

OpenSAFELY Oversight Board Meeting Notes 3-2-21

Ratified at 13-5-21 Oversight Board

OpenSAFELY Oversight Board Meeting Notes 3-2-21	1
Attendees	1
Agenda	2
Notes - under Chatham House Rule [except BG/NS/JM/AM]	3
[Item 3] Presentation on OpenSAFELY progress	3
[Item 4] Review Onboarding Process	6
[Item 5] Brief verbal update of Patient and Public Involvement and Engagement (PPIE)	8
[Item 6] Terms of Reference	8
[Item 7] AoB	9
Actions	10

Attendees

Date & Time	03/02/2021 10am-12pm
Goals	Update Board on work to date on OpenSAFELY Discuss Onboarding process for OpenSAFELY Agree ToR of Board

Attendees	<ul style="list-style-type: none"> ● Prof. Nigel Shadbolt (Chair) [NS] ● Dr. Ben Goldacre (OpenSAFELY PI) [BG] ● Seb Bacon (OpenSAFELY CTO) [SB] ● Jess Morley (OpenSAFELY; Policy Lead, DataLab) [JM] ● Amir Mehrkar (OpenSAFELY Head of IG and external relationships) [AM] ● Mark Coley (GP, BMA) ● Kevin Minier (Expert patient representative) ● Jeni Tennison (Open Data Institute) ● Johnny Stewart (GP, RCGP) ● Rony Arafin (ApHA) ● Stephen Evans (OpenSAFELY researcher; Professor of Pharmacoepidemiology, LSHTM) ● Simon Madden (OpenSAFELY NHSE/X Sponsor; Director of Policy & Strategy, NHSX) ● Chris Bates (OpenSAFELY-TPP; Director of Research and Analytics, TPP) ● Shaun O’Hanlon (OpenSAFELY-EMIS; CMO EMIS) ● Sam Smith (MedConfidential) ● Alex Freeman (Executive Director, Winton Centre for Risk & Evidence Communication, University of Cambridge) ● Sean Kirwan (OpenSAFELY NHSX IG; Senior Data Sharing and Privacy Manager, NHSX) ● Wendy Harrison (OpenSAFELY NHSE IG; Senior Lead – Data Governance / Deputy Head of Corporate Information Governance, NHSE) ● Hilary Winstanly (OpenSAFELY Admin Support)
------------------	---

Agenda

Documents circulated in advance:

- [OpenSAFELY Oversight Board TOR](#)
- DRAFT How OpenSAFELY works
- Onboarding documents:
 - [Onboarding process flowchart](#)
 - [Principles of OpenSAFELY](#) (updated [Website version](#))
 - [Onboarding new users to OpenSAFELY](#) (Updated [Website version](#))

Item	10:00 - 12:00	Topic
1	10:00 - 10:05	Introduction from the chair (NS)

2	10:05 - 10:15	Introductions from attendees (ALL)
3	10:15 - 10:45	<p>Presentation on OpenSAFELY progress - including discussion (BG)</p> <ol style="list-style-type: none"> Goals "Principles of OpenSAFELY" (Updated Website version) document. Progress against project milestones, notably April 2021 and Nov 2021
	10:45 - 10:55	Break
4	10:55 - 11:15	<p>Review Onboarding Process (Updated Website version) (BG/SB/SM) Flow Diagram</p> <p>An initial discussion of how data access will be managed, following on from previous discussions with NHSE, specifically: what access tasks are managed by OpenSAFELY team; what access tasks are managed by the Data Controllers for the data that the OpenSAFELY software runs across (in this case NHSE).</p> <p>Consider how this may change over time (e.g. during and after COPI, and with other datasets).</p>
5	11:15 - 11:25	Brief verbal update of PPIE (JM)
6	11:25 - 11:45	Discuss and agree ToR (ALL)
7	11:45 - 12:00	<p>AOB (ALL)</p> <ol style="list-style-type: none"> release of data by Local Authority or STP information (AM)
8	Actions	

Notes - under Chatham House Rule [except BG/NS/JM/AM]

[Item 3] Presentation on OpenSAFELY progress

BG Presentation on OpenSAFELY progress

- Key achievements to date:
 - [Published papers](#)
 - Platform running analyses across full pseudonymised records of 40% UK population
- Principles of what we were trying to achieve:

- Provide access to high volume of data, no data downloads, no unnecessary access to disclosive raw patient data
- All code to be shared for reuse
- Data curation and data management to be conducted in a pragmatically standardised framework: create once and share to everyone
- Low start up cost
- Completely transparent to everyone on the use of the data
- Platform completely portable and modular
- How platform was built with TPP / EMIS; rationale for keeping data inside GP IT systems
 - Brief summary of how data is represented within GP systems and transformation process to convert into research ready dataset
 - Creation and use of <https://codelists.opensafely.org/> for codelist curation
 - Example of code to generate a COPD patient cohort and synthetic dataset
 - Brief explanation of the levels of researcher data access and how this works
- Two brief demos:
 - OpenSAFELY jobs runner <http://jobs.opensafely.org/jobs/>
 - OpenSAFELY documentation <https://docs.opensafely.org/en/latest/>
- Example of vaccine effectiveness research study
 - Standard vaccine effectiveness template
 - Standard working environment including all the variables
- Example result from [NHS Service Restoration Observatory](#)

NS opens floor for discussion:

Question: What are the different levels of access in OpenSAFELY?

BG describes the levels:

- Level 1 = only TPP (or EMIS) has access. Full raw data - identifiable data. This is where linkage occurs. [Access by TPP data processor information only]
- Level 2 = real identifier replaced with pseudonym, still one row per event (but without free text information or documents; only coded data and removal of restricted/sensitive codelists) [Access by platform developers where strictly necessary; No access by researchers]
- Level 3 = 1 row per patient to answer research questions [Access by platform developers where strictly necessary; No access by researchers]
- Level 4 = summary aggregate data to answer research questions [Access by approved researchers]

BG notes that it is challenging to promote the benefits of OpenSAFELY without causing alarm of prior (and still existing) data sharing in health. The norm is that everyone downloads the level 2 data.

Question: What is the attitude to different levels of commercial use of patient data that could be enabled through the use of OpenSAFELY?

BG - difference between early work in OpenSAFELY and future of commercial use of data; currently OpenSAFELY purpose to respond to Covid-19 pandemic. In future, where the

public is satisfied, it would be good to explore how we can run analyses for commercial use, whilst still maintaining patient privacy. But this still requires ethical discussions. Currently, this is all out of scope of OpenSAFELY.

Question: How resilient is OpenSAFELY for changes in datasets and how you access data?

BG notes this is a big question. Key challenges:

- We need an environment in which we have access to pseudonymised event-layer data. We have built OpenSAFELY to be portable. It can run wherever the data resides.
 - We are happy to build OpenSAFELY in NHSD or NHSE. CSUs already have data lakes where we would also be happy to build.
 - We are keen to preserve the arrangement with EHR vendors. Not least because the secure environment without having to build it, but also because we get an ongoing and close relationship with the vendors who have huge skills and expertise around the data they themselves collect and manage.
- Funding to keep the project alive. It is simple to get funding for research money for single research questions. It is almost impossible to find open competitive funding to build open source software to help 100s or 1000s of people achieve their health data science ambitions. This is a major problem that we would like to tackle. Software and informatics have a low status in UK. This is different in US.
- We would like to be part of a broad collaborative ecosystem of teams like ours.

Question: Are there any safeguards in place that could prevent study codes to be carefully crafted to output identifiable data?

BG - all the risks are heavily reduced by the infrastructure. However, by writing your code in the open also reduces the risk. You would have to write several queries; have to be approved on to the platform;

We hope to write code that would also check the code meets the type of research question you have requested to run; this would help when many researchers join the platform

Opportunities to use mathematical techniques to check for uniqueness e.g., k anonymity that can be used to check for potential problems

Code is open and on GitHub people can see what you doing

We do small number suppression on tables and where there are tables that update over time we think it is best to do this through rounding e.g., every number to the nearest 5 as a barrier to puzzle attacks.

Every output requires two reviewers and so if none of the other checks had succeeded, the risk would be caught at this point.

We are building collaborations with researchers and developers on this issue.

Question: How reliant is OpenSAFELY on COPI for research?

All managed by NHSE currently under COPI notice.

Backup plans include: GP practices consenting to share data into TPP's ResearchOne; adding OpenSAFELY onto NHS D TRE or other TREs.

Question: Where might we sign-post our GP users to understand the security and anonymity issues that OpenSAFELY is managing.

BG - this should involve PPIE; OpenSAFELY website has information; our agenda content will be posted on the website

BG highlighted challenges - created appropriate versions to suit different audiences

NS - important to help allay people's fears to privacy concerns.

BG - summaries project milestones (see [TOR appendix](#))

Also early discussions with NHSD and others on whether OpenSAFELY can be helpfully deployed as a software layer to facilitate and enhance their existing TREs.

Online <https://docs.opensafely.org/en/latest/> provides extensive information about the platform

OpenSAFELY team has confidence that 10 analyses will be achieved by Nov 2021.

[Item 4] Review Onboarding Process

BG notes:

Onboarding people to OpenSAFELY is much more complicated than onboarding to a traditional TRE; in particular because for reasons of security and patient privacy OpenSAFELY does not allow downloading of data extracts or uncontrolled access to the event-level records in a remote desktop.

- External users need to be able to work with GitHub, docker etc.
- Need to be capable of working as computational data scientists and to an extent software developers, as much as traditional epidemiologists e.g. in Stata / R with data downloaded onto a laptop.

The OpenSAFELY team is small but we have an aspiration to generate graphic user interfaces that help people generate summary tables and graphs in ways that are familiar to them which, under the bonnet, produces OpenSAFELY code for others to reuse including study definitions etc.

We are providing support to 3 external teams for Jobs Runner and other features. We are looking to time box this through a series of intensive weeks.

We want good quality work to be done on the platform. This is because at least during the early phase people are likely to hold us liable for the quality of the research conducted by other groups.

We have set out some expectations in friendly terms for pilot users. In return for support to onboarded researchers and projects during the pilot phase we need users who can deliver *several* research outputs (not one-off papers) because of the time investment required on our side. For the early pilot users (but not the final grand service we are working towards) we also expect collaborative working. We also expect contributions back to the platform development as referenced in the principles document (such as codelists curation, documentation input, sharing of analytic code, blogs, etc.) as per the Principles document, and as per working practices of those in the platform already. There has been early very positive feedback from researcher groups to engage in this way; there are other research groups keen to learn more about this form of open collaborative data science research.

The group was presented with the proposed onboarding assessment process; this involves two key checkpoints outlined:

- NHS England evaluate appropriateness of project purpose, priorities of NHSE and ethics
- Datalab evaluation: data science skills;

Assessment is by application forms and evaluation by NHS E and Datalab.

Importance of NHSE oversight discussed: initially NHSE may manage applicants by the current single point of access mechanism for the C19 datastore and then review with the DataLab team regarding capacity for OpenSAFELY staff to onboard and support applicants, depending on the level of interest.

A possible third checkpoint was discussed: how to prioritise research teams so that most beneficial research gets carried out. This depends on the current capacity of the OpenSAFELY team to onboard researchers.

The formal grant application milestone describes a goal of onboarding five researchers; what is the level of interest?

BG describes:

- Some individuals are already writing some code against OpenSAFELY, although they cannot execute the code; this shows skills and demand in the community. There is a possibility that OpenSAFELY could execute such code on their behalf.
- The level of involvement and support required from DataLab team can vary between research groups in part dependent on the complexity of research question;
- Describes pilot candidates under discussion

Question: Is there anyone from HDRUK in the pilot groups of interest? Is anyone from HDRUK on the board?

Many of the people involved in the project and the board have some affiliation with HDR.

Question: Can the platform support experimental research as well as observational research?

Yes. BG notes that it is possible to develop code using the dummy data.

Question: Would groups managing RCTs be able to use OpenSAFELY data to analyse the records of patients in their trials, giving richer data?

In theory, yes; as a framework it could be used for any patient-level data.

Question: Who is governing the onboarding process and able to hold people accountable for dragging their feet?

This is a common problem for all data access. OpenSAFELY can make it more transparent about how long it takes to go through the process e.g. on the Jobs Runner. This functionality is being developed; also by publishing where people are in their application process, everyone else can also hold OpenSAFELY accountable. Once this feature is available the board can obtain useful metrics regarding the onboarding process as well as pace of delivery of results/outputs.

NHSE is the data controller and so would ultimately govern the process.

COPI is likely to be extended; onboarding should continue as planned.

[Item 5] Brief verbal update of Patient and Public Involvement and Engagement (PPIE)

Redevelopment of website; commission of videos to explain to the public (eg why access needed to large volumes of data); engaged and supported by Understanding Patient Data.

Plan / update to be circulated with the next meeting.

[Item 6] Terms of Reference

Question: the decision making authority of the board; who are you asking the question - who is the accountability board vs. the executive running the project? Is this an accountability board, advisory or a management board?

BG notes:

- This specific group [OpenSAFELY Oversight Board] was established in response to a requirement from Wellcome, to give oversight on delivery of OpenSAFELY as a software project; this board cannot give approval on which researchers get access to the data, but can give valuable input on related issues such as how privacy risks can be reduced to the extent that data access is safer than current norms. There are other lines of accountability already in place to govern access e.g. NHSE.

- OS is not a charity or company, but a collaboration.

Suggested this board therefore functions as an advisory group, not an oversight board.

NS notes: this group is not about the group permissioning who does or does not get access to the data, but rather helping to guide OpenSAFELY to [achieve its objectives](#). As a set of interested individuals we want to give constructive and critical feedback as critical friends; consider to highlight this issue of advisory in the ToR.

Suggested that more technical representation might be needed to achieve quoracy

[Item 7] AoB

Question: AM - should we be able to present data by STP?

Depends what you want to do with this data?

BG: once the data is in the public domain this then is out of our hands. BG gives an example for vaccine uptake by STP shows quite significant discrepancies. NHS as a large organisation can be difficult to cascade such variances to those groups confidentially to help them address local inequalities; but at the same time we are mindful that while public dissemination can be an effective route for internal comms, we do not want unnecessary and unhelpful press coverage that lends itself to unreasonable criticism during a health emergency.

Observing the risks mentioned above; what we have heard from Local Authorities in pandemic is on the whole LA/STPs find this information difficult to obtain and would find this helpful.

What will the STP analysis achieve? If you are an STP/PCN, they want to know you are “below/above” average compared to others. In terms of “doing something with the data” might vaccinations by STP be too high or not?

BG - considers this is actionable information.

Is the ethnicity explicit? [BG replied, this is explicit.] Concerns there may be overt racism if this comes out in public domain: if different ethnic groups are perceived not to uptake vaccination this could lend themselves on the receiving end of racist criticisms. Noted that this would be a risk for any national data on vaccine coverage broken down by ethnicity not just OpenSAFELY.

Recognised there is a variety of views. It is a difficult question. These decisions must be taken into account from an NHS E perspective (as data controller), acknowledging the balance of its role in policy, commissioning and as data controller.

Actions

Item	Action	Owner	Status Notes
1	Create Chatham House style shareable minutes	AM	Complete
2	Make clear in the ToR the remit of the Board i.e. that the Board is Advisory to help OpenSAFELY achieved it stated goals as part of Wellcome grant requirements	JM/AM	Complete (see TOR)
3	Reconsider quoracy to ensure enough technical capacity	JM/AM	Complete (see TOR)
4	Share paper on PPIE	JM	Verbal update for May agenda
5	After discussion email recommendation: TOR to be updated with papers being labelled as “for information”, “for discussion”, “for approval”.	JM/AM	Complete (see TOR)

END