

HPHI Oversight Committee Meeting

Michaelmas Term Meeting – 21st November 2017, 4pm -5.30pm

1. Present

Professor Ed Bullmore (EB), Dr Guy Williams (GW), Professor Bertie Gottgens (BG), Dr Lydia Drumright (LD), Richard Bartlett (RB), Dr Paul Calleja (PC), Andrew Leedham (AL)

2. Apologies

Professor Kevin Brindle (KB), Dr Ari Ercole (AE), Dr James Brenton (JB), Dr Ferdia Gallagher (FG)

3. Minutes of previous meeting

Minutes of the previous meeting were reviewed.

4. Matters arising

A) EB to discuss the HPCS-SG Committee with Caroline Edmonds (CE) and Paul Alexander and ask them whether he should represent the School on this Committee

Completed, although it was noted that Paul Alexander as chair of that committee hasn't been provided with a secretary and is reluctant to hold a meeting until one is provided. As such there haven't been any meetings of the committee to date.

B) GW to contact PS and Stuart Rankin (SR) to discuss support for migration of users to the HPHI and decommissioning of legacy systems in WBIC and BCNI.

This has progressed well, although GW needs to talk to PC about details of the support arrangement.

AP1 – GW and PC to meet

C) BG to advise his staff to contact PS and SR and they would be signed up in the usual way.

BG has spoken to his team and there have been meetings between members of BG's team and P. Sumption to facilitate their use of the system.

D) JB and BG to consult users and find early adopters who could act as exemplars.

BG explained this was part of the conversation with getting members of his team set up with the correct access on the system. Now this has happened they would be able to start using it.

LD explained there had been a presentation given to James's team/other members of CI team and he has instructed his team to use it, at the moment, however, they aren't as they have access to and are already using a similar resource in the CI.

The Committee felt the resources needed to be opened to more people.

E) PC to put together various costings models to be discussed at the next Committee meeting.

Discussed later in the meeting.

F) JB and BG to discuss storage costs and bring their ideas to the next Committee meeting.

This has not happened. PC mentioned HPCS storage now has a fully costed model and suggested that any model adopted by HPHI should be based on this.

5. Terms of Reference

Carried over to next meeting

6. Review current usage of both imaging and genomics side (Inc. feedback from exemplars covering actions 3 & 4 from previous meeting)

WBIC

The new WBIC systems on HPHI have now replaced the legacy systems and all software has been transferred. Users are happy with the speed but there are some issues with documentation. The level of support provided by the HPCS team, however, has been good although there is still a lot of outstanding work to migrate data from the old system to HPHI.

GW highlighted that some new use cases from BCNI and some external sources to the original grant scope are providing some challenges. In general, however, the WBIC usage was felt to be good news.

It was noted that a resource from ISS is now embedded in WBIC to help with desktops, but HPHI support will be given by the HPCS team. This will address the outstanding migration work including some BCNI systems. Once this work is complete WBIC compute and storage needs will be provided by HPHI and all old kit can be decommissioned. The following figures relating to the WBIC usage of HPHI were discussed:

WBIC Cluster Utilization

95% availability, 37% utilization

16,238 core hours used out of potential 45,906 possibly core hours.

Storage

9.3TB used out of 208TB allocated Nexenta storage

400TB used out of 693TB Lustre storage

The stats highlight the need to monitor storage usage, given 2/3 of lustre storage has already been used. GW has a "to do" list to modify storage policy to, initially take to imaging group. Felt that some of this storage is being used for unrequired data. An archive policy is being considered where users should only keep data they need long term in a specified archive area.

It was noted that 2/3 of storage being used is not a big problem, as storage can be expanded relatively easily from HPCS infrastructure if required. PC stated that his team is happy with the relationship with WBIC.

Brenton Cluster

BG has only really been able to engage with this in the last couple of months. The cluster was meant to serve single cell genomics which only really came online in the summer. BG feels that aspects of it are very useful to users, but it's been difficult to know what is available and how to use it although meetings with Paul Sumption and AL have helped address this.

BG has been informed that 1.4pb of the storage is not backed up and therefore not that useful. PC clarified that it can be backed up to tape, although they don't routinely back up very large data sets.

It is felt that that when the Stem Cell Institute move into their new building (August-October 2018), their existing kit may no longer be available, and certainly will suffer from significant downtime. HPHI therefore will be very useful and so needs to be set up for this user community within the next 6 months to help with this transition process.

It was noted that between the submission of the grant and the availability of HPHI, JB's team had gained access to alternative resource which they are currently using.

The Committee asked what help BG needs to enable him to use the resource more. Policies are required on how to use it and an allocation of storage is requested. It was also suggested that there is a medium-term need for a charging model, that then need to be communicated to research group leaders, to be taken account of in any future grant applications.

LD commented that all the sequencing, research and clinical data from the Integrative Cancer Medicine Programme is planned to go in a database there, so there may be more demand from here in time.

BG commented that over the next 12 months he would probably require and fill a third of the 300tb storage.

AP2 - BG to ask his community to start using HPHI resources.

7. Review cost recovery including costing options and sustainability (covering action points 5 & 6 from previous meeting).

The Committee was asked whether there is a requirement or need to recover some of the costs.

BG thought it might be easier to build into new grant applications costs for core hours and storage which would include replacement capital costs and support.

PC commented that the WBIC portion is a functioning platform and the use case makes sense, however, the other portion of HPHI is now technically one generation out of date compared to CSD3, so any cost model would need to take into account that people will want to use faster newer processing.

There is still a need for a cost model so that people become used to paying something for the resource and not just getting it for free. The cost model could involve selling HPHI resources at a loss or at least at a lower rate than the latest HPCS cluster. PC suggested that it should provide a comparable cost to that for obtaining results on the newer system. There needs to be a parity in price per performance.

A strategy was suggested to get people with no money onto the system, initially for free, proceeding to a two-tier system with charging once the resource fills up. This will also generate a community who are used to working with this type of equipment.

The Committee asked PC to provide a cost model which harmonises prices across the two systems so the differences to users are invisible. Quality of service differentials also need to be mapped and put in place for when the system is full, so people can start paying for the high priority tier.

AP3 - PC to provide cost model and cost per core hour which is consistent across platforms to be shared with JB, GW and BG.

BG pointed out that HPHI also needs the flexibility for user to access various queues.

LD raised the question about how storage would fit into this approach. If the resource is free initially, how do you control people's storage use? What if they want large chunks of storage? It would, therefore, be a good idea to define an amount of storage up front that users can have for free.

It was noted some of the users of HPHI on the WBIC tenancy are not linked to the grant and that WBIC have a cost model that's based on charging per scan.



GW stated that it would be useful to have some storage apportioned to an institution that they can have as a long-term resource, to act as storage in perpetuity.

PC pointed out that although storage prices tend to reduce exponentially and there is a myth that this means the cost tends to zero. In reality this isn't true as it doesn't take into account the cost of the resource required to support it.

The Committee considered whether HPHI should be opened up as a clinical school resource. It was felt that, at the moment, it is sensible to concentrate on defining a plan to make it a resource that works for BG's group. It was therefore decided to focus on refining the resource for BG's and WBIC community first and then review usage at the next committee to decide on whether to open it up to the broader community.

PC mentioned the Hadoop and OpenStack infrastructure. The SBS community are currently pushing heavily for a virtual server service on OpenStack. His plans are to turn it into one model for all users and set this up as a standard platform providing virtual servers for research.

AP4 – PC to prepare a paper on how to use Openstack to provide virtual servers for research.

LD raised the issue of moving anonymised scanned data into HPHI. This is possible if the image doesn't identify the patient and there is a secure platform. As soon as a secure space is ready, this option will be available. GW and LD are working on the linkage project for clinical data and images.

The Committee suggested that EB should meet with JB and that AL attend as well to discuss usage of HPHI. PC is also to be invited to this meeting. It was noted that in the meeting there is a need to remind JB about his integrated cancer medicine requirements.

AP5 – AL to request a meeting for EB with JB and invite PC.

8. Report on WBIC and BCNI migration (covering action point 2 from previous meeting)

Not covered - to be carried over.

9. Report on ISO 27001 safe havens

The Committee was informed that the first internal project meeting was held the previous week, although funding for the project is not yet in place. A project manager has been secured and potential members of project board will be invited by PC shortly.

EB asked whether the aim is to end up with one safe-haven. RB stated that once a safe haven is provided by UIS the Clinical School will migrate to this. Although PC mentioned that the migration of existing safe-haven data is not currently in scope for the project.

GW explained governance and policies are separate from the ISO certification on the platform and how his current safe-haven tenancy is approved under the NHS toolkit. If there is a platform that is certified, and the use case is compatible then it enables users to build their safe haven tenancies on that platform.

LD asked about where IGT 2 will sit and stated that they will build a tenancy on the UIS certified platform.

PC explained access controls and layers of software security that will be part of the platform.

It was established that, at the moment, safe haven does not matter to BG and his use of HPHI.

RB explained if it was certified under the NHS IGT toolkit then it would be a platform that could be used across the Clinical School. It was noted that the project will need to gather the requirements of the initial users to make sure the platform is fit for purpose

RB requested that Rich Hutchinson from MRC Epidemiology be included in the project but if RB is on project board then he could represent their interests as well as Clinical School interests.

PC cautioned that the project team need to be careful not to broaden the scope of the project too far as this could delay the delivery of any solution.

GW would also like to be on the committee, as long as this does not make it too unwieldy. LD also needs to be involved as GW will be using ethics as defined by her.

The £57k in the HPHI budget assigned for a safe-haven was discussed and RB was asked to facilitate the release of this as appropriate. It was acknowledged, however, that this is small proportion of the required funds for the project. The rest of the money will need to be found and the solution needs to be sustainable. The University needs to look at this as an opportunity and not just as cost, if this project doesn't happen then it will miss out on research opportunities that won't be possible without it.

10. Committee Structures

Minutes of HPHI (Imaging) Management Committee (unconfirmed)

Not Covered

11. Review requests for use which fall outside the original intended use

West Cambridge 9.4T MRI Facility - the board was asked if funding is available for this new use case, which, is in keeping with the way WBIC are already using HPHI, but is outside the uses described in the original grant.

EB responded that there wasn't any funding, but the Committee gave the green light for them to use it now and think about recovery costs in future grants.

12. Confirm any other strategic deliverables not yet agreed/confirmed which the HPHI should deliver (deferred from last meeting).

Not covered

13. Date of next meeting

End of Feb.

Action Points arising from meeting

1	Meet to discuss the details of the support agreement for WBIC.	GW & PC
2	BG to promote usage of HPHI in his group	BG
3	Prepare an update on usage in respective areas for review at next committee	BG & GW
4	Provide cost model and cost per core hour which is consistent across platforms to be shared with JB, GW and BG	PC
5	Arrange a meeting between EB and JB (including AL and PC).	AL
6	Provide a paper on how to use OpenStack for virtual servers for research. The hardware is delivered but there needs to be a definition of how to make a service out of it.	PC
7	RB to talk to Cassie Bradley (PM for secure research project) about safe haven.	RB