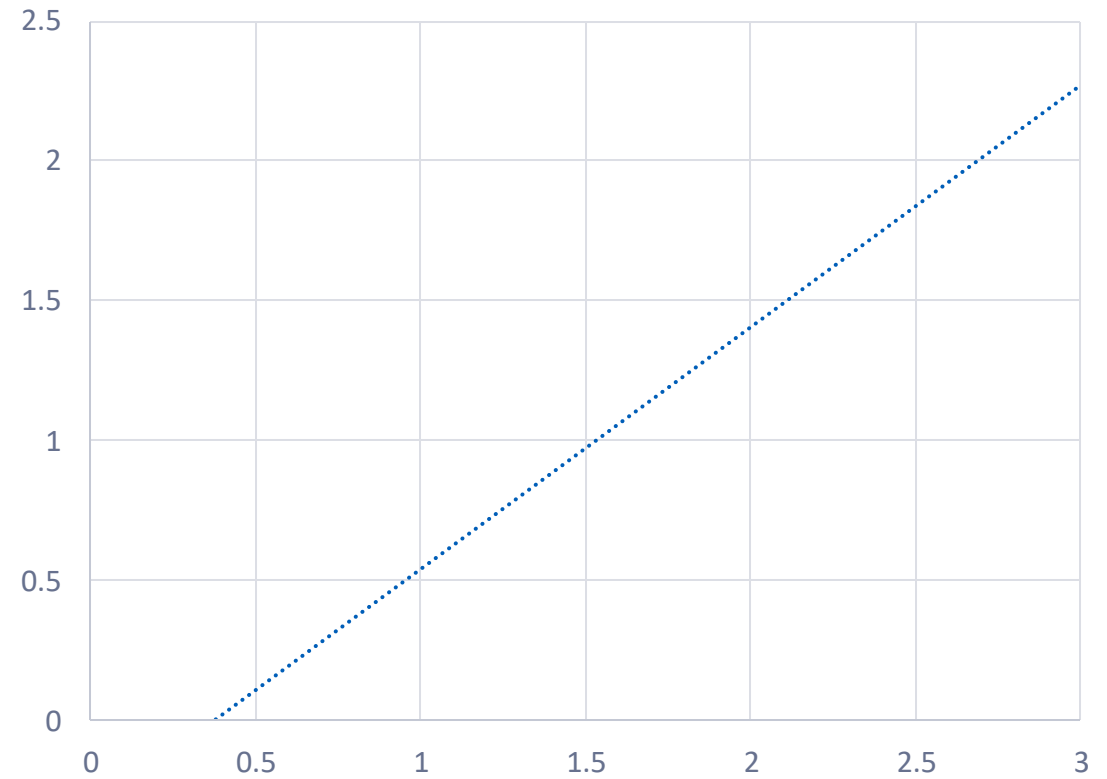
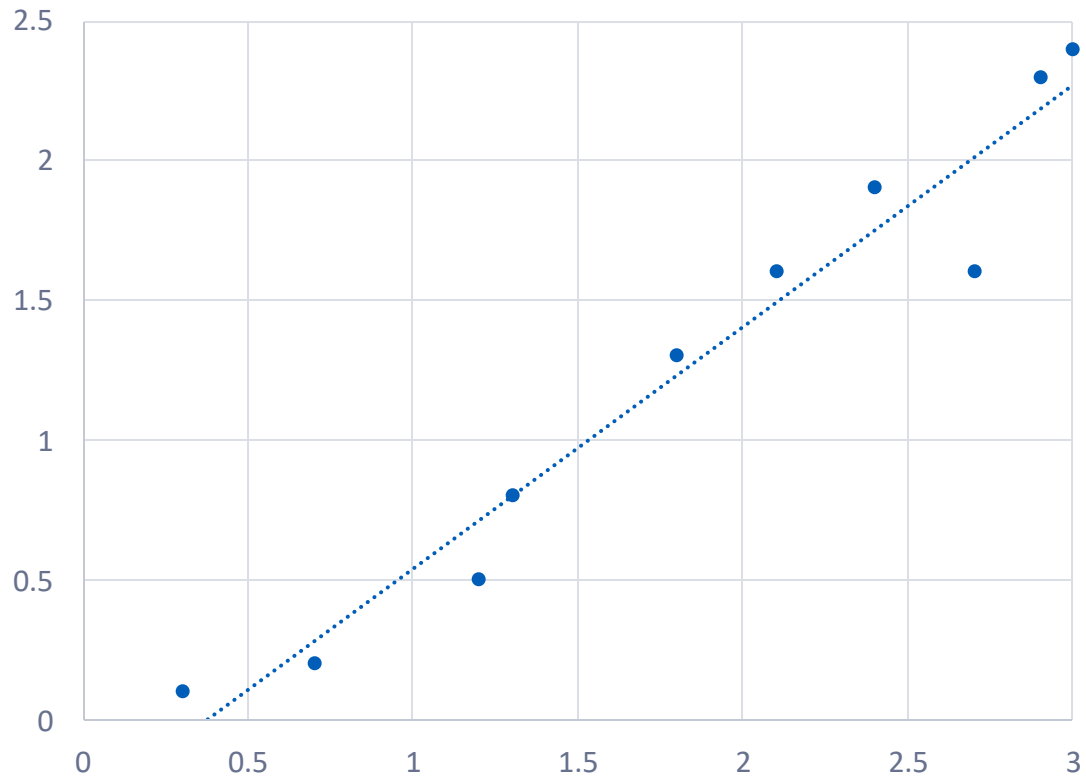
Graphs



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

IGV Browser

LibreOffice

Labkey

R

Research Environment Documentation

Participant Explorer

Terminal Emulator

Visual Studio Code

Welcome Pack

Panel App

Git GUI

Emacs

Research Registry

RStudio

Genomics
england

Applications

Places

System

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 – can be zipped

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

Research Network or Discovery Forum

File(s) to export

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

Description of file content

What the software does and any data included in the scripts

Research registry ID

To check it matches your registered project

Transfer request justification

If you plan to publish and share

Programming language

If multiple, select the most common

Software license

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

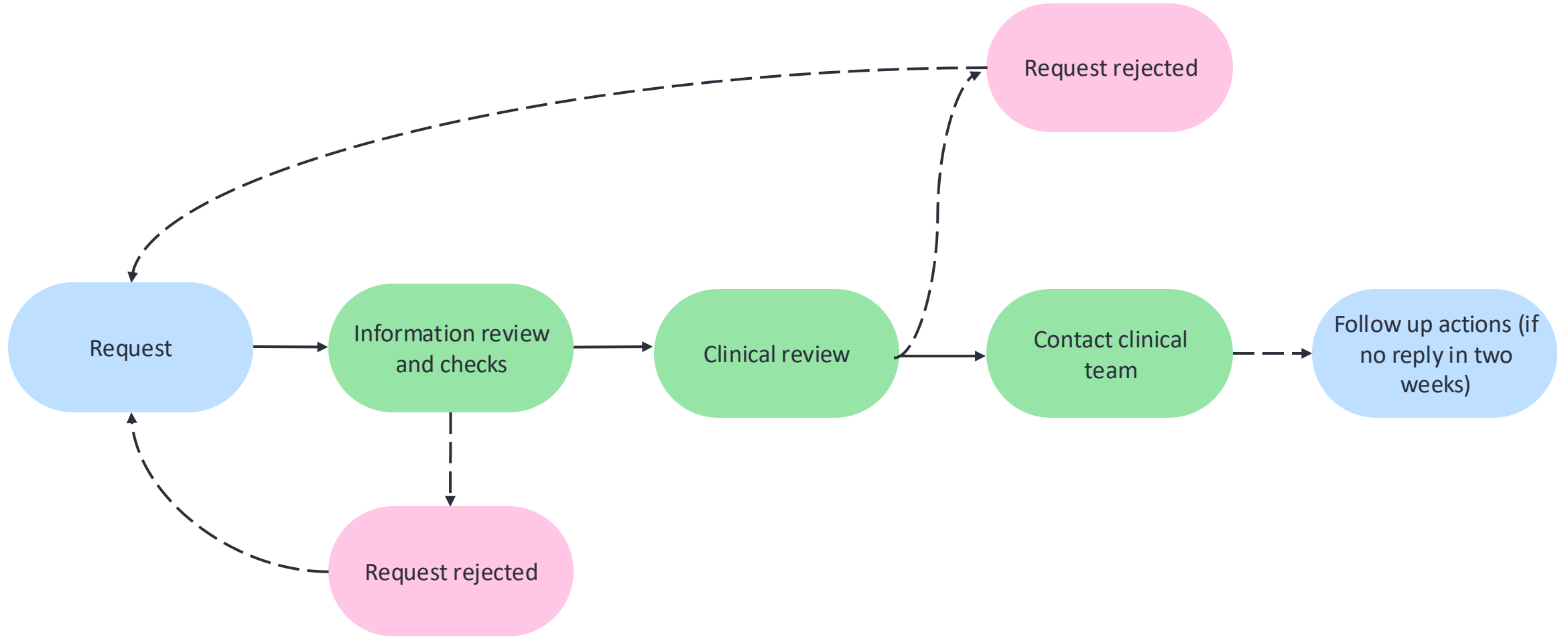
To check it matches your registered project

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?

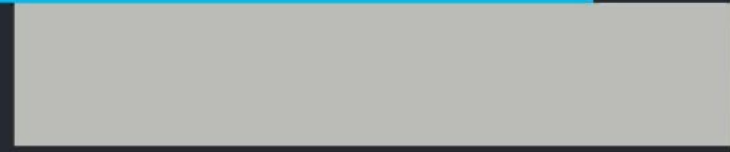
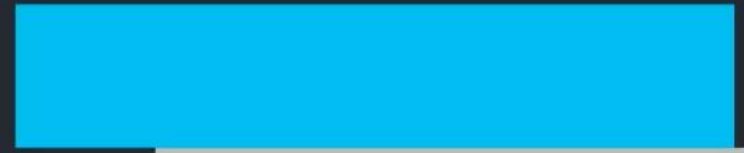
What happens to my request?



Using Airlock demo

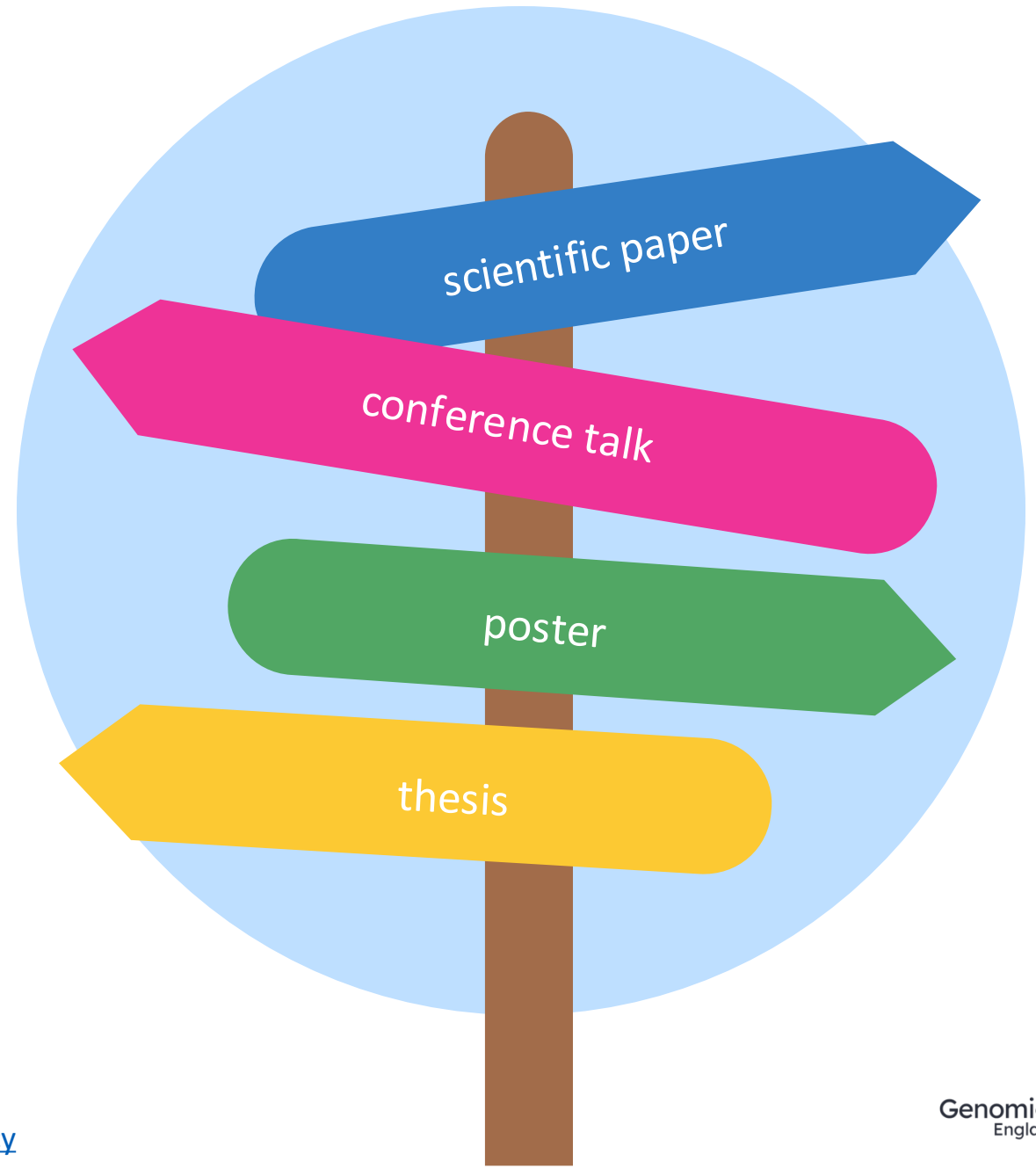


- Computer
- Firefox
- Panel App
- RStudio
- eperry's Home
- Git GUI
- Participant Explorer
- Terminal Emulator
- Old Firefox Data
- GVim
- R
- Text Editor
- Airlock
- IGV Browser
- RE Messages
- Visual Studio Code
- CloudOS
- IVA 2.0
- Research Environment Documentation
- Welcome Pack
- Document Viewer
- Labkey
- Research Registry
- Trash
- Emacs
- LibreOffice 7.6
- Rocket Chat
- Ensembl
- Open Targets
- Link to emily

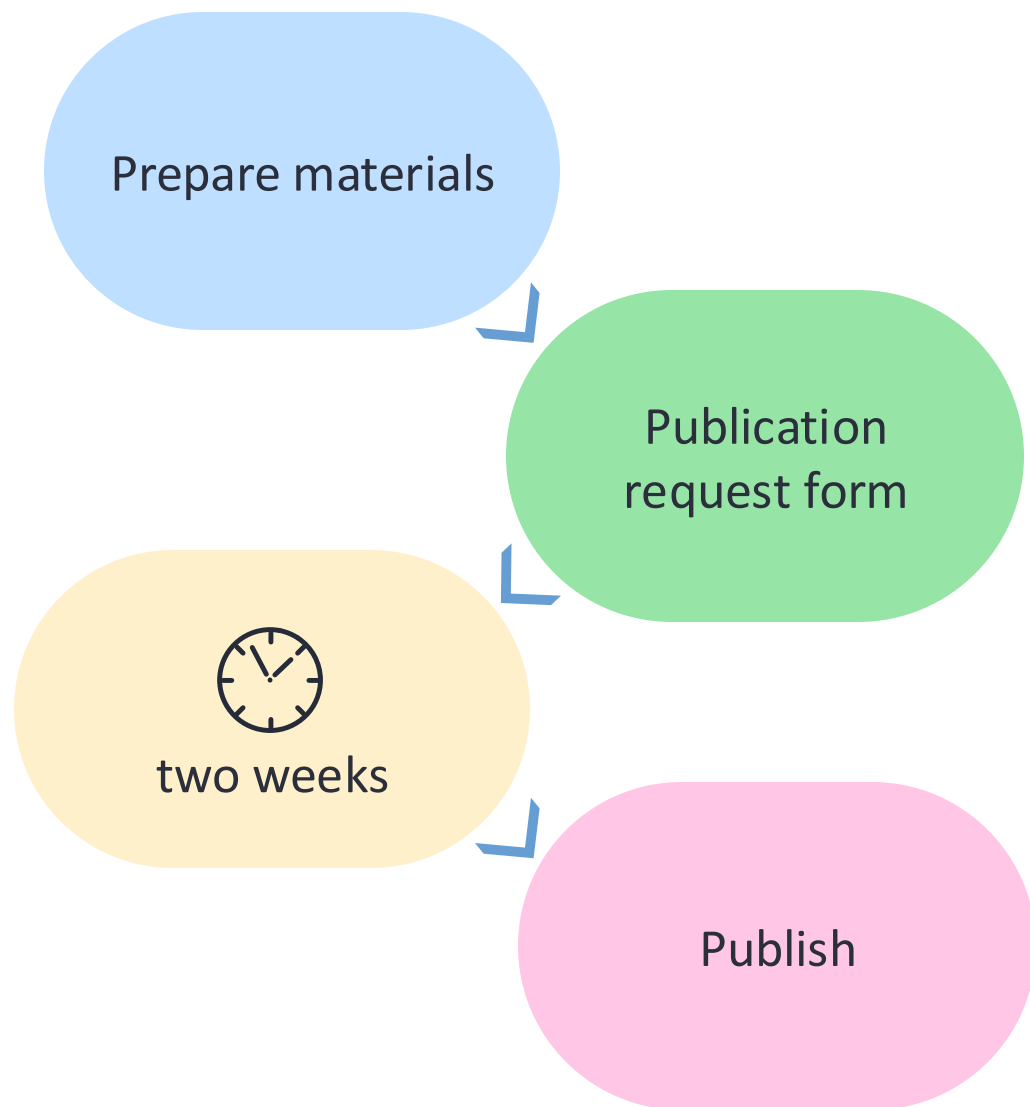


Publication guidelines

- 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



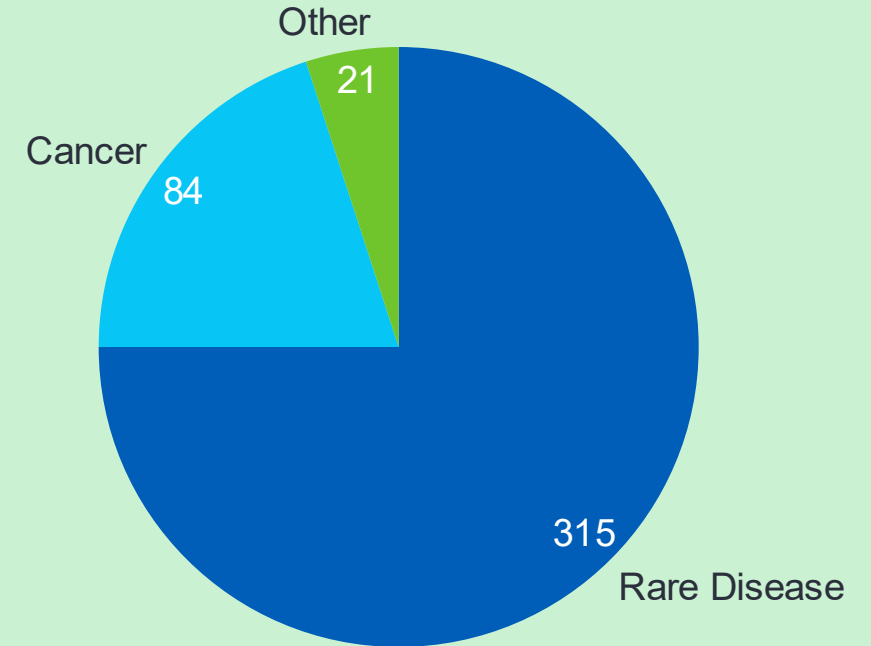
Approval for publication



Publication Policy Updates 2025

- Data Access Statement
- Submitting data to public repositories or knowledgebases (e.g. GeneMatcher, ClinVar, PhamVar etc.)
- Revised acknowledgement text
- Clearer sections on post-publication requirements and what happens when in the event of a policy breach

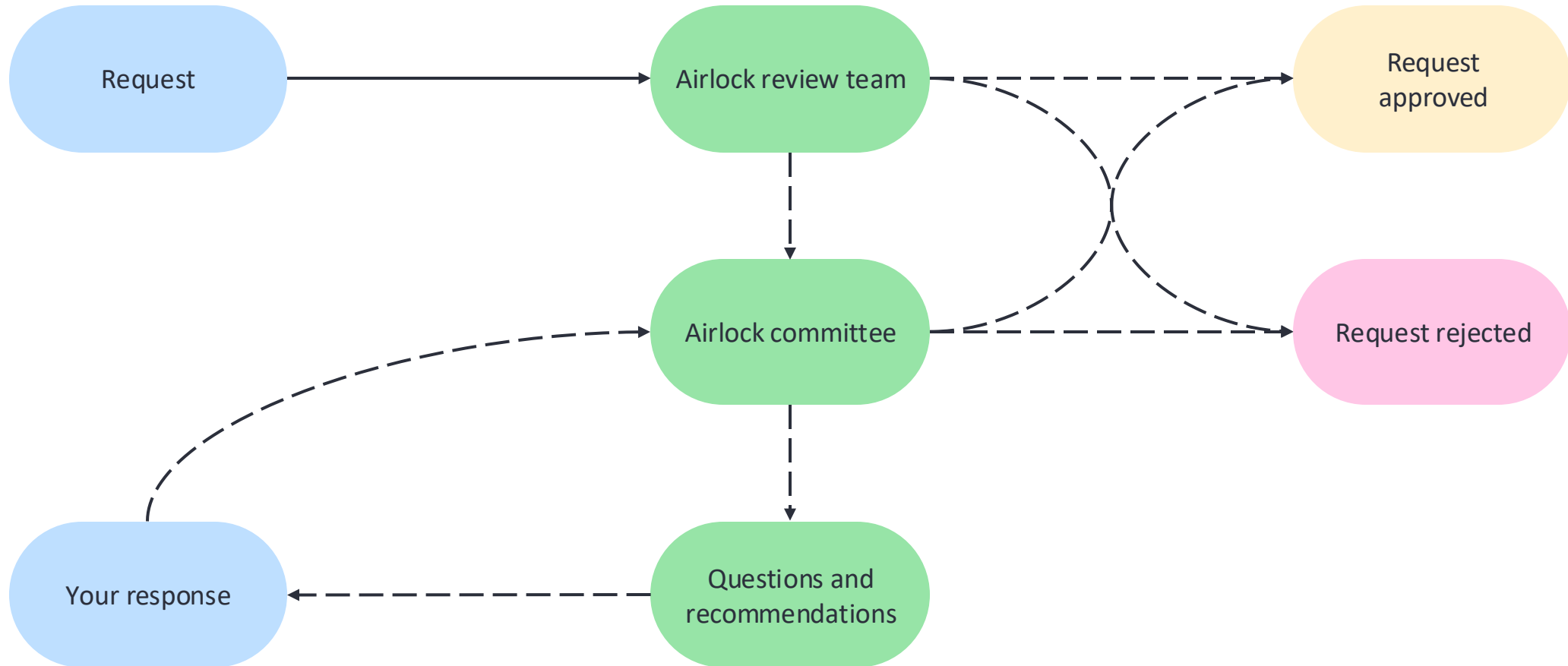
Publications from use of NGRL data (all time)



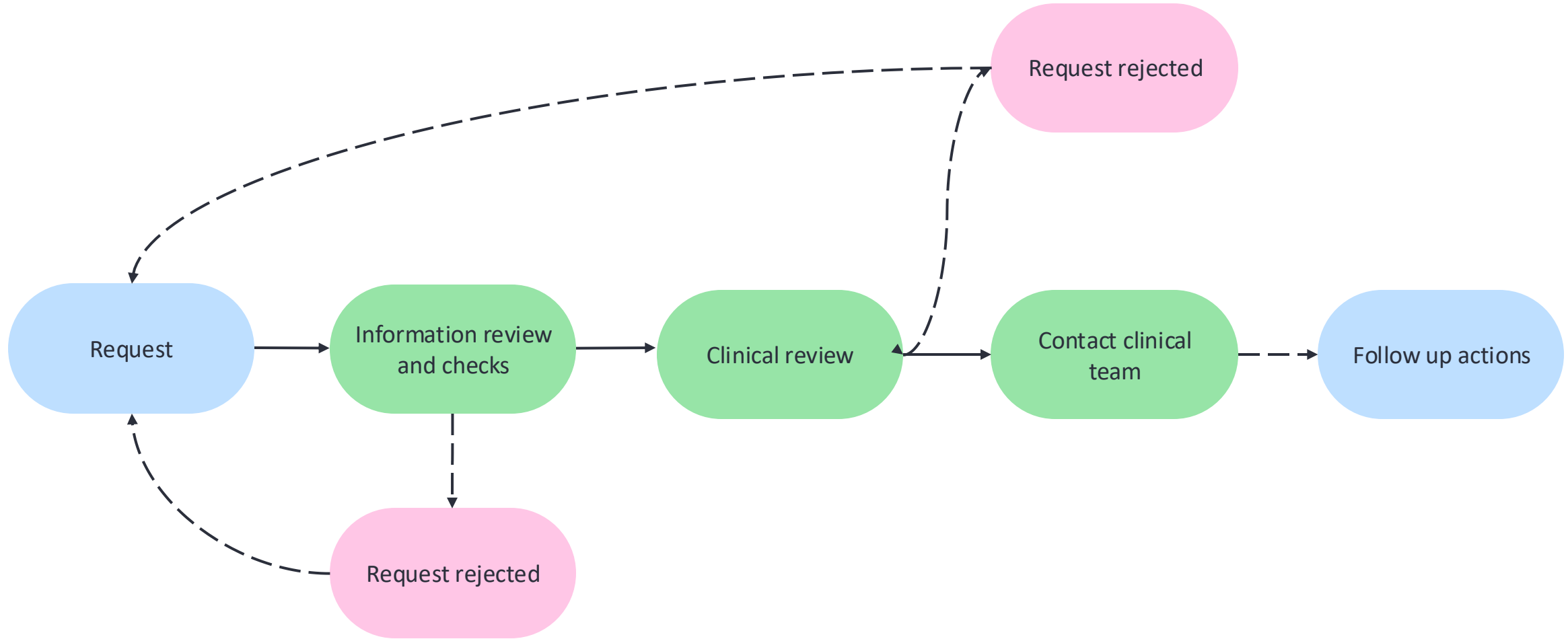
Publication request demo

6. Sharing data with reviewers and examiners in the RE

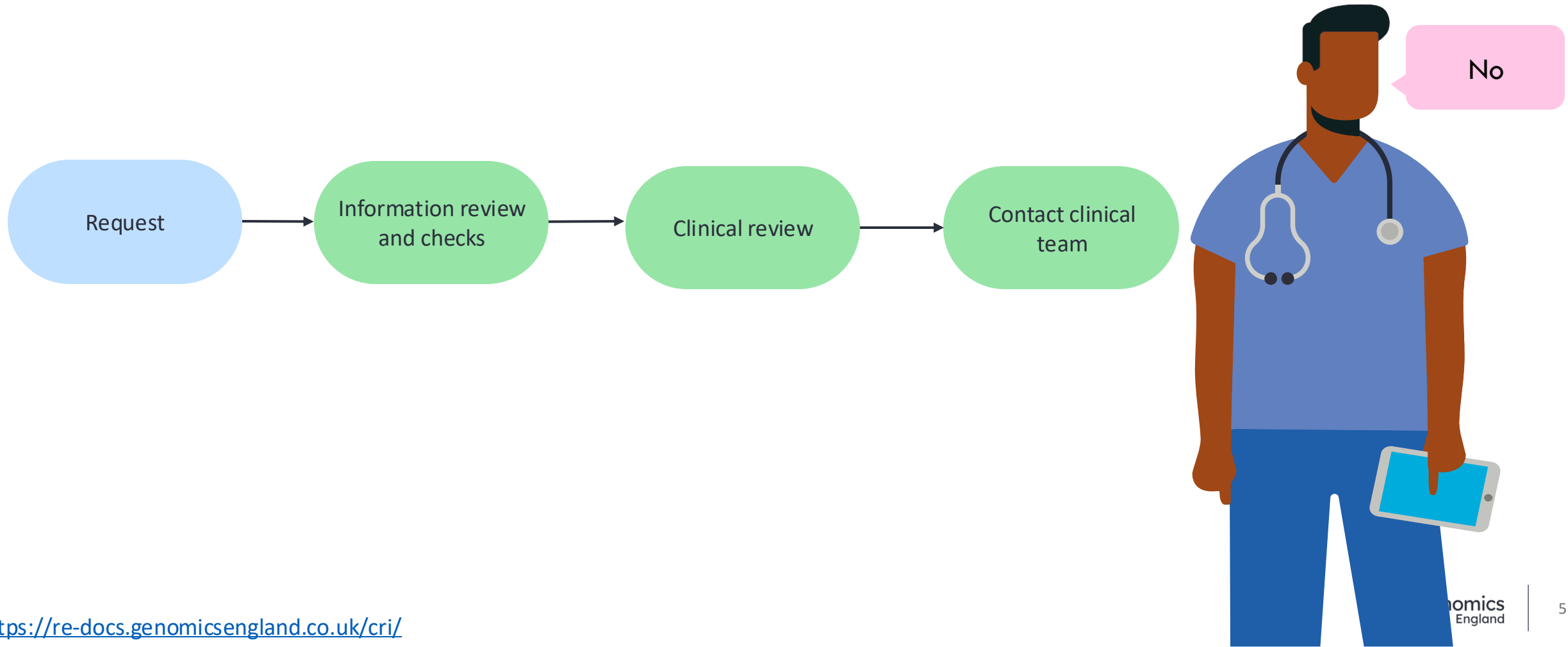
Airlock request rejected



Clinical collaboration request rejected



Clinician chooses not to collaborate



Show your examiners your results in the RE

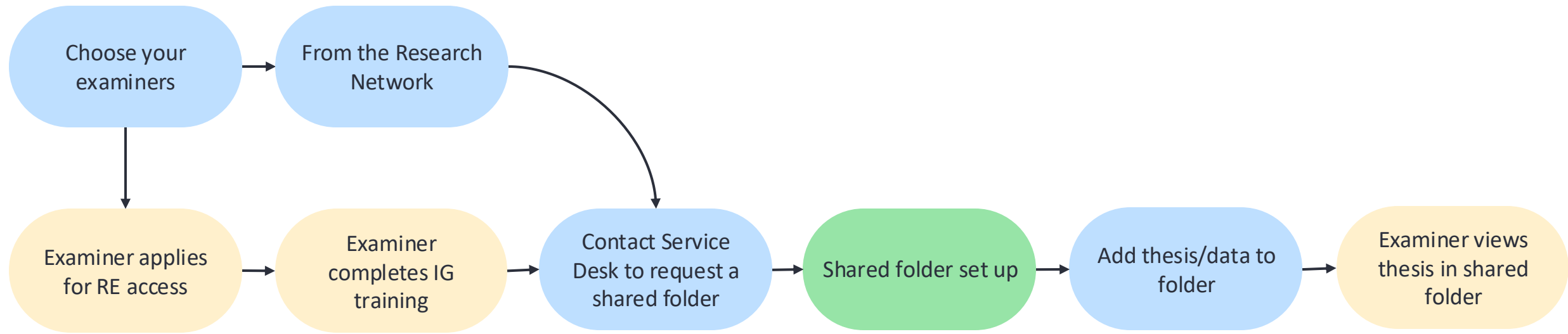


Redacted thesis sent out



Full thesis inside RE

Show your examiners your results in the RE



Some results in the RE



Redacted paper published



Data in the RE

Data Access Statement template



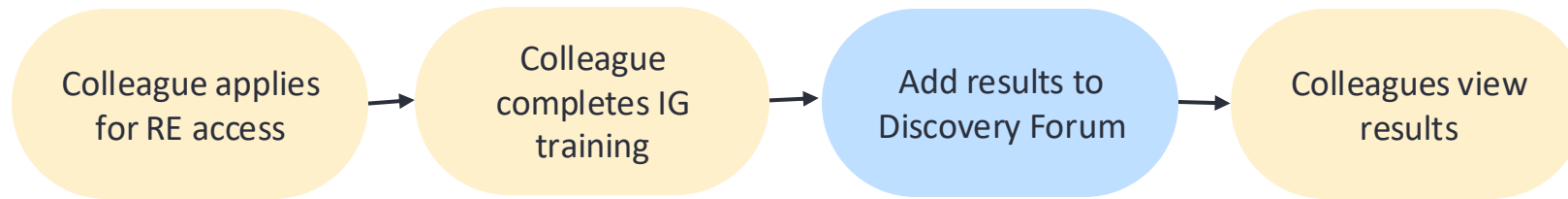
Data from the National Genomic Research Library (NGRL) supporting this study are available within the secure Genomics England Research Environment. Data used or referred to in this publication include:

XX

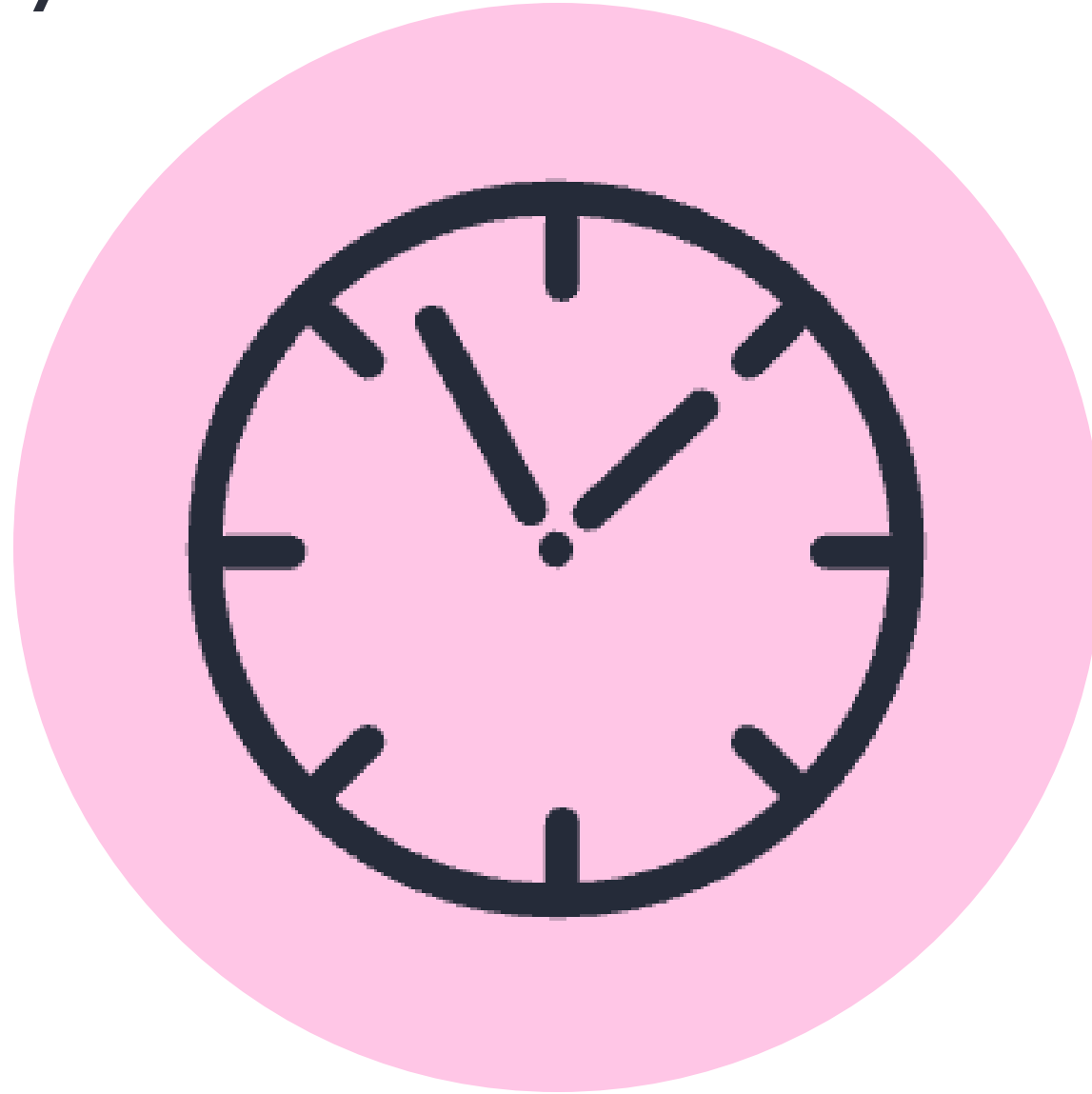
Access is restricted to approved researchers who are members of the Genomics England Research Network, subject to a data access agreement and participant-led governance. For more information on data access, visit:

<https://www.genomicsengland.co.uk/research>

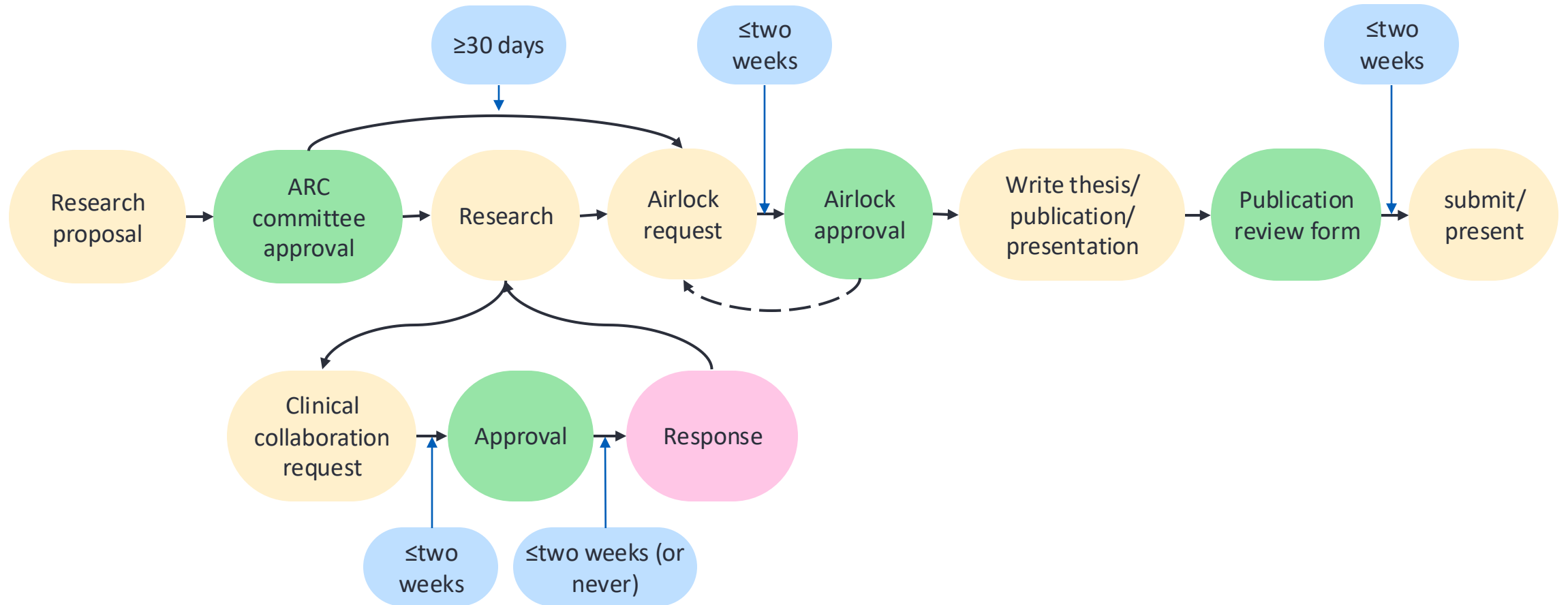
Share your results with your company in the RE



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

3rd Tuesday every month

Introduction to the RE

21/10

18/11

16/12



Materials
from past
training all
online

Training sessions

11/11

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

9/12

Running workflows on the HPC and Cloud

12/1

Using the Research Environment for clinical diagnostic discovery

9/2

Importing tools and data to use in the Research Environment



Materials
from past
training all
online

Feedback



Thank you

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