



**23 February 2017, Bay 13, room 311**

Present: Professor Ed Bullmore (EB), Dr Guy Williams (GW), Dr Nicola Wilson (NW), Dr Lydia Drumright (LD), Dr Ferdia Gallagher (FG), Dr Ari Ercole (AE), Robin Uttin (RU), Richard Bartlett (RB)

## HPHI OVERSIGHT COMMITTEE

### 1. Apologies

Professor Kevin Brindle (KB).

### 2. Minutes

The minutes of the first meeting were approved.

### 3. Matters arising

An invite had been sent to Dr Paul Calleja inviting him to attend this meeting or send a representative (see previous minute action 12(c)) but no response had been received, despite two follow-up emails. It was noted by EB that UIS representation was key to the effective operation of the committee, and LD took an action to contact Paul Calleja and discuss UIS representation at this committee [**action 8(a)**].

Dr James Brenton (JB) had accepted the meeting request but was not present, and had not sent apologies. His absence was noted, and LD took an action to contact JB and arrange a meeting to get an update from him [**action 8(b)**].

### 4. Review and approve (or amend and approve) the terms of reference

The Terms of Reference have been amended (as per action 12(a) from previous meeting) and were approved, subject to confirmation of UIS representation. All previous academics invited to the committee had been contacted to confirm their willingness to continue to attend (as per action 12(b) from the previous meeting), three had agreed to continue to attend, the remainder were happy to step down.

### 5. Confirm the current state of readiness of the HPHI elements

GW summarised his report (see attached paper) on the Neuroinformatics portion of the HPHI. It had originally been due in March, but was delayed until Autumn due to a lack of access to the





hardware. The old system functionality has been replaced, and the user feedback so far has been good. Data migration is still outstanding. Work with 'foot soldiers' at UIS (e.g. Paul Browne) had been very productive. GW stated that they need at least a fraction of a Computer Officer (CO) in post to complete the outstanding work. There is a proposal from the review of WBIC IT that a 0.5fte post is dedicated towards application support, probably matched by a 0.5fte post for hardware support. Once a decision has been made seconded staff may be in post within 6 weeks (though the exact skill set of the post holder to be agreed, Linux expertise is definitely required). A one year 1.0 fte post for data migration is already advertised.

EB had spoken to a user of the system who had flagged barriers to entry, but these were confirmed as primarily platform dependencies in the scripts which were easy to fix, and support from UIS (Paul Browne) was extremely effective in resolving such issues.

There was no report from JB on the Bioinformatics (Genomics/Single-Cell) part of the HPHI. There was lengthy discussion about the Bioinformatics system and planned developments in High Performance Computing at UIS which can be summarised as follows;

- The HPHI Bioinformatics base hardware has been procured and installed, and provides some facility, which some SBS staff are already making use of.
- The intention was that eventually the secure data hosting will migrate to the HPHI (where they want to do anonymised data analysis and link it to imaging/genomics).
- A new UIS HPC facility called Peta5 (funded by EPSRC grant) is expected to ultimately subsume the HPHI facility
- LD has worked with Paul Calleja on a governance model for HPC, and a joint Clinical Medicine/Biological Sciences committee is proposed, jointly chaired by LD and Prof Nabeel Affara (Pathology), into which the HPHI Committee would feed.
- Paul Calleja had requested that SCM develop a cost recovery model (with a free element and chargeable elements to replace the equipment at the end of its life). EB confirmed the Neuroinformatics system would be cost recovered through charging for scans.

The following actions were agreed during the discussion;

- LD will request an executive summary of the Peta5 facility and proposed governance model from Paul Calleja, to be presented by him to the next meeting of this committee [**action 8(c)**].
- EB will do a presentation on the proposed Neuroinformatics cost recovery model as a potential model for the Bioinformatics portion [**action 8(d)**].
- The next meeting will be held in the Cancer Research UK Cambridge Institute, with JB attending to represent his portion of the system (and others from CRUK-CI if required) [**action 8(e)**].

## 6. Discuss how the free allocation of computation and storage should be calculated

Given that the lifetime of the equipment is 5 years, EB queried when that 5-year period starts (given that the equipment is not yet in use) as this would impact any cost recovery model. EB actioned UIS to confirm what would they need to recover and on what timescale [**action 8(f)**]. GW



has done forecasts for Neuroinformatics, and has discussed the 'doomsday' scenario with Paul Calleja, who said they could be flexible on retaining some essential data (~50TB) if required. GW asserted that long term data curation is essential while storage itself is less of an issue as it is getting cheaper; provision of compute is where the challenge is but can be scaled with availability of funding. LD stated that because UIS saved 20% on Peta5 as a result of SCM involvement Paul Calleja has assigned Health Informatics 20% of the £9m, which would serve SCWSBS and other collaborators within Health Informatics. EB noted that there was some inefficiency in the committee structures now, which should be revisited at a later date [action 8(g)].

### **7. Review any requests for use which fall outside the original intended use**

There was general discussion about how the existing hubs of imaging data relate, and it was agreed that any barriers between such hubs should be addressed, and longer term everyone with imaging or clinical data would be using a single system that incorporated Peta5 and HPHI. LD warned that Peta5 is only hardware, no staff funding, and consequently LD/GW were submitting a bid (in April) for posts to do work on provisioning services using the Peta5 infrastructure. RB asked whether it was now possible to extend the CSCS ISO 27001:2013 ISMS to include HPHI, LD said she would raise this with Paul Calleja and Paul Sumption [action 8(h)].

EB raised the question of WBIC clients outside the main facility, GW confirmed that data from the Behavioural and Clinical Neuroscience Institute (BCNI, part of Psychology in the SBS) would move to the HPHI Neuroinformatics, but there is a safe haven requirement for some of their data, GW not clear yet on how much of that is identifiable (and the new post referred to in section 5 above would identify that).

EB confirmed Psychology, BCNI, Anaesthesia and animal imaging are not in scope yet, and whilst the School should be inclusive about who uses HPHI (and successor systems) this would need to be built that into our cost recovery model (that model to potentially include all human and animal imaging, MRI/PET/CT scans, which will require projection of demand, including use by Radiology and PDN). This committee would oversee the creation of that model.

EB confirmed BG should have use of the Bioinformatics system (once complete), as he contributed funding which resolved overspend in another area. LD confirmed she would ask JB who should engage with Paul Sumption in UIS in the same way GW did for Neuroinformatics [action 8(i)]. In response FG suggested Dom McIntyre (who runs the core facility in CRUK-CI), and whoever replaces Peter MacCallum (as Head of IT & Scientific Computing).

### **Date of next meeting**

The next meeting will be held at the CRUK-CI, date to be confirmed [action 8(e)].

### **8. Actions**

- (a) **LD meet with Paul Calleja reference UIS representation at this committee**  
LD to confirm with Paul Calleja what representation UIS would have on the HPHI Oversight Committee



- (b) **LD contact JB reference update on HPHI Genomics/Single Cell**  
LD to meet with JB and ascertain the state of readiness of the HPHI Genomics/Single-Cell system
- (c) **LD contact Paul Calleja reference presentation at the next meeting**  
LD will request an executive summary of the Peta5 facility and proposed governance model from Paul Calleja, to be presented by him to the next meeting of this committee
- (d) **EB present Neuroinformatics cost model to next meeting**  
EB will do a presentation on the proposed Neuroinformatics cost recovery model as a potential model for the Bioinformatics portion.
- (e) **RB arrange next meeting**  
The next meeting will be held in the Cancer Research UK Cambridge Institute, with JB attending to represent his portion of the system (and others from CRUK-CI if required).
- (f) **LD contact Paul Calleja reference cost recovery model**  
LD to request from Paul inclusion of what costs they need to recover to sustain the HPHI function (wherever that may be) and on what timescale (including confirmation of when HPHI is considered to have begun its five year life).
- (g) **RB add committee structures to the agenda for the next meeting**  
To be discussed after the presentation from UIS [action 8(c)] above].
- (h) **LD discuss extension of CSCS ISO 27001 ISMS to include HPHI with UIS**  
LD to raise this question with Paul Calleja and Paul Sumption
- (i) **LD discuss appropriate implementation contact for Bioinformatics with JB**  
Discussion to include consideration of Core Facility Manager and Head of IT & Scientific Computing at CRUK-CI.