

## **Technical Information on the DPUK Data Portal Analysis Environment**

DPUK allows approved researchers access to cohort data using a virtual desktop infrastructure. Users will download VMware Horizon Client to their local PC in order to connect to the DPUK infrastructure, using two-factor authentication. This involves a mobile authenticator passcode as a first step, and username + password combination as a second step.

### **macOS**

macOS Users will be likely to need to change their settings at point of logon. After completing the two-factor authentication, you will be presented with your choice of desktop to access. Before selecting the desktop, you should explore which is most optimal for you. This is done by selecting the cog symbol in the top right of the window. Where 'Connect via' appears, please select one of three options (Microsoft RDP, PCoIP, VMware Blast) to confirm which allows connection.

To provide a solution to the 'Ctrl+Alt+Delete' issue with macOS devices, the virtual desktop has a 'Send Ctrl+Alt+Delete' function in the floating menu bar at the top of the desktop. Please note as a general point: disconnecting (simply closing the virtual desktop) does not shut down the desktop, but selecting disconnect **and** log off will do so however).

### **Standard Technical Desktop Specification**

DPUK deploys a standard desktop for researchers seeking to access standard phenotypic data on the Portal (clinical observational data for statistical analysis rather than omics for example).

The standard desktop has the following specification: Windows 7, 8GB RAM, and 4 CPUs. Pre-loaded with R, RStudio, SPSS, SAS, Stata (16), Python, Eclipse, SQL Server Management Studio, Microsoft Office.

Statistical software, such as R and Python, can connect to its official library/package/index directory to enable configuration of software on a per-user basis.

As standard, DPUK will endeavour to provide the most up-to-date version of software for use within the analysis environment. Should an older version be necessary for analysis, please contact the Data Portal team via <https://helpdesk.hiru.swan.ac.uk> or [helpdesk@chi.swan.ac.uk](mailto:helpdesk@chi.swan.ac.uk). In some cases, it may be necessary to provide a dedicated user desktop with the required software, whereas in other cases DPUK would seek to provide the correct version within the shared desktop suite.

DPUK can provide a larger scale desktop, which contains the same pre-loaded suite of software, however has 32GB, RAM, 8 CPUs. There is a final extra-large desktop, which contains 128GB RAM and 16CPUs, which is likely to be more suitable for studies in the machine learning and statistics relying on heavier computation loads. There are a limited number of these larger desktops available on the DPUK Data Portal, and therefore to increase availability for the study, the Data Portal team may discuss costs for the increased desktop specification licenses with the study team.

Windows 10 virtual desktops are in test, and will be deployed later in 2020 for routine access.

General storage is scalable according to the study requirements, with basic access to data stored on our systems free to all users. Studies needing large amounts of storage outside of the standard shared network shares to bring in their own larger scale data will be subject to potential extra costs which will be discussed with researchers as part of the set-up of the desktop.

Research teams accessing cohort data have a shared folder (study number and name on the S drive) to be able to share collaborative work and individual researchers have their own personal directory (username on the P drive) to save their private files.

Non-Windows based desktop systems such as Linux are also available. However, as this requires additional setup and bespoke build, we would like to discuss with applicants their requirements in order to provide as usable system as possible for planned analyses.

#### **Technical Specification Form**

On the DPUK application form, there is a technical specification section. We would like all applicants to highlight the type of study planned, select which desktop specification they feel will be required for their study, and if there are any bespoke requirements needed (such as the installation of software that we do not provide as standard listed above). Having this information at an early stage will allow the Data Portal technical team to discuss the build of extra requirements and tailor the system to the user(s).

Access to High Performance Computing and GPU usage for various analyses within various domains is possible, and we would request that this is alluded to in the hardware/software requirements of the technical specification form. The Data Portal team can then look to prepare the necessary operating system and software configurations to link to the HPC cluster. Depending on the use case and time spent using the available HPC cluster or GPU, there may be extra costs charged to the study, however the Data Portal will facilitate initial discussion with the requesting applicant in order to look at funding options.