

EXPERIENCE WITH THE
EUROPEAN GENOME-
PHENOME ARCHIVE



John Dawson & Rachel Thompson



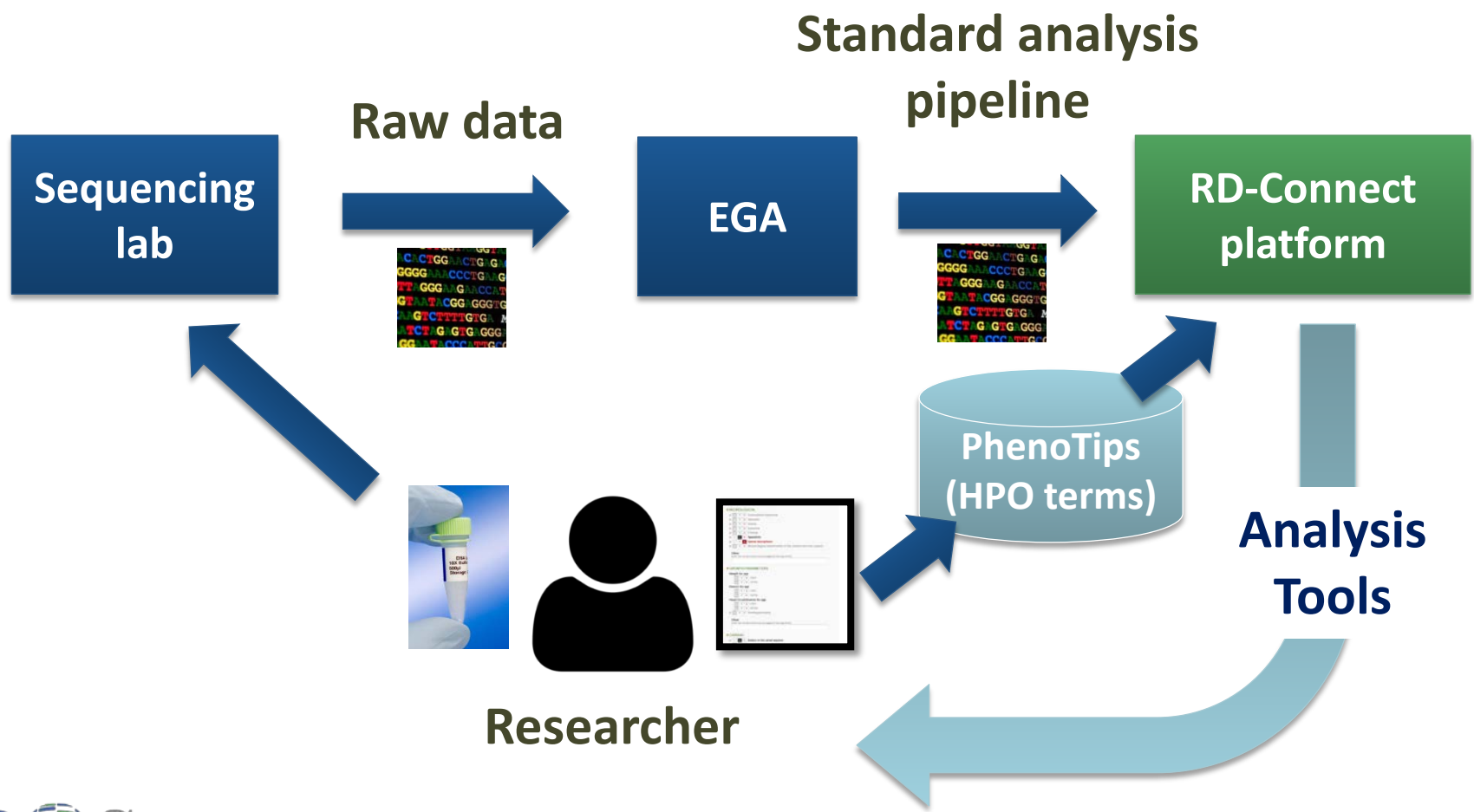
What is the EGA?

2

- The **European Genome-Phenome Archive** is an online repository for sequencing data co-hosted by EBI (UK) and CRG (ES)
- Used in RD-Connect as the permanent archival repository for the raw data generated in associated projects
- Also stores (some) phenotypic information
- Has robust security and access mechanisms and long-term sustainability – excellent for archival purposes
- Does not have analytical capacities required by our collaborating end users – still a need for the RD-Connect platform
- Further details – talk to Dylan or Paul (EBI) or Auldad (CRG)



(Genomic) data flow to RD-Connect



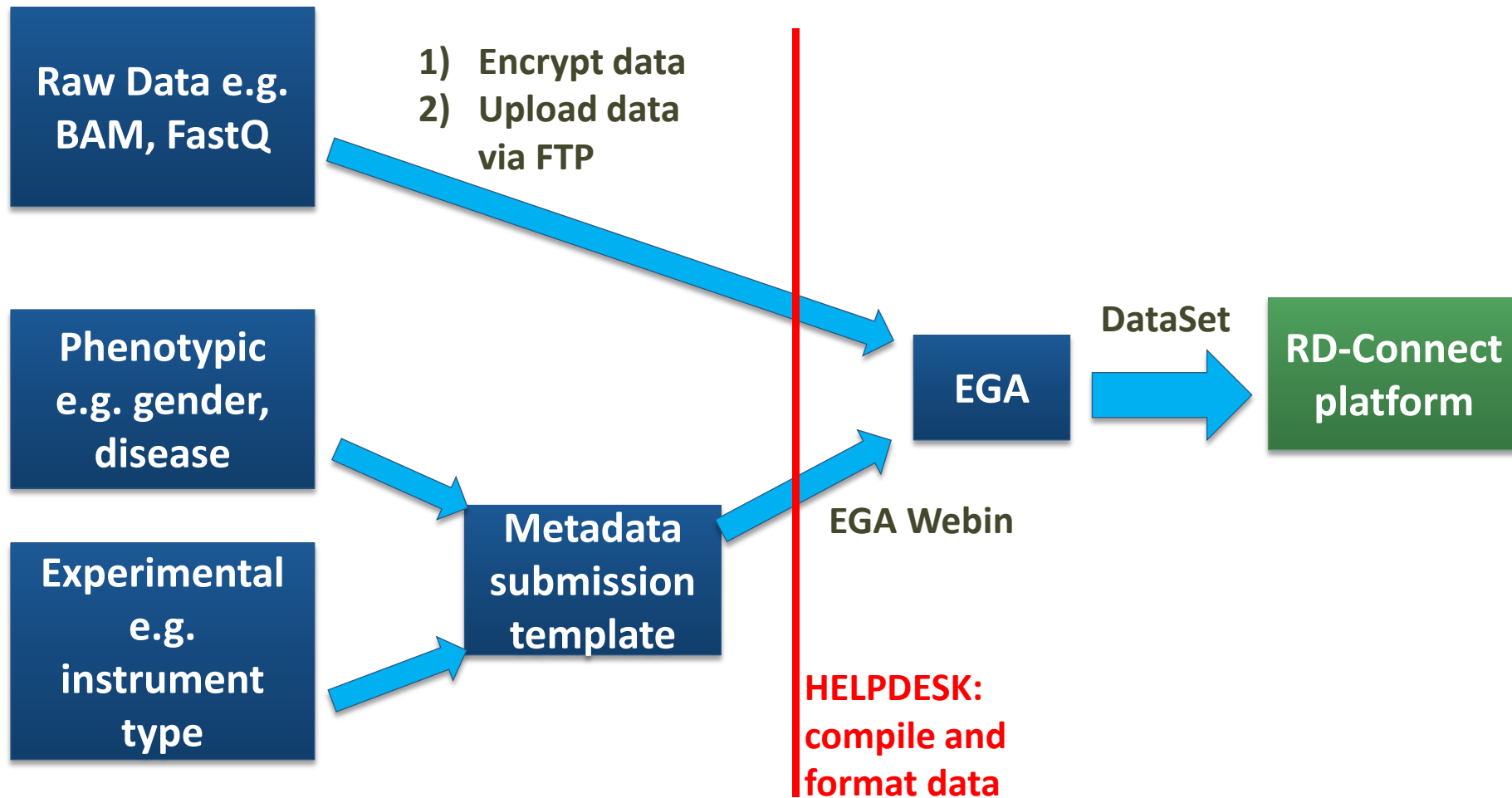


Data flow – helpdesk input

4

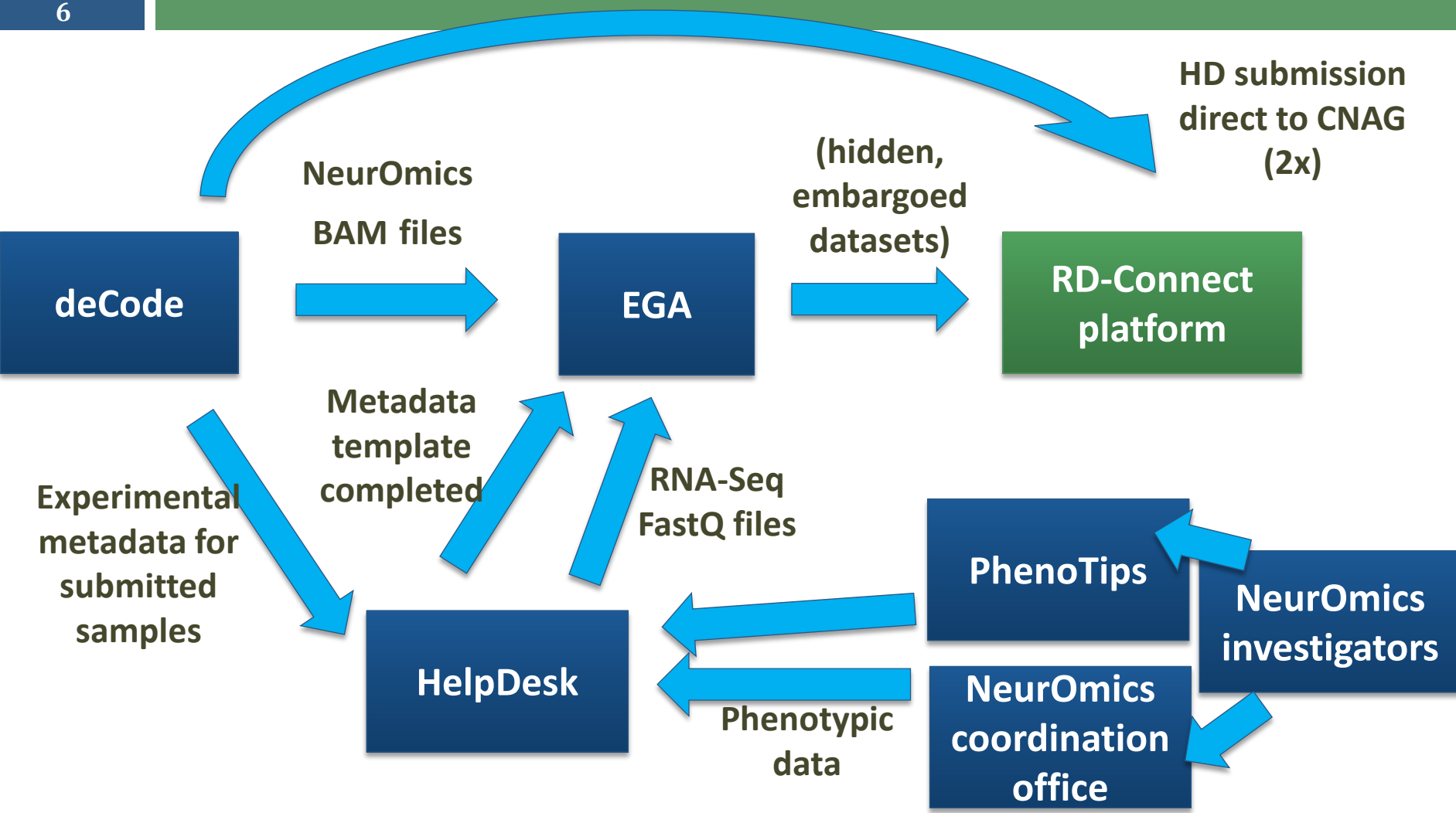
- EGA submission process quite complex – barrier for many RD partners unfamiliar with the system
- EGA has own helpdesk support – very good but don't have capacity to be fully hands-on with RD-Connect partners (have to support all EGA submitters) and cannot support e.g. hard drive submissions
- We want it to be as easy as possible so people more willing to share data
- RD-Connect helpdesk jobs:
 - Follow data flow, iron out kinks, streamline process, create SOP for data sharing
 - Provide extensive hands-on support (“hand-holding”) or even complete submissions on behalf of partner projects
- More data shared, more useful the platform becomes

Submitting the data (theory)



Submitting the data (actual)

6





Complications

7

- Time consuming; back and forth
- Having to use CLI for encryption
- Tracking data throughout process
- Once submitted to the EGA extra metadata can't be uploaded in the same way
- Many IDs throughout the process
- Linked samples (exome plus genome plus RNASeq on same patient): creates 'secondary' samples that are difficult to find in PhenoTips or EGA
- Needed to transfer data to CNAG before embargo periods had expired – still "hidden" in EGA – special mechanism needed
- File size! Need terabytes of space; dedicated server?



Complete datasets

8

- Compile uploaded data into Datasets for CNAG to download
- Provide all accession numbers for each sample
- 567 WES NeuroOmics at CNAG now
- 65 WGS NeuroOmics ready for download to platform
- 117 WES NeuroOmics uploaded/pending dataset completion
- 397 WES coming from mitochondrial/neurogenetic team at UNEW
- ~1000 WES coming from MyoSeq project at UNEW
- EURenOmics, other interested parties



Interested parties

9

Name	Address	Email
Prof. Wolfgang Mueller-Felber	Center for neuromuscular diseases in childhood, Dr.V. Hauner Children's Hospital, Ludwig Maximilian University, Munich	wolfgang.mueller-felber@med.uni-muenchen.de
Dr. Katharina Vill		katharina.vill@med.uni-muenchen.de
Dr. Astrid Blaschek		astrid.blaschek@med.uni-muenchen.de
Mauro Javier Oruezabal Moreno	Jefe de Servicio de Oncología Médica, Hospital Universitario Rey Juan Carlos, c/Gladiolo s/n- 28933-Móstoles, Madrid	mauro.oruezabal@hospitalreyjuancarlos.es
Geoff Woodward	Bristol Genetics Laboratory, Pathology Sciences, Southmead Hospital, Bristol	geoff.woodward@nbt.nhs.uk
Jan Geryk	Dept. of Biology & Medical Genetics, Motol University Hospital, Czech Republic	jan.geryk@fnmotol.cz
John Vissing	Department of Clinical Medicine, Rigshospitalet - Neurocentret, Blegdamsvej 9, 2100 København Ø	john.vissing@regionh.dk
Prof. Sophie Hambleton	Primary Immunodeficiency Group, ICM, 4th Floor, Catherine Cookson Building, Medical School, Newcastle University	sophie.hambleton@ncl.ac.uk
Prof. Constantinos Deltas	Laboratory of Molecular and Medical Genetics, Department of Biological Sciences, University of Cyprus, 1, Panepistimiou Avenue 2109 Nicosia, Cyprus	deltas@ucy.ac.cy
John Sayer	Institute of Genetic Medicine, Newcastle University, Central Parkway, Newcastle NE1 3BZ	john.sayer@newcastle.ac.uk
Prof. Rita Horvath	Institute of Human Genetics, Newcastle University, International Centre for Life, Central Parkway, Newcastle upon Tyne	rita.horvath@ncl.ac.uk



Next steps

10

- Questions on additional metadata important for NeurOmics but not requested by EGA
- **XML via REST** for remaining metadata
- Other types of omics data
- SOP for data flow