OpenSAFELY Oversight Board Meeting Notes 13-05-21

Ratified at 30-09-21 Oversight Board

Attendees

Date & Time	13/05/2021 13:00 - 15:00
Attendees	- 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) - Chris Bates (OpenSAFELY-TPP; Director of Research and Analytics, TPP) - 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)
- Natalie Banner (Understanding Patient Data)
- Jenny Firth (Deputy Director, Data Policy, NHSX)
- Hilary Winstanly (OpenSAFELY Admin Support)

Apologies

- Wendy Harrison (OpenSAFELY NHSE IG; Senior Lead – Data Governance / Deputy Head of Corporate Information Governance, NHSE)
- Shaun O'Hanlon (OpenSAFELY-EMIS; CMO EMIS)

Agenda

Documents circulated in advance:
- Meeting notes 03-02-2021 OpenSAFELY Oversight Board
- Updated ToR (with suggested changes - **for approval**)

Item	13:00 - 15:00	Topic
1	13:00 - 13:05	Introduction from the chair (Nigel Shadbolt) ● Request for AOB ● Ratify meeting notes from 03-02-2021
2	13:05 - 13:20	Discuss and agree ToR (ALL)
3	13:20 - 13:50	New developments and progress against milestones (Ben Goldacre) ● Resource ● Staff ● Datasets ● Research outputs ● Tech (Sebastian Bacon) ● Onboarding of external researchers (Amir Mehrkar)
4	13:50 - 14:05	Patient and Public Involvement and Engagement (Jessica Morley)
5	14:05 -	BREAK

	14:15	
6	14:15 - 14:50	Open Discussion (Nigel Shadbolt)
7	14:50 - 15:00	AOB (ALL) - none raised
8	Actions	

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

[Item 1] Introduction; review of prior board notes

- 03-02-2021 meeting notes ratified
- Board members approve meeting notes to be placed on OpenSAFELY website

[Item 2] Discuss and agree TOR

- NS emphasised the role of the board, in particular it is advisory.
- Board discussed the following points and changes to the TOR:
    - OpenSAFELY (and its team at the DataLab), whilst primarily working with NHS England, are also working to deploy the OpenSAFELY platform elsewhere. AM to discuss with Wendy Harrison (who was not present).
    - Agreement to re-order 'strategic advice' role from second point to first point, emphasising the board's role to support OpenSAFELY in meeting the Wellcome grant objectives.
    - Can the board actually 'ensure' that we continue to have access to data post COPI?
        - 'Ensure' is too strong a word as (for example) the board is not a data controller for all the data sets; perhaps 'advocate' or 'promote' to recognise a broader influencing role.
        - Finally settled on 'help ensure'
        - Attention was drawn to members balancing their board advisory role alongside their work as individuals in their 'day jobs'.
    - Added wording to emphasise integrity and ethics:
        - 'To ensure that OpenSAFELY is working to earn public trust by upholding high standards around quality, ethics, and transparency'.
    - Decision to mark papers *"for information"; "for discussion"; "for approval".*

[Item 3] New developments and progress against milestones

Staff (update by BG)
- 7 new researchers, many of whom already embedded; positives and challenges with regard to managing a bigger team.

- Developing a team manual: very helpful & documentation is good.
- Main recruitment drive is for epidemiologists; but there are barriers to recruitment, such as:
    - Contract duration, which we are trying to extend to 2 years
- Senior Project manager (Becky Lumbard) starting at the end of June; will be very helpful to support JM and the rest of the team.
- Software developer team expansion
    - Tripled in size
    - Hires are excellent, all bringing new skills and expertise, everyone very thorough
    - Note one of our key values is 'helpfulness' and have been very successful in recruiting for this
    - Willingness of software developers to take pay cuts is testament to the value of the project

Datasets (update by BG)
- OpenSAFELY now operates across 58 million patient data; billions of rows of patient information.
- ISARIC data agreed for ingest ~100K patients admitted during the pandemic
    - Collaborating with ISARIC researchers who will work inside OpenSAFELY
    - Research topics discussed
- ONS Covid Infection Survey data agreed for ingest - random sample of general population who provide a Covid-19 swab and associated symptom information.
    - 10s of thousands of households
    - Important as helps address confounders; clozapine example discussed
    - Includes questionnaire on Long Covid (helps for cross validation as to whether patients who are self-reporting also have data in their electronic health records).
        - Such cross validation currently only possible with OpenSAFELY
    - Also contains a flag of whether a vaccine recipient is a health care worker
        - Will allow, for example, head to head comparison of Pfizer and AZ for vaccine effectiveness studies, including addressing various confounders.
- OpenSAFELY continues to explore ingestion of other datasets with interested data controllers:
    - COG-UK: describes which variant of Covid-19 each person has; important for monitoring vaccine escape; trying to work with HDR-UK to obtain data
    - Birth cohort datasets
    - PHOSP-COVID dataset - part of long covid research
        - Helpful for validation studies, for example, for those same patients, could you detect signals of long covid in other datasets?

Question: On the datasets does OpenSAFELY have accountability, responsibility and liability on data cleansing?

- BG: We use the data provided by other data controllers who take on such accountability as the data provider; but our work can help highlight and improve data quality because our code is openly shared and can be reviewed by others.

- BG: By working with OpenSAFELY tools and our team, data providers and researchers are able to re-use tools which helps to limit errors

Question: But what about the underlying raw data, such as that filled out by GPs?

- BG: Interesting philosophical point: the GP electronic health record (EHR) data was not made for us and research. We are using it as a byproduct of its intended main purpose - direct care.
    - This can result in potential ambiguity as to what a specific clinical code represents without the associated free text consultation for context: for example, is a salbutamol inhaler prescription denoting an underlying diagnosis of asthma, COPD or even a suspected diagnosis?
    - Does this represent an issue of data quality or the challenges of understanding and learning how to manage complicated EHR data?
- BG: Vendors also have a role in helping to identify and alert us to which parts of the system may have better or worse data quality; this can help as a useful mitigation.
- BG: In reality there are some situations where published papers have results that may misinterpret the clinical event data because, for example, clinicians have not been sufficiently involved.
- BG: Important to recognise that our OpenSAFELY codelists are openly shared, but also our cross validation checks; by sharing the underlying study code we are helping to expose (and discuss) any concerns amongst the research and clinical communities.

Question: How do we ensure records for someone who has opted out of health research do not get selected or used?

- BG: Under COPI it is a possibility that someone is in a research cohort where they have objected for their EHR data being used for research and planning.
- BG: NHS England manages all the information governance provisions for OpenSAFELY, including the data sharing agreements, in line with the COPI regulations and guidance. OpenSAFELY works within these rules.

Datasets continued (update by BG)
- OpenSAFELY within NHS Digital
    - OpenSAFELY team have honorary contracts with NHS Digital to build a data management and analytics layer; early days but feeling confident with delivery timelines
- OpenSAFELY within Integrated Digital Care Records (ICDRs)
    - A large number of ICDRs are supplied by Graphnet/SystemC
    - Often cover populations of 1-2 million people
    - They approached us for a data management layer in their environment
        - Graphnet/SystemC demonstrated they have used the OpenSAFELY open source code and built it in their sandbox
    - They are now working with their data controller organisations to provide OpenSAFELY for their ICDR customers
    - This raises the real potential of deploying federated analytics across millions of patient records with the records remaining in situ in their source ICDR

- OpenSAFELY functional in EMIS
    - We have accomplished [a federated analysis with TPP and EMIS](#)
    - Significant achievement because the underlying datasets in both TPP and EMIS environments have different data models
    - Now, using OpenSAFELY, if a researcher writes one study analysis code, this single code can be executed across both OpenSAFELY-TPP and OpenSAFELY-EMIS back end environments; the small number suppressed outputs are combined to create a result for 58 million patients

Question: Will this be an environment that uses secondary data for direct care?
- Lots of ICS seem to have a fine line between data for secondary data and direct care datasets
- From an OpenSAFELY perspective (as providers of an open source platform) it is important for us to understand our partners appreciate their responsibilities; although we will not be the data controller

Research outputs (update by BG)
- Description of research using federated analytics (e.g. benefits of keeping large volumes of disclosive data in situ)
- Research papers from DataLab/LSHTM available at www.opensafely.org/research
- An overview of some of the studies was given
- Some of the research pipeline was discussed, such as:
    - Vaccine effectiveness; Vaccine declines; Vaccine escape; Vaccine safety
- Overview of [service restoration observatory](#) outputs discussed
- Paper on [codelist attribution](#) informed by the OpenSAFELY work
- Production of short data reports - [first one on care homes](#)
- Brief update on Goldacre review related to OpenSAFELY

Tech (update by SB)
- Update on development roadmap - will bring this to the next board meeting
    - Planning to create roadmap with 6 week timeblocks of work
- Current areas being considered as a focus of development work:
    - Security
    - Dashboards - public dashboards reporting on KPIs, but also where patients and the public can review all the IG approvals etc.
    - Making it easier and faster for people to work on the data; the aim is to create a graphical user interface, for example for NHS analysts

Onboarding of external researchers (update by AM)
- [Summary of number of applicants and organisations](#) which have applied to be pilot external researchers
- [Suggested next steps](#)on external researcher roadmap
- Email board to arrange a sub-committee meeting of members interested in proposing KPIs

Resource (update by BG)

- Very easy to obtain funding for research questions but difficult to obtain funding for software or platform development that supports research; hoping to address this system issue in the Goldacre review
    - This remains an ongoing risk and constraint for our ongoing ability to maintain, develop and deploy the OpenSAFELY software platform
- Description given of the grants obtained, such as:
    - UKRI £1.3m principally research (split Oxford/LSHTM, but coming to a close) - not for platform development;
    - UKRI NIHR Long Covid £650k - not for platform development;
    - NIHR Long Covid two bodies - not for platform development;
    - Wellcome £2.3m over 3 yrs (for core resource to keep project running, but this needs to be supplemented to keep core team working and growing)

[Item 4] Patient and Public Involvement and Engagement

- More involvement of BG and JM once Goldacre review finished and senior project manager in role.
- JM participation in Citizens' Juries over last 6-8 weeks
    - Waiting for final report [Now published here; blog here]
    - AM was observer - very helpful to understand the transparent engagement with public
    - Jurors asked a wide range of questions
    - We were initially unsure about how easily we could explain the difference between OpenSAFELY being a software platform versus access to the underlying data; in the end felt distinction was well received and understood
- We have commissioned a video producer for a video / animation explaining the platform; this will be informed by the Citizens' Juries and wider team engagement
- In the tech team we have recruited a front end website developer who will help us improve the ability for the public to navigate our website and content
- We also produce a lot of widely viewed public explainers such as http://www.thedatalab.org/blog/189/opensafely-the-origin-story/

Question: you talked of jurors being interested in the platform; will you consider how OpenSAFELY fits into the bigger vision of healthcare research?

JM: Yes, and a lot of that will be informed by the Goldacre review; another challenge is how do you surface that "different", in terms of how OpenSAFELY is to other TREs, is good whilst attempting to be positive about differences and the diversity that is currently in existence.

[Item 6] Open Discussion

NG offered the opportunity to explore the particular challenge to obtain funding for core infrastructure and platform development rather than specific research projects.

Question: How do we promote the recognition and sustainability of platforms like OpenSAFELY? Are there other places, beyond the UK even, amenable to providing funding? Or is the UK OpenSAFELY's principal location to obtain funding from?

- BG: Would personally like to focus on the UK for a number of reasons, including wanting to deliver a benefit for the UK itself, but also familiarity with UK systems and processes.
- BG: Examples of other countries where they build platforms; it feels that in the UK there is a lot of money spent on projects that do not deliver. There seems to be a need to have difficult negotiations with organisations who receive a lot of money (compared to OpenSAFELY) but who have not delivered tangible outputs.

Suggestion: Is it an option that for any research project a proportion of the costs must be set aside for maintenance/infrastructure rather than research outputs alone?

- BG: Doesn't always work; OpenSAFELY is principally here to entice users to carry out secure research; it is difficult to ask users to then support the infrastructure and would potentially work counter to encouraging users to move en masse to work in more secure and open ways.
- BG: Ironically OpenSAFELY has shipped a lot of outputs despite difficulty in attaining funding; it does become uncomfortable to talk about other organisations that have received a lot of money but haven't shipped outputs.

Suggestion: Should we widen our horizons, say to include STPs or social care? Perhaps OpenSAFELY might receive more funding by pursuing alternative sources?

- BG: We have knocked on so many doors; we are not on the breadline, but we definitely want to help more people.

Suggestion: If we drafted a letter from Nigel and the Board to key stakeholders to gain support, would the Board be able to consider supporting a letter?

- A general consensus that the board would want to support such a letter.
- TPP would very much like to continue supporting e.g. the server costs £250K; these are not insignificant costs.

Suggestion: Might crowdfunding be an option?

- NS: we will need a multi-pronged approach but there is a serious case to be made for supporting a particular type of infrastructure; all too often these decisions are left to private institutions who do not have the public interest at heart. The board should advocate for this.

Question: If you were to write such a letter, would it raise any questions around the GP Data extraction planned by NHS Digital in July? Why are two GP data extractions safer than one?

- BG: to be clear, we support the goals of NHS Digital; however, we do not want the GP data to be disseminated. As such, we are helping to build OpenSAFELY within

NHS Digital. It will be an absolute disaster for the full GP dataset to be shipped off site from NHS Digital. There are many tools that make dissemination unnecessary. OpenSAFELY offers a safe alternative for working with such disclosive data.

Suggestion: There is research from the Open Data Institute looking at funding mechanisms.

- ODI have a workbook that we could work through: https://theodi.org/article/introducing-the-odi-sustainable-data-access-workbook/

BG discussed the proposal in the OpenSAFELY principles document, under the heading *"Contributing to best practice around open science",* to automatically make public the GitHub repository (to include the study definition, codelistsand released results) 12 months after any code [i.e. the first code] has been executed against the underlying patient data.

- BG: We expect all code to be released as soon as a project completes; but this proposal is particularly relevant for abandoned projects.
- Consider 1 month warning before the 12 month deadline is up; all exceptions to be brought to the attention of the OpenSAFELY Oversight Board. The ability for exception also supports changes in professional or family circumstances e.g. maternity leave.
- This is analogous to the pharma industry; we can learn a lot about failed projects to prevent wasting resourcing; this is an important aspect of learning.
- Board supports the principles document.
- AM: so far no external researcher group has suggested our principles and processes are unworkable or unreasonable.

BG discussed the importance of advocacy material to help change culture to promote modern collaborative data science. This will also partly occur via the Goldacre review and partly by sprinting to deliver the platform and OpenSAFELY research outputs. However, we have not yet produced all the important content due to time and resource constraints.
- We could divert team resource
- We could employ professional writers
- We could invite commentaries
But using internal resources comes with the opportunity cost of improving platform development; or onboarding new users onto the platform.

Question: Are there certain institutions we should be targeting?

- Experience of Open Data Institute is that communications is important
- A communications strategy is critical so that you target resources effectively; it is important because there is a lot to learn from OpenSAFELY; important to replicate and pass the learning to elsewhere.
- Also employing professionals has risks: they may not adequately understand what you do; it will take up your time to explain.

Suggestion: Could you use the people who use the platform as ambassadors? Sending them out to conferences and sharing the information?

- BG: Such ambassadors are useful in some, but not all, cases: there are challenges with working within a Trusted Research Environment (TRE), such as OpenSAFELY, where researchers need to learn how to use tools such as GitHub and Docker, as opposed to systems where they simply obtain the data onto their local machines.
- Why not choose as your ambassadors those who have learnt to use the system well?
- NS: I wouldn't be surprised if you need to bring in external capacity.

Actions

Item	Action	Owner	Status Notes
6	Uploaded May board meeting notes to OpenSAFELY website	AM	Complete Link: https://www.opensafely.org/governance/
7	AM to discuss relationship of OpenSAFELY wording in TOR with NHS England / Wendy Harrison	AM	Complete TOR amended from: 'The OpenSAFELY team working on behalf of NHS England is collaborating closely with Data Controllers on establishing the best methods for evaluating potential users' skills and projects…' To: 'The OpenSAFELY team is collaborating closely with Data Controllers (principally NHS England) on establishing…'
8	Create public dashboard of all approved external researcher pilot applications	AM	Complete https://www.opensafely.org/approved-projects/

9	Email board to arrange a sub-committee meeting of members interested in proposing KPIs.	AM	**Open**
10	Draft a letter from Nigel and the Board to key stakeholders to gain support?	BG	**Open**

END