

1<sup>st</sup> November 2021

---

## Agenda Items

---

1. Apologies
2. Review last meeting notes
3. RG report back on QA in ConcePTION project
4. APM report back on legal paper
5. DS conference organising
6. Quality Assurance
6. DataSHIELD community governance / organisation
7. Other Sustainability Issues
8. Next meeting date
9. Any other business

---

## In Attendance

---

**Present:** Andrei S. Morgan (ASM), Angela Pinot de Moira (APM), Becca Wilson, Kim Cajachagua Torres, Simon Parker, Stuart Wheeler, Yannick Marcon, Elaine Smith.

**Apologies:** Juan Ramón González, Artur Rocha, Paul Burton, Rosa Gini.

---

## Review

---

Minutes were reviewed and agreed.

---

## Previous Actions

---

1. SP reported back from his legal team
2. ES circulated adverts for the conference
3. Conference speakers contacted and accepted:
  - SW/ES - Sido Haakma
  - APM contacted Mette Hartlev, Copenhagen Legal Professor to give keynote
4. ES set up meeting for Abstract decisions/entertainment planning (18<sup>th</sup> October at 2pm BST)

---

## Legal Advice (SP)

---

Having spoken with his legal team regarding a review of DataSHIELD, SP reported that the team would need a more focused approach. If the approach was to use DataSHIELD to access data internationally, this would constitute data sharing and would take the team one week to complete at a cost of a few thousand Euros.

SP also consulted UKRI contacts to ask about alternative options for institutes who can't resource TRES. UK Data Service may help with accreditation and SP will investigate further.

Accreditation is a massive gap for everyone; a framework is needed to enable data sharing.

Without a clear idea of what is being transferred, for a formal quote there would need to be a precise scope. As there is no certificate which states our software is compliant, could we produce a technical report which would be published on our website or produce a paper? We could pay for this analysis as BW has

funds to cover this, but we don't want to overlap with Copenhagen's DataPlus research project. APM will set up a meeting with the Copenhagen team to investigate, inviting relevant DAB members including ASM, BW & SP.

---

## Conference

---

- BW reported that as costs were minimal, they would be covered by our DataSHIELD Conference budget in Newcastle.
- Links to Remotely Green (social networking pre-conference event) would shortly be posted on Eventbrite
- Delegates attending are from a diverse background and include UK grant funders.
- Diversity of abstracts submitted is good.
- SW, ASM, KCT, SP and APM happy to act as Zoom co-hosts
- ASM, APM, SP happy to chair sessions
- Polling during conference
- From the data gathered on Eventbrite, BW suggested sharing the demographics at the conference. SW and ASM offered to help – BW to send anonymised data.

---

## Quality Assurance

---

- Rosa Gini's advice on QA in the [ConcePTION](#) project has yet to be presented to the DAB. ASM will contact.
- Installing new packages in consortia presents problems. Whilst packages that have been developed within the DataSHIELD team have been fully tested, there is no guarantee that new package developments do not pose disclosure risks. SW advised that the DataSHIELD team follows a process for non-disclosure and ensuring high quality.
- Challenges:
  - Without a governance policy for the whole project, there have been no roles outlined, no code reviews or statistics policies. This is a huge job.
  - It's hard to enforce developers to do checks – they're often focused on wanting to use the packages.
  - Statistical Methodology reviews need a statistician to complete them and cannot be automated. Code review models are much easier and could follow testing guidelines.
  - In the future, combinations of packages may interact and produce disclosure risks. Should the developer be responsible?
  - Disclosure comes in many ways – simple error messages can lead to disclosure.
  - A code review and audit takes time and funding. A starting point would be defining a governance policy: we need transparency. Can we prove dsBase is safe?
  - In the Open Software world, software either gets updated or it's dropped. Abandonware can have serious implications for research projects. A protocol for funding applications is needed.
- Solutions:
  - A model to use is dsBase – which can be built upon. Users need to prove they've done testing, before being ingested into dsBase (n.b. this is not the current model). This may incentivise testing, as researchers are focused on developing new functions rather than testing.

- BW suggested that core packages with dsBase are separate from community packages which could be installed at the user's risk. However, all fully tested packages sitting in one software repository would be better.
  - YM suggested an 'incubation system' which whilst taking time to set up, may be a good approach to mitigating disclosure risks.
  - Should the core team have responsibility for every package development OR should they have an advisory role and offer guidelines "this is the certification you need to do...?"
  - A developer could pay a fee of e.g. £5k for testing their new packages, which could provide a revenue stream.
- Examples:
    - In RECAP, every institution had to satisfy their own requirements which meant lots of different set-ups. As a result of Brexit, GDPR will inevitably change in the UK.
    - Using a GUI and its review section, this be used to identify what needs focusing on, to then include in a core package.
    - In Rock, its functionality allows separating profiles. Node sharing in LifeCycle (6.1. and 6.2) can be problematic if different versions are being used.
- Further solutions:
    - Could create tiers:
      - Stable version
      - 'Test' version which is riskier and might be disclosive with small risk of bugs
      - 'Wild west' version – cutting edge technology
    - Keeping small and focused may be better:
      - Stats review
      - Peer-reviewed articles
      - Review of Stats methodology
    - Recommendations on how to get approval – provide a checklist for non-disclosive packages: if this isn't completed, it can't be used.
    - All packages on our website should be transparent; metadata should be visible, with issues reported on GitHub for 3 years. There might be a live page with automatic daily updates and information available e.g. "release x used by x nodes for 3 years." This could include the citation of papers with names of developers who could be contacted by adopters directly. However, as researchers often change jobs, this might not be possible.
- Short-term 'fix'
    - The website packages list should be web-based. It's important to communicate to users that there may be disclosure risks. There needs to be a warning at the top of the page on the website "Disclaimer: no responsibility in using packages" and a similar "no warranty statement" SW to action. SW has created the packages page of the website, where he has defined what should be included in package development. The answers from developers sometimes omit disclosure reviews, which is concerning. <https://www.datashield.org/help/community-packages>
    - SW will also add summary information in columns, which will highlight which packages have not been tested. This may pressurise developers to implement testing. (Not the responsibility of DAB to add disclosive information).

- With a Governance meeting in the spring, these QA issues can be addressed and working groups created (looking at the wider perspectives of users/adopters and not just developers).
- Governance Meeting date set: March 17<sup>th</sup> 2022. This will be posted on Eventbrite.

---

## Other Sustainability Issues

---

- EPSRC Software Maintenance Grant application focuses on adapting disclosure controls in DataSHIELD (outside Health domain).
- 2 grants for UKRI 'DARE' aimed at updating the core infrastructure for UK, enabling access to multiple sites. Aim to scope and build in the next 3 years, the landscape for the next 10 years.
  - SP advised that the grant should aim less at highlighting an automated approach, as funders not looking for this – better to focus on computer aided decision-making.
- The Nottingham led DARE bid shows federated analysis working, also applicable in non-health setting.
- Would new Horizon 2020 (Feb. 22) call be applicable?
- All funding applications should include DataSHIELD: community guidelines are needed which cost in time and people to tailor new functions.
- Definitions of community teams is needed: who are the “core development team” “core comms team” – to be defined in Governance policy.

---

## Actions

---

1. UK Data Service may help with accreditation, SP will investigate further
2. APM to set up meeting with Copenhagen team to discuss scope of the legal paper. SP & BW to be invited
3. BW to share anonymised data with SW and ASM who will produce charts/graphs on attendees, which BW will share at the conference
4. ASM to contact RG
5. BW to send SP grant proposal for any comments
6. SW to add disclaimer to community packages page on website and highlight packages not tested

---

## Next Meetings

---

Monday 6<sup>th</sup> December 2021 13:00-14.00 GMT (14.00-15.00 CET) – special post-conference DAB discussion (Zoom link sent in calendar invite)

Monday 7<sup>th</sup> March 2022 13:00-15:00 GMT (14:00-16:00 CET) – full DAB meeting (Zoom link sent in calendar invite)

Thursday 17<sup>th</sup> March 2022 9:30-15:30 GMT (10:30-16:30 CET) – Special Governance meeting – please register:

<https://www.eventbrite.co.uk/e/datashield-community-governance-workshop-tickets-208582023447>