



PhenomUK Data Strand Report

Report identifying and mapping hardware and storage needs

Authors:

**Valerio Giuffrida (valerio.giuffrida@nottingham.ac.uk)
Sotirios A Tsaftaris (s.tsaftaris@ed.ac.uk)**

4th January 2024

Executive Summary

We selected two potential providers for hosting the PhenomUK Digital Research Infrastructure. We have been discussing with SCAN and EPCC for several months to present them with the community needs for this scoping project. The EPCC proposal resulted in being considerably more cost-effective. This reduction in costs originates in that we rent resources as part of the EPCC EIDF offering, which also includes cost-effective data storage.

Foreword

This report is a follow-up of the *Report mapping the community requirements on Software and Data* that our team has already produced. The reader is advised to refer to that document before proceeding.

I. Introduction

Following our recommendations on the software framework to be used to ingest, organise, and manage the data generated by the plant phenotyping community in the UK, we need to identify and obtain a hardware solution to accommodate the software stack we aim to develop and deploy as part of the scoping DRI.

The diversity of hardware currently available on the market is considerable. To narrow down the choices, we leveraged the information we gathered during:

1. The interviews we conducted with researchers, stakeholders, and developers having specific questions on hardware/storage needs.
2. The responses we collected with the Mentimeter poll we presented to the community during the first PhenomUK-RI occurred in September 2023 probing options on hardware/storage/compute current and future needs.

With this information at hand, we outlined below minimal desiderata on the hardware side of the DRI that is necessary to complete this scoping project:

- At least 500TB of storage for user data.
- Computational power to run (inference and training) AI models (GPUs).
- Computational power to run the web interface and database for the software framework.

We then proceeded to identify suitable providers in the UK, the details of which are provided in the next section.

II. Identified Suppliers

We identified two suitable providers in the UK:

- Scan Computers International Ltd (shorted as SCAN).
- Edinburgh Parallel Computing Centre (shorted as EPCC).

► **SCAN:** located in Bolton (UK), SCAN is renowned for providing high-spec computer hardware to consumers and businesses, including several universities all over the UK.

► **EPCC:** located at The Bayes Centre, EPCC is a supercomputing centre based at The University of Edinburgh. They have considerable experience with HPC facilities, operating also UKRI/EPSC supercomputing facilities. While their aim is at providing supercomputing facilities, they can still host the DRI for this scoping project.

Consultations with both providers occurred in several forms over two months: (i) email exchanges; (ii) MS Teams meetings; (iii) face-to-face. This report will show the outcome of this consultation process with both providers.

III. Proposed options and quotations

After showing and discussing our desiderata with both providers, we will summarise their proposed options together with the budget required for procurement.

► **SCAN Proposal:** SCAN detailed their proposal in a 6-page document. The major takeaways from that report are summarised in the following table:

Item	Description	Qty	Unit Price	Sub Total
GPU Server	NVIDIA DGX-A100	1	£171k	£171k
Networking	NVIDIA SN3700VS2FC Ethernet Switch	2	£23.5k	£47k
Network Cabling	NVIDIA Networking 200 GbE cabling (10 x 3M AOC & 4 x 0.5M DAC)	1	£15.5k	£15.5k
Orchestration	Run:AI DGX Bundle	3	£14k/year	£42k
Storage	Peak:AIO Storage Solution (RAID 6) (19 x 15TB NVMe Flash Drives)	2	£82.5k	£165k
Storage (backup)	Peak:AIO Storage (RAID 6) (10 x 61TB NVMe Flash Drives)	1	£52k	£52k
Installation	SCAN setup cost	1	£7.5k	£7.5k
Total				£500k

Table 1: SCAN proposed hardware solution costings (reported numbers are rounded for the sake of clarity).

SCAN advised that procuring the items listed in Table 1 can be done within 6-8 weeks from the day the order is placed. This solution is valid until the end of January 2024 (more details on this deadline are provided in Section IV).

An important consideration that must be taken into account with this proposed solution is the cost incurred to host all of this hardware solution. SCAN has facilities in the UK capable of hosting such hardware with a cost of approximately **£90k/year**. An alternative and cost-effective solution would be to host this option within the *University of Edinburgh IT Infrastructure*.

► **EPCC Proposal:** Before introducing the details of the EPCC proposal, it is important to highlight that EPCC hosts several national-wide supercomputers, such as ARCHER2.¹ They have considerable experience running HPC infrastructures. For the PhenomUK DRI, we leverage an offering by EPCC that falls under the EIDF arrangement which has been privately developed by UoE and EPCC and is not using any UKRI funding. The following table summarises the EPCC offer:

Item	Description	Qty	Unit Price	Sub Total
GPU Server	NVIDIA DGX-H100	1	£335k	£335k
EIDF Infrastructure	Access to Virtual desktops, Kubernetes, and 500TB of data storage ²	3	£1.5k/year ³	£4.5k/year

¹ EPCC will be the host of the first exascale supercomputer of the UK.

² This cost is calculated for each developer accessing the infrastructure. We accounted for 3 users to accommodate our team (2 Software Engineers and 1 Data Scientist).

³ This cost is subject to annual increase due to inflation, calculated at 5% / annum.

Support	Staff cost for the first two months (0.25FTE for 2 months, 0.05FTE till the end of the project)	1	£13.4k	£13.4k
			Total	£353k

Table 2: EPCC proposed hardware solution and access to their infrastructure costings (reported numbers are rounded).

We also asked for an estimate to extend the use of this solution for 2 years after the end of the scoping project. The additional costs to maintain this solution are summarised as follows (we utilised a different table colour for the sake of clarity):

Item	Description	Price
EIDF Infrastructure	Access to Virtual desktops, Kubernetes, and 500TB of data storage ⁴	£10k
Support	Staff cost for the first two months (0.25FTE for 2 months, 0.05FTE till the end of the project)	£15k
		Total
		£25k

Table 3: EPCC additional costings to keep maintaining the hardware solution and access to their infrastructure for 2 years after the end of the PhenomUK-RI scoping project (reported numbers are rounded).

IV. Consideration

As mentioned in the previous section, SCAN estimation does not include hosting. To reduce the required 90k/year, we asked the University of Edinburgh IT Service about the possibility of hosting a GPU server. After several discussions, we found out that their infrastructure cannot accommodate a DGX-H100 due to limitations to their power supply. For this reason, we asked SCAN to quote for DGX-A100, which is the NVIDIA solution preceding the newer H100. However, DGX-A100 will be out of production starting from February 2024, which justifies the hard deadline SCAN provided us to place the order.

The EPCC GPU server is more expensive, but it includes state-of-the-art hardware, which makes this option more up to date. In addition, EPCC cost is reduced with respect to the SCAN option, because we will be renting part of their existing EIDF infrastructure (including storage). Hence, we do not need to buy extra hardware for data storage and interlinking devices (e.g., switch) and related cabling to get their option up and running.

V. Recommendation

After several internal discussions with our team, we assessed that the EPCC solution is viable. Given the difference in cost, we recommended to the Management Board the **EPCC option**.

Acknowledgements

We thank all the collaborators for their tireless and invaluable contribution to the Data Strand: Feng Chen, Andrew Wood, and Danilo Bueno. We also thank Jamie Higham and the SCAN team for their support throughout these months. Finally, we thank Prof. Mark Parsons and Kostas Kavoussanakis and their EPCC team for all the support, suggestions, and guidance they provided to help us detail our needs following what they have already in place.

⁴ This cost is calculated for each developer accessing the infrastructure. We accounted for 3 users to accommodate our team (2 Software Engineers and 1 Data Scientist).