Health Data Research UK's (HDR UK) Response to Goldacre Review Team

Response to follow-up questions on Trusted Research Environments and Data Management - sent on 11 May 2021

Many thanks for your note and I hope that the review is progressing well. The following are responses to your questions regarding Trusted Research Environments (TREs) and federated analytics.

Please find responses to your questions below. To maintain transparency as other organisations may also be interested we will also publish this on the following page here as it's in response to the document on TREs.

Goldacre Team – Question 1

Firstly, on scope, there are some projects we were hoping to hear a little more on that involve TREs, data management, and federated analytics, especially: British Heart Foundation Data Science Centre (BHF Data Science Centre) led by HDR UK; CVD-COVID (if this is regarded as separate from BHF DSC); and ICODA. For each of those we would be keen to know: the funding sources; the amount of funding from each; what it was spent on; and who has been doing the work? We would also be keen to see public URLs for any methodological work, code, tools or output. As before, we are trying to get a good overview of the current and planned TRE landscape.

HDR UK - Response 1

The BHF Data Science Centre CVD-COVID is in the NHSD TRE.

The work has been undertaken by NHS Digital staff working with members of the CVD-COVID Consortium.

The data dashboard is available here with additional information available via the HDR Innovation Gateway. Code is available here and further outputs are available here.

The International COVID-19 Alliance (ICODA) is convened by HDR UK.

It is supported by a grant from the COVID-19 Therapeutics Accelerator, a large-scale initiative initiated by the Bill & Melinda Gates Foundation, Wellcome, Mastercard with additional support from Minderoo Foundation, and other donors to speed the development of and access to therapies for COVID-19. Other funders to the ICODA include Microsoft, through its AI for Health program.

The ICODA "Workbench" has been separately commissioned by the Funders. It is provided by Aridhia Informatics and allows researchers to discover, access and analyse global multi-dimensional datasets while respecting confidentiality and privacy.

The Workbench brings together a combination of tools and processes and can connect to multiple regional or national data infrastructures used by ICODA members.

There are a range of partners, including UK TRE providers such as SAIL Databank, UK Secure eResearch Platform and Genomics England.

HDR UK has no involvement with funding of the ICODA TRE.

Goldacre Team - Question 2

Under NCS D&C is listed "1,204 COVID pre-prints using health data and 131 published papers". This is a small issue but it's not clear to us if these were using D&C resources, or selected by some other route?

HDR UK - Response 2

This includes the 330 research projects in the national data infrastructure and the additional health data research outputs identified by the HDR UK Paper/Preprint scraping code. The details are reported through the SAGE reports, latest here. Papers are also exported into the HDR Innovation Gateway periodically here and into Zotero to help researchers re-cite them in future papers/preprints.

Goldacre Team - Question 3

On page 4, as part of D&C work TREs were asked to "develop technical standards and components to enable advanced analytics across TREs (federation capabilities)". Could you tell us a little about what this was, and where we can best see the work?

HDR UK - Response 3

TRE Technical capabilities alignment survey is here and the TRE collections is here.

Goldacre Team - Question 4

Investment to date on "cohort discovery" is £126K (excludes investment through CO-CONNECT and Data & Connectivity). Please can you outline the additional spend from these two sources? Overall, beyond simple cohort counts, and in the space of federated analytics more broadly, it seems that CO-CONNECT, D&C, and the HDR Innovation Gateway have some overlap in ambition, plans, and possibly implementation. It would be good to understand the thinking around this, and the totality of the work, appreciating of course that overlap is often fertile and positive.

HDR UK - Response 4

CO-CONNECT is a UKRI funded programme that was funded prior to the development of the National Core Studies.

The Principal Investigators are Aziz Sheik, Emily Jefferson, Philip Quinlan and Susan Hopkins.

BC Platforms are the main technical partner who supply the software that drives the federated search. It is utilising their BC|RQuest, BC|Link and BC|Insight tools to help deliver a scalable, secure and powerful data discovery and analytics package to enable the datasets to be brought together without all the data having to reside in a single location.

It is aligned and co-ordinated with the development of the Health Data Research Innovation Gateway's (HDR Innovation Gateway) Cohort Discovery.

The ISCF Gateway investment to date has been to enable the user interface to be available through the HDR Innovation Gateway and enable direct link to dataset metadata and ability to enquire about access to the data. Further HDR Innovation Gateway development work is planned to support the integration of the cohort query into the data access request.

The Data & Connectivity National Core Studies programme collaborates with CO-CONNECT partners (specifically the main national data custodians / TRE providers) and so D&C is enabling the cohort discovery functionality being deployed through CO-CONNECT to cover a wider range of datasets than the serology data assets that are the focus for CO-CONNECT.

All three programmes share the objective of providing federated cohort search / discovery capability across a wide range of data assets for researchers, with the HDR Innovation Gateway as a common platform for accessing the functionality.

Goldacre Team - Question 5

Regarding the section on future work for federated analytics, this is clearly very ambitious, but we were unclear on its status, especially which aspects are planned and which are ongoing; and in particular whether it is unfunded, or partly or wholly funded. Please can you share some details on the funding

sources, the approximate amounts, which organisation or team is doing the work, and any early estimated goals and timelines?

HDR UK - Response 5

We are currently developing this with the research community and will involve you in the development of this.