

26th November 2020

In Attendance

Present: Andrei S. Morgan (ASM), Angela Pinot de Moira (APM), Artur Rocha, Becca Wilson, Juan Ramon Gonzalez, Kim Cajachagua Torres, Rosa Gini, Simon Parker, Stuart Wheeler, Yannick Marcon, (Elaine Smith).

Apologies: Paul Burton.

Background of Board members

[Andrei Morgan \(Chair\)](#) – INSERM, Paediatrician (specialising in Neonatology) Epidemiologist, *RECAP* Preterm (Paris)

RECAP Preterm project – working to create a platform with very preterm birth cohorts around Europe using DataSHIELD to share data and perform analyses. Personal interest in free software development with many years' experience.

[Angela Pinot de Moira](#) – University of Copenhagen, Epidemiologist, *LifeCycle* (Copenhagen)

Working in *LifeCycle* for past 3 years on early risk factors for asthma and allergy. *LifeCycle* is a Cross-cohort collaboration building a data resource (the Child Cohort Network) – built on DataSHIELD (user of DS for 2 years). Fair data and fair research is really important.

[Artur Rocha](#) – INESC TEC, *RECAP Preterm*, *EUCAN-Connect* (Porto)

Senior researcher and System Centre Coordinator at INESC TEC (non-profit research institute) – building infrastructures for open science with a focus on personalised health research. Encountered DS while working in *RECAP Preterm* and believes in the DS way of doing things, it's fair and fulfils GDPR with importance of privacy preserving analysis.

[Becca Wilson – Liverpool University](#) Innovation Fellow HDR UK (Liverpool).

Background in planetary and space research, health informatics researcher for last 7 years where started working with Paul and DS team.

[Juan Ramón González](#) – Assoc. Research Professor, ISGlobal - Institute for Global Health (Barcelona) and Professor of Maths at Barcelona University. *ATHLETE* project

Leading a package to analyse exposome data in different cohorts across Europe; developing new tools and methods to analyse this data (jointly with Paul Burton). Worked with DS for 2 years and collaborated with Yannick on Resources – focusing on omics data. Just developed the Shiny app for DS. Goal is to extend DS capabilities for omics data.

[Kim Cajachagua Torres](#) - Medical doctor, graduated in Peru. Doing PhD in Generation R at Erasmus, (Rotterdam) in substance abuse and consequences in cardiovascular and metabolism? Planning to use DataSHIELD in PhD project.

[Rosa Gini](#) – Public Research Institute with access to Big Data Tuscany, Italy. *ConcePTION* project

Expertise in multi-database studies – merging different cohorts from disparate areas. Infrastructures necessary to drug safety. Main project - accessing the reaward safety of COVID vaccines. Participating in *ConcePTION* – methodology of studying safe drug use in pregnancy: started with DS community at that time. Working in collaboration with Morris Swertz on merging pipelines with Maelstrom. Very strict privacy regulations on data in Italy.

Simon Parker – Data Liaison Manager at [DKFZ](#) German Cancer Research Centre (Heidelberg) (previously Cancer Research UK)

German Human Genome archive – data coming from multiple sources. Other experience of SDC and disclosure control with outputs, where DS would fit (automated role – suppressions). An SDC pipeline will be interesting to explore.

[Stuart Wheeler](#) – Computer scientist at Newcastle University with experience of building distributed systems, HP Medware lab. (Newcastle).

Became involved with DS through projects sharing and managing data work. Now involved in testing, quality assurance, continuous integration and managing the software releases.

[Yannick Marcon](#) – Consultant – Epigeny (own company), PI at Maelstrom. Software engineer.

Involved with DS from the beginning whilst working on Maelstrom at McGill University. BioSHaRE program – had Repository system called OPAL and collaborated with Paul. Improves infrastructure – making DS more open – so not only based on Opal. Also on data analysis – data can be stored anywhere. Collaborates with Juan and his innovative ideas to enable new scalable analysis – will be working on the scalability of DS in the year ahead. Good to have certification – building the trust of the users.

History

DataSHIELD was conceived at a conference at McGill University in Canada by Paul Burton and his team from Leicester University. Computation methodology and infrastructure (piggybacking on Opal). Data sharing – distributed method, not much used at the time. [BioSHaRE](#) project connected research across EU in 2013/14 which enabled the proof of concepts of DataSHIELD to be built. BioSHaRE tested functionality – had good use-cases, which were published, forged new collaborations and the project grew from there.

2013 Bristol University – maintained software but few staff to develop it. Horizon20/20 projects, InterConnect and Tom Bishop became involved. Tried to think of alternate uses outside academia, e.g. for commercially sensitive data

Training courses & workshops were developed, as well as Wiki to help new developers, and a peer-support model in the Forum. The website and communications were professionalised. UK Software Sustainability Institute – Becca Wilson (BW) gained funding from them for further development to move towards a sustainable community project.

Paul's group moved to Newcastle in 2017 and wider community really started to develop around this time, particularly from the DataSHIELD workshops that were held where people got the opportunity to meet other users of the software.

Milestone in the project's history – today's Advisory Board meeting.

Strengths and Weaknesses of DataSHIELD: what should we be working towards?

- Courses and videos are ideal for new starters
- Wiki is really useful and Forum is good with quick responses, with good community help
- Once you are familiar with DS, analysis is simple (especially compared with other projects doing federated analysis)
- Software training – diagram for Information Officers e.g. how to create Data Protection Impact Assessments.
- When performing analyses, opening R – errors occur and it's hard to install packages
- Learning curve is steep at the beginning
- DataSHIELD analysis – needs to be made easier
- Problems when using different computing languages
- Servers often go down, which is frustrating for users
- Error messages could be clearer
- It should be easy to do simple things, and should be possible to do more complex ones

Ideas

Schrems II (data sharing between EU and US)– would DS be a good solution? Potential market for international collaboration, especially as UK leaves EU.

Forge community links – who might be interested in DS potential – particularly with international transfers?

For rare case analysis, one country might not have enough data

As a Board, collaborate with other projects – build bridges with federated analysis; enable interoperability as the “one system fits all” scenario is false.

Huge possibilities with biobanks and platform with genomics

Need to work on certification and IG acceptability – e.g. how to “validate” new packages created by others?

Terms of Reference and Decision Making (may take 6 months – 1 year to decide)

- What’s the minimum number of people to make meeting viable?
- How long should people be on the Board? If it’s an arbitrary 3 years, then no continuity.
- All have the power to collaborate and seek funding – as the group diversity spans disciplines
- As this is a community project, want ratification from the community itself at next DataSHIELD Workshop
- Vision/aim of the Board?
- What’s the expected work the Board will do?
- How can new people join?
- What commitment is involved?
- Number on the Board – chose 9 (with Paul Burton and Andrei Morgan an extra 2) – small enough to make decisions and diverse enough to bring expertise from different fields
- Decision making – consensus vote for now but need a way to resolve differences of opinion if they should arise.
- DataSHIELD already have both Equal Opportunities and Diversity statements
- Chair role is seen by ASM as collaborative, not hierarchical.

Future Agenda Point Ideas

Finances

Finances are an important consideration for the Board. Funding streams are precarious, particularly with the historical reliance of DataSHIELD on research funding which has meant that much of the time the project has been underfunded. This is likely to be an important focus for the Board – to consider strategies to maintain funding and ensure that costings for projects etc are appropriate. However, as an *advisory* board, we can only make recommendations, we are not able to tell anyone what they must do.

RECAP Preterm example of 5 key focus areas for sustainability:

- Comms
- Governance
- Infrastructure maintenance
- Outreach
- Capacity Building

All of these areas benefit from money/financial backing – but finances on their own are not a sufficient requirement for sustainability. Indeed, it's actually very beneficial to think about how to keep things going with an absolute minimum because this allows a lot of growth when there are times of greater abundance.

There was agreement that we need to think regularly about how the project can be financially sustainable. For example, how we might set-up a more formal body (e.g. a foundation) that is able to deal with concrete financial aspects could be something that we want to think about in the future. Currently, as DataSHIELD is reliant on research funding, diversifying is also an important idea – how can DataSHIELD be used outside of academia? Separation of funding into that which is required for the operational side (code, servers, infrastructure, etc) from the funding needs more orientated to building community (e.g. admin, training materials, communications and website hosting, outreach etc) may also be helpful.

Actions

1. All Board members to send an email with a photo and brief comments for website (**All – please complete by 7th January and send to ES**).
2. Create a section under the “About” section of the website, about the DataSHIELD Advisory Board, what we are about and how to contact us (**ES**) (NB: will contain Terms of Reference and other information once this is drafted).
3. Draft of Terms of Reference to be created for review at next meeting (**BW to coordinate; SP, AR, APM, KC, SW in particular expressed interest in helping with this; AR suggested he may be able to ask advice at a later stage of the Data Protection Officer at INSEC TEC**).
4. All Board members to think about requirements for other documents and longer term coordination of the board, e.g. length of commitment to Board, routes to membership for people to join, equal opportunities/diversity policies.

Next Meeting

Monday 18th January 2021 13:00 UTC, Zoom
(Zoom link sent in calendar invite)