



**National Federated
Compute Services**
NetworkPlus

Spring Conference
26 - 27 February 2026

Identifying Barriers for Biology Researchers Using Federated HPC Service

Wahab Kawafi, BriCS (Bristol Centre for Supercomputing)





Project team

Project Leads

Wahab Kawafi (UoB)

Danielle Paul (UoB)

Laura Shemlit (RFI)

David Horsfall (NCU)

Andrew Dowsey (UoB)

Project Team Members

Jon Lees (UoB)

Ravi Teja (RFI)



**What issues do biologists face in
using HPC?
Software, interface, or data?**





**National Federated
Compute Services**
NetworkPlus

Motivation

Why are we focusing on biologists?



Motivation

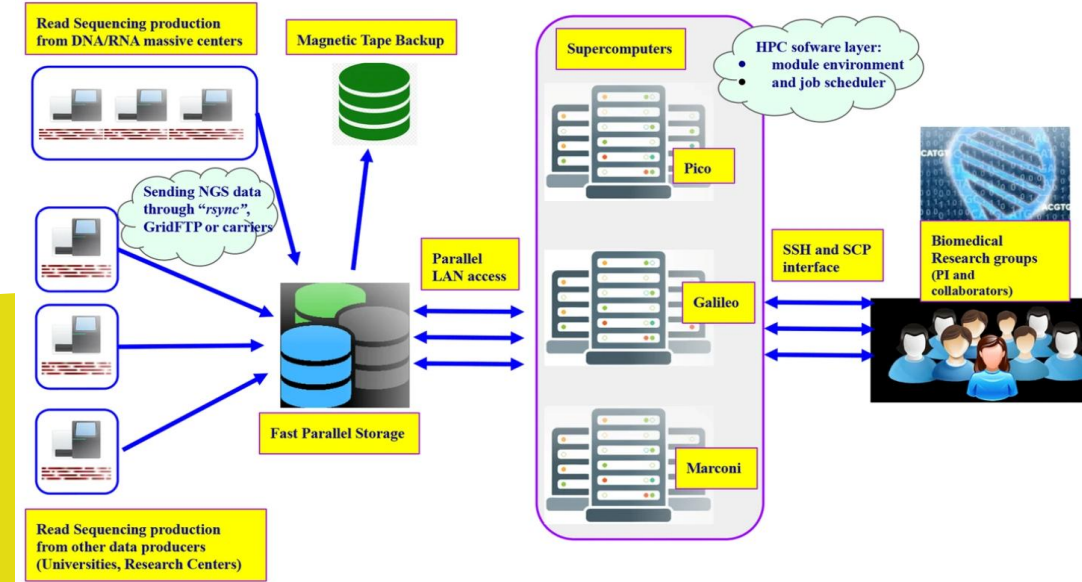
- For scientific computing, computational physics and chemistry have mature and well supported software stacks.
 - Biologists face relatively lower HPC literacy partly due to the newer influx of AI models.
 - Focusing on NON-sensitive biology research. (i.e. not TREs/SDEs)
 - Structural biology, genetics, OMICs, engineering biology, animal research.
- We aim to apply the knowledge gathered to other scientific communities / domains.
 - Biologists benefit from dedicated HPC services such as EBI-EMBL, Cornell BioHPC, IMBBC Zorbas HPC, CINECA ELIXIR-IT HPC, and Bristol's BlueCryo.
 - How can federated HPC services meet the demand of domain specific HPC centres?

Software, interface, data

We look at Elixir-IT HPC@CINECA and Cornell BioHPC as examples:

1. Software: preinstalled software (1312 @ Cornell).
2. Interface: Graphical user interfaces (RDP, VNC, X-windows, Jupyter, R-Studio)
3. Data: Preloaded with reference datasets (e.g. PDB, UniProt) and facilities to ingress data from scientific equipment.

From: ELIXIR-IT HPC@CINECA: high performance computing resources for the bioinformatics community



BioHPC Cloud Software

There are 1312 software titles installed in BioHPC Cloud. The software is available on all machines (unless stated otherwise in notes), complete list of programs is below, please click on a title to see details and instructions. Tabular list of software is available [here](#)

Please read details and instructions before running any program, it may contain important information on how to properly use the software in BioHPC Cloud.

3D Slicer, 3d-dna, 454 gsAssembler or gsMapper, 7zip, a5, ABRicate, ABrujin, ABySS, AdapterRemoval, adephylo, Admixtools, Admixture, AF_unmasked, AFProfile, AFsample2, AGAT, agrep, albacore, Alder, AltIV-Perl interface, AlleleSeq, ALLMAPS, ALLPATHS-LG, AlphaFold, Alphafold3, alphapickle, AlphapullDown, AlphaScore, Amber, AMOS, AMPHORA, amplicon.py, AMRFinder, AMRplusplus, analysis, ANGSD, AnnotaPipeline, Annovar, ant, antiSMASH, anvio, apollo, arcs, ARGWeaver, aria2, ariba, Arlequin, ART, ASEQ, aspera, assembly-stats, aster, ASTRAL, atac-seq-pipeline, ataqv, athena_meta, ATLAS, Atlas-Link, ATLAS_GapFill, atom, ATSAS, Augustus, autocyclus, AWS command line interface, AWS v2 Command Line Interface, axe, axel, BA3, BactSNP, bakta, BAMdash, bamm, bamsnap, bamsurgeon, bamtools, bamUtil, barcode_splitter, BarNone, Basset, BayeScan, Bayescenv, bayesR, baypass, bazel, BMap/BBTools, BCFtools, BCL convert, bcl2fastq, BCP, bdbag, Beagle, beagle-lib, BEAST, BEAST X, Beast2, bed2diffs, bedops, BEDtools, bettercalls, bfc, bgc, bgen, bicycle, BiG-SCAPE, bigQF, bigtools, bigWig, bioawk, biobakery, biobambam, Bioconductor, bioinfo-notebook, biom-format, BioPerl, BioPython, Birdsuite, biscuit, Bismark, Blackbird, blasr, BLAST, BLAST_to_BED, blast2go, BLAT, BlobToolKit, BLUPF90, BMGE, bmtagger, bonito, Boost, Bowtie, Bowtie2, BPGA, Bracken, BRAKER, BRAT-NextGen, BRBseqTools, BreedingSchemeLanguage, bresseq, brisqtyp, brocc, BSBolt, bsmmap, BSseeker2, btyper3, BUSCO, BUSCO Phylogenomics, BWA, bwa-mem2, bwa-meth, bwtool, cactus, CAFE, CAF5, caffe, cagee, canu, Canvas, CAP3, caper, CarveMe, catch, cBar, CBSU RNAseq, CCMetagen, CCTpack, cd-hit, cdbfasta, cdo, CEGMA, CellRanger, cellranger-arc, cellranger-atac, cellranger-dna, cellsnp-lite, centrifuge, centrifuger, centroFlye, CFM-ID, CFSAN SNP pipeline, CGAT, CheckM, CheckM2, chimerax, chimerax, chimerax, chip-seq-pipeline, chromosome, ChromHMM, chromosomer, Circlator, Circos, Circuitscape, CITE-seq-Count, clam, ClermonTyping, CLIMATHET, clues, CLUMPP, clust, Clustal Omega, CLUSTALW, Cluster, cmake, CMSseq, CNVnator, coib, coinfinder, colabfold, COLMAP, CombFold, Comparative-Annotation-Toolkit, compat, CONCOCT, Conda, Conform-gt, conpair, Cooler, coolpuppy, cooltools, copyNumberDiff, cortex_var, CoverM, crabs, CRISPRCasFinder, CRISPResso, crisprn, Cromwell, CrossMap, CRT, CSP2, cuda, Cufflinks, curatedMetagenomicDataTerminal, cutadapt, cuteFC, cuteSV, Cytoscape, czid, dadi, dadi-1.6.3_modif, dadi-cli, danpos, DAS_Tool, dashing, DBSCAN-SWA, dDocent, DeconSeq, Deepbiner, deepplasmid, DeepTE, deepTools, Deepvariant, defusion, degenotate, delly, DESMAN, destruct, DETONATE, dfast, diamond, dipcall, diploSHIC, discoal, Discover, Discover de novo, distiller, distrust, DiTASiC, DIYABC, dmtcp, dnmtools, Docker, dorado, DRAM, dREG, dREG.HD, drep, Drop-seq, dropEst, dropSeqPipe, dsk, dsat, Dsuite, dTOX, duphold, DWGSIM, dynare, ea-utils, EagleC, earlgrey, ecCodes, ecopcr, ecoPrimers, ectyper, EDGE, edirect, EDTA, eems, EgaCryptor, EGAD, egapx, eggNog-mapper, EIGENSOFT, elai, EIMaven, emacs, EMBLmyGFF3, EMBOSS, EMIRGE, Empress, emu, enfuse, EnTAP, entropy, epa-ng, ephemer, epic2, ermineJ, ete3, EukDetect, EukRep, EVE, EVM, exabayes, exonerate, ExpansionHunterDenovo-v0.8.0, eXpress, FALCON, FALCON_unzip, Fast-GBS, fasta, FastAAI, FastANI, fastcluster, fastGEAR, FASTK, FastME, FastML, fastp, FastQ Screen, fastq-mulx-1.4.3, fastq_demux, fastq_pair, fastq_species_detector, FastQC, fastqspliter, fastsimcoal2, fastspar, fastStructure, FastTree, FASTX, fcs, FEELnc, feems, feh, Ffmpreg, fgbio, ficle, figaro, Fijii, Filltona, fineRADstructure, fineSTRUCTURE, Fit-SNE, FlaGs2, flash, flash2, flexbar, Flexible Adaptor Remover, flexidot, Flve,



**National Federated
Compute Services**
NetworkPlus

First Case Study & Interviews

Gathering requirements and identifying barriers



Case Study: Relion and Alphafold on Isambard-AI

1. A project utilizing Cryo-EM and Alphafold derived protein structures
 2. Difficulty in installing Relion (No RDP)
 3. Web based alternative (Doppio) not available on aarch64
 4. Developer (STFC) doesn't have access to aarch64 development environment (or CI-CD)
- Installing Alphafold and obtaining UniProt not difficult
 - The main barriers are:
 - No access to X11 forwarding (RDP or xShell)
 - No developer access to aarch64

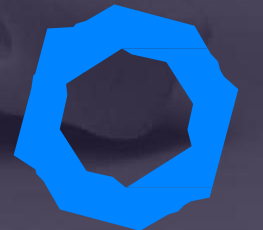




User Interviews

- Currently Interfaces are limited to SSH and Jupyter
 - Users have an expectation software installations to be a struggle on HPC systems.
 - But interface seems to be the main barrier.
 - Preloaded reference databases are not required, but dedicated Data Movement tooling is.
- A unique demand of developers on HPC services has become apparent.
 - Attracting Developers to HPC systems can improve support of scientific software.
 - There are currently no allocation mechanisms for developers to use HPC systems.

What interfaces and allocation mechanisms do Users and Developers need?





**National Federated
Compute Services**
NetworkPlus

Access to RSEs

**Do allocated projects have access to an
RSE resource?**



RSEs

- We identify 3 access mechanisms to RSEs
 1. HPC-RSE
 2. Central RSE team
 3. Embedded RSE
 - Many of the difficulties that researchers face are the focus of RSEs
 - E.g. Software portability
- Should access to RSE be a requirement for being granted an allocation on Federated HPC?

Thank you!

Next steps: continue with user interviews, demo the Interactive Drug Design tutorial.

Wahab Kawafi

a.kawafi@bristol.ac.uk