

# UoM UK Biobank User Community Meet-up

November 2018

George Leaver, Pen Richardson, Gillian Sinclair, [Research IT](#)

Dr Hui Guo, [Division of Population Health](#)

Chris Boulwood, Chief Information Officer, [UK Biobank](#)

# Why is today happening?

FAQ: Is there a UKBIOBANK group at the University where I can

Ask for advice? Share good practice?

Collaborate, avoiding duplication of effort in data clean up?

- Yes! Mailing list - [UOM-UKBIOBANK@listserv.manchester.ac.uk](mailto:UOM-UKBIOBANK@listserv.manchester.ac.uk)
  - To join: <https://listserv.manchester.ac.uk/cgi-bin/wa?SUBED1=uom-ukbiobank&A=1>
- This meeting is aimed at kickstarting a community here at UoM, and includes
  - Research IT facilities related to accessing and processing UKBIOBANK datasets
  - Dr. Hui Guo - a user experience of using these facilities
  - Chris Boulwood, Chief Information Officer from UKBIOBANK – “Past, Present, future”
  - *Your* community - Q&A and discussion, next steps.

# Future: What and How?

- What do you want from this community?
  - Exchange of hints 'n' tips, good practice?
  - Help with data clean up? Or a particular piece of software or workflow?
  - Internal collaborations?
  - Guidance on applying to UKBIOBANK? A closer relationship with UKBIOBANK???
  - Training?
- More meetings?
  - Frequency? Format?
- The mailing list – what to use it for?
  - General discussion? Asking for help?
- Other suggestions

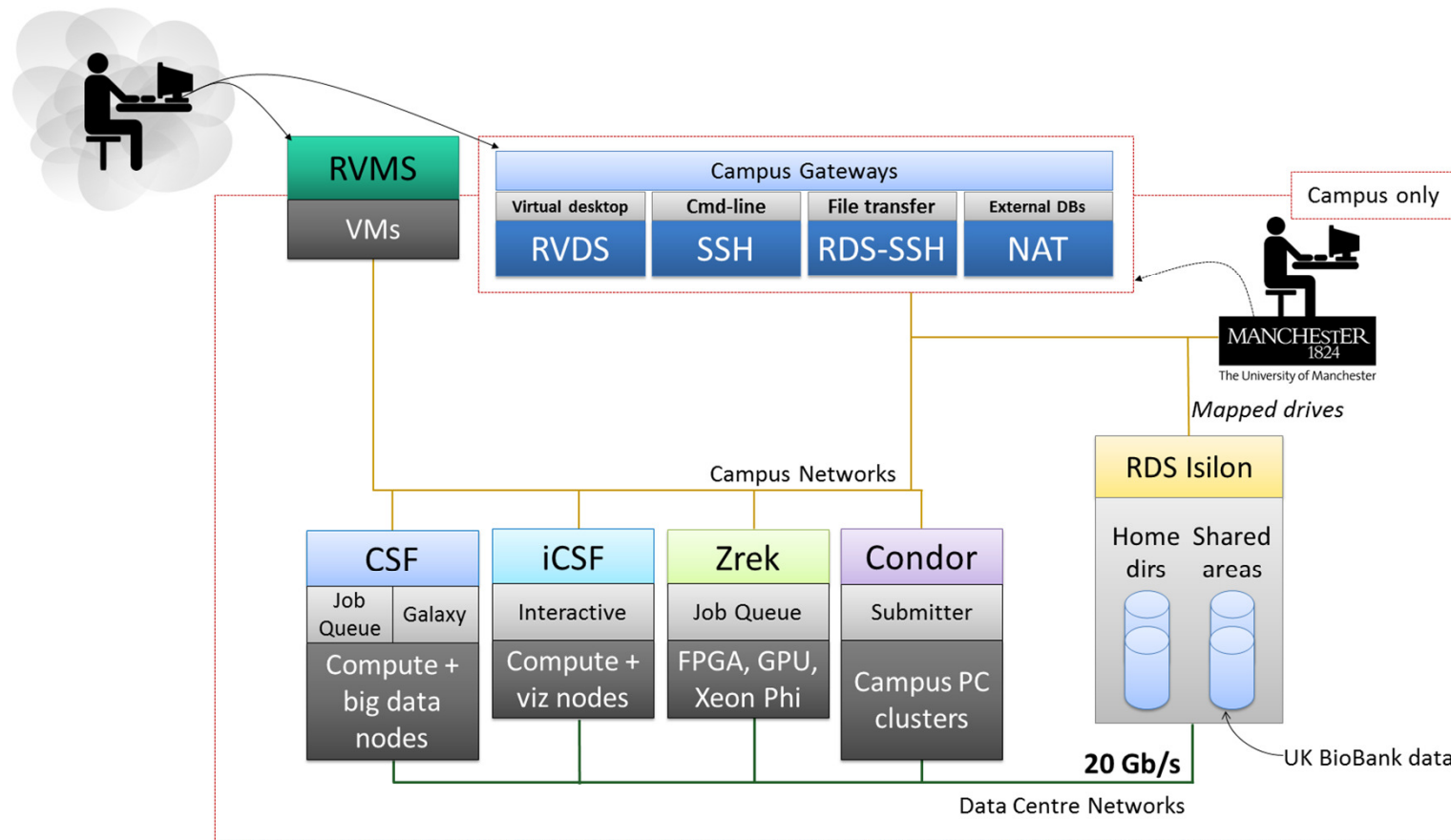
# Future: Who?

- We are looking for a volunteer researcher to drive the community
- Research IT can help with
  - Access to centrally hosted datasets
  - Managing the mailing list
  - Logistics for meetings – e.g. organising a room
  - Publicity about the group and the meetings
- Research IT are keen to keep a close relationship with the community
- Research IT staff are not
  - Users of the datasets or software used to process / analyse it
    - This limits our understanding of your needs and issues!
  - A replacement for the UKBIOBANK access and support teams
- Interested? Feel free to offer your services today, or contact us after the meeting to discuss it further.

# Research IT

- What? Research Infrastructure (& people), Res Software Engineers, RDM
  - Compute, storage, app support / software engineering, "data compliance"
  
- Why? A number of our users requested storage for a UK BioBank dataset
  - All waiting for 500,000 participant release (~24TB needed)
  - Download → verify → decrypt → verify → decompress → verify
  - UK BioBank advised that an institutional copy could be created
  
- How can Research IT help ...?
  - We can provide you with your own storage (up to 8TB free)
  - We can host datasets centrally where there is enough demand
  - We can provide lots of *compute* to help with processing

# Computationally Intensive Research *Ecosystem*



# Our Hosted Datasets

- Currently the *500,000 participant* datasets (v3)
  - Imputation data: EGAD00010001474
  - Genotyping data: EGAD00010001497
- How to obtain access to our copy
  - Principal Investigator (PI) informs UK BioBank of their intent to use an institute-held dataset *during the application review process (preferably)*
  - PI *must* name people who will access the data on the *Material Transfer Agreement* (MTA)
  - Research IT will provide access following confirmation a user is on the MTA

[access@ukbiobank.ac.uk](mailto:access@ukbiobank.ac.uk)

# Where can datasets be accessed?

- On the Computational Shared Facility (CSF) and friends
  - ~10,000 CPU cores to run data processing jobs
  - Large memory compute nodes (128GB, 256GB, 512GB, 1TB)
  - GPUs (restricted access)
  - Lots of bio-inf apps: bedtools, BioPerl, Bioconductor, BLAST, Bowtie, BWA, FastQC, ...
  - Free-at-point-of-use access or access via a contributing group
  
- On your desktop
  - Map a *network drive* (e.g., a new R: drive on Windows)



## Further info

- Centrally hosted dataset info  
<http://ri.itservices.manchester.ac.uk/hosted-data-sets/ukbiobank/>
- Use on Computational Shared Facility  
<http://ri.itservices.manchester.ac.uk/csf>  
<http://ri.itservices.manchester.ac.uk/csf-apps/software/applications/ukbiobank/>
- Subscribe to UoM-UKBIOBANK@LISTSERV.MANCHESTER.AC.UK email list  
<https://listserv.manchester.ac.uk/cgi-bin/wa?SUBED1=uom-ukbiobank&A=1>
- Subscribe to UKB-GENETICS@JISCMAIL.AC.UK email list  
<https://www.jiscmail.ac.uk/ukb-genetics>
- UK BioBank  
<https://www.ukbiobank.ac.uk>

# Presentations

- UK Biobank data access and software for genome-wide association analysis  
Dr Hui Guo, Division of Population Health, Health Services Research & Primary Care,  
School of Health Sciences, University of Manchester  
*Experience of accessing the UK Biobank genotype data on the central server maintained  
by Research IT, with an introduction to some useful software for genome-wide  
association analysis.*
- UK Biobank: Past, present, and future  
Chris Boulton, Chief Information Officer, UK Biobank  
*A brief review of the history of UK Biobank, a brief review of the history of UK Biobank,  
and a summary of other future developments currently under discussion.*

# Over to you...

- Specific questions received
- Possible discussion topics
  - Data clean-up
  - Datasets
  - Using central compute platforms
  - Feedback to UK BioBank
  - Training
  - Good practice / hints-n-tips
  - Workflows
- Community next steps
  - Researcher-led?
  - Future meetings?
  - Format of event?