

**MINUTES OF THE ELEVENTH MEETING OF THE EXPERT ADVISORY GROUP
ON DATA ACCESS (EAGDA)**

10:30-15:30 TUESDAY 18 APRIL 2017, WELLCOME TRUST

Present:

James Banks (chair)
Mark Elliot
Mark Guyer
Paul Flicek
Tim Hubbard
Andrew Morris
Onora O'Neill

Funders:

Audre (ESRC)
Rebecca Leithall (ESRC)
Rachel Knowles (MRC)
Katherine Littler (Wellcome Trust)
Natalie Banner (Wellcome Trust)

Apologies for absence:

Paul Burton
John Hobcraft
Bartha Knoppers
Melanie Wright
Mark McCarthy
Ros Eeles
Fiona Reddington (CRUK)

Also attending:

Graeme Laurie, Annie Sorbie (David Townsend by phone) on behalf of 'Data Terms' team
13:00-13:45

1.0 Welcome

1.1 JB welcomed the group and relayed apologies for absence.

2.0 EAGDA Review

2.1 JB provided an overview of the chair's paper and the need for EAGDA to reflect on and discuss what the group should be doing and focusing on. The funders' paper (Annex 3) is useful for context but the funders are keen to hear EAGDA's views independently.

JB outlined the two elements to the paper: an overview of the rapidly changing landscape and the variety of other bodies that are acting in the space of data access, use, governance and management; and a series of questions to stimulate EAGDA's thinking about its role, how it should be structured etc.

2.2 EAGDA discussed several aspects to the paper and the future of the group:

Wider context

- The UK needs a robust, reliable research health data infrastructure that is scalable, with reducing complexity and increased coordination.
- Many other bodies share the funders' concerns about data use in a wider context that clinical research and medicine. It's vital to consider the issues within this broader context, e.g. we have no definitive understanding of the impact the GDPR and its post-Brexit operation will have on data use for research, or what compliance will look like for data sharing across UK-EU.
- The legal structure for data use is bedrock and vitally important to understand but is unclear at the moment and funders should be thinking carefully about its impact on research and activities that impinge on research.
- The research field is not isolated from broader issues e.g. in the private sector over the responsibilities of data platforms over content hosting; the 'right to be forgotten' etc, and funders should have a grasp of the implications of these sorts of issues on research data use.
- Data sharing is actually getting more difficult and there are opportunity costs to not acting strategically and coherently to address the barriers and challenges now.
- Data runs through many key initiatives and groups: GO Science, the Office for Life Sciences, learned societies, NIHR, DH, the NDG, NHS Digital, ONS, ICO, the UK Life Sciences Industrial Strategy working group and the Turing Institute are all highly relevant stakeholders.
- Scotland has made substantial positive progress in setting up trustworthy, robust governance, oversight and ethical frameworks for health and biomedical data use for research, as a result of close collaboration with policymakers.
- There is currently little coordination between the multiple groups working on issues relating to data use and governance in research and conversations continue to be siloed. There's inertia and people have good reasons to maintain their siloes so funders need to take these challenges seriously and work out

what problems they want to address.

- It is anticipated that more of the process of data sharing will be formalised into contracts in future, where there were previously informal arrangements. Introducing contractual obligations will affect the way these systems are set up and governed.

2.3 Scope

- This is a timely point to consider EAGDA's role and remit, which has been restrictive as it has focused on cohorts and longitudinal studies when many of the issues it addresses have broader reach and significance.
- The reason for the initial focus on cohorts was the perceived value of these long term resources and the need for cross-disciplinary approaches.
- Is this still valid given complex questions emerging about archiving, data discovery, encryption technologies and data infrastructures etc that require similar cross-disciplinary thinking?
- Do funders want to prioritise better use of high risk or high value datasets?
- What constitutes high value: amount invested in creating a resource; cost of reproducing a dataset; uniqueness; social value?
- The boundaries of what constitutes health or health-related data are becoming more blurred as different data types can increasingly be linked across sources. If the group is to expand its interests beyond cohort and longitudinal studies, clarity is needed on what its limits should be.

2.4 The current EAGDA model

- EAGDA was prescient at the time it was set up but has been disconnected from delivery and implementation, and from strategic funding priorities.
- The group is not necessarily getting feedback from the funders on what is changing and how EAGDA's advice is being taken up, or why it is not. The extent to which funders' policies and practice is being informed by EAGDA's discussions and advice is not clear.
- Is there buy-in at an executive board level for the work EAGDA is doing? This strategic power is vitally important if EAGDA's work is going to be useful. EAGDA needs clarity on where value for funders lies.
- EAGDA raises issues the funders wouldn't necessarily see or be able to discuss themselves, and this is valuable. The second step is consensus building about how to take issues forward and act on them in a coherent way – this takes real resources and convening power which is what EAGDA has not managed to do so effectively to date.
- EAGDA is not equipped to map the landscape of data use and management at the moment or fulfil a watching brief on law, technology and ethical issues. This would require both resource and operational and strategic commitment from a wider group of funders to cover off the breadth of the space.
- EAGDA to date has had three types of role: an advisory 'service' role for funders; undertaking fact-finding initiatives and evidence gathering; and demonstrating leadership in discussions about data use and management [comment via written response].
- What does it mean to be 'advisory'? The group needs to be clear on who is going to act and how as a result of advice is given. This will necessitate a

deeper connection to funders and government, especially if the constituency EAGDA serves is widened beyond cohort and longitudinal studies.

2.5 **Recommendations on scope and model**

- There are many UK bodies working in this space: the focus of an advisory group should be to provide big picture thinking and advice across disciplines.
- Any such group should be international in scope but this requires substantial focus, resource and the right independent membership to be part of the conversations e.g. about international standards for de-identification or big data more broadly.
- Funders appear to want to both expand the group's remit but also for it to provide more focused and practical outputs. These requirements are in contention.
- EAGDA wanted to push back to the funders the question of what scope most interests them, recognising that the cost of increased scope could be coherence and the ability to provide practical advice.
- The challenge here is to reconcile the competing demands of producing outputs that are specific enough to be directly useful but broad enough to be widely relevant while understanding that the issues are constantly broadening.
- What does it mean to be 'advisory'? The group needs to be clear on who is going to act and how as a result of advice is given. This will necessitate a deeper connection to funders and government, especially if the constituency EAGDA serves is widened beyond cohort and longitudinal studies.
- A future EAGDA could work on a two-part model:
 - 1. Raising issues funders wouldn't necessarily identify.
 - 2. Consensus building about how to take issues forward, bringing a wider audience on board to work out how to do things about it, e.g. convening, consensus building, connecting with other players in the landscape.
- To act on these issues in the right way takes real strategic resources and convening power. This needs willingness to organise and buy in at board level.
- Identifying evidence gaps and issues on the horizon is challenging given these siloes: there is no single list of key issues in data ethics, for example, as most interested parties are concerned with specific issues e.g. data protection/privacy, data security, standards etc depending on their own priorities and perspectives.
- There is now an 'industry' in data use and management: many more people are interested in this space than in 2012. This is positive but has the disadvantage that multiple conversations take place in the absence of clear and agreed principles, which means there is duplication of discussions and no substantive normative moves forwards. A group should seek to alter the 'ground state' so that progress can be made beyond these early discussions repeated in different contexts.
- A future group should look to connect between siloes, connect better with the law and with new technologies.
- If funders want the group to have some authoritative oversight of other initiatives in this space, convening and bringing together different perspectives,

including different disciplines and data types, it is questionable why it is this group of funders and not a significantly larger group involved – especially with the creation of UKRI on the horizon.

- One potential way of working could be to convene groups of people to identify where there are policy-relevant evidence gaps and then recommend that the funders ought to do something to address those, e.g. by working to encourage harmonisation or convening different organisations to work towards adopting standards etc.
- If EAGDA funders wish to continue the group with its present group of funders, the group will only really be relevant if it has a more public-facing role and can push its deliberations and recommendations out better and further, both to stimulate real progress and hold the funders to account.
- There are two feasible directions for future activities:
 - Being outward facing, reaching out into the research community and beyond (e.g. into policy making) and making messages more powerful and coherent. This would necessitate a larger group of funders and stakeholders around the table to bring the right convening power across disciplines.
 - Given the complex landscape, focus on helping funders get their priorities and strategies right and coherent with inwardly-directed advice for them, involving practical guidance.
- A discussion with Sir Mark Walport on coordinating discussion about the role of this sort of group could be valuable at this point in UKRI's planning and development.

The funders thanked EAGDA members for their insightful comments and queries.

2.6 **ACTION:** Secretariat to write up comments and views on this for quick distribution, to feed into final report to funders, due for June 2017. Funders to take a decision on future plans and report back to EAGDA in Autumn.

3.0 'DataTerms' project

Graeme Laurie introduced the legal and literature review the team had undertaken, exploring the concept of data ownership and how that is interpreted. Key insights were:

- Data sharing is generally understood as a positive thing, and 'ownership' of data is perceived as a barrier to good data sharing.
- Allusions to ownership appear to be proxies for control, data security and the protection of privacy.
- Appeals to data as 'property' are related to arguments about IP (patents, copyright etc).
- The DataTerms project anticipates it will be useful to unpack the references to ownership as there is a conceptual and practical confusion about privacy and property, both of which are about seeking control over data but which have very different legal bases.
- From the literature review, the survey is being developed, which will incorporate broader issues of incentives and opportunity costs in an effort to unmask the behaviours underlying appeals to property and ownership concerns.
- The survey will seek a broad range of representation internationally and from different parts of the research system – aiming for around 150 participants across registries, genomics, longitudinal studies, regulatory bodies and non-

medical dataset users. Focus groups will also build on the survey findings. The full report will be available in October.

- 3.1 EAGDA members thanked the team for the update and asked questions:
- Ownership references are made colloquially, i.e. in reference to 'my' data, which does not distinguish between 'data that are mine' and 'data that are true of me'. It will be important to recognise entrenched and rhetorical uses of property terms and where these do not genuinely indicate that people are thinking about data in terms of a legal notion of property.
 - The issues raised may apply to all kinds of data, beyond health data. The project should ensure it is clear why the focus is on health data, i.e. that there are additional regulations and confidentiality issues to manage in relation to these types of data.
 - Is it plausible that poor/sloppy language is driving overly restrictive behaviours? It may be beyond the scope of the project but exploring GPs attitudes towards their control of their patients' data would be interesting in this regard.

- 3.2 The DataTerms team responded to the query on language by suggesting that the law is often presented as a mask for not sharing (e.g. appealing to data protection law to deny access, even where legitimate). The law may support different behaviours and not mandate sharing data, or offer incentives to do so, so there is latitude for different interpretations of what is appropriate.

Lessons for data sharing may be learned from other fields, e.g. physics, where collaboration is far more established because there is a more collaborative funding model and large scale infrastructures are required: this is a fundamentally different way of thinking about research, coordination and reward.

- 3.3 **ACTION:** DataTerms team to roll out survey – EAGDA members to snowball to relevant others as needed. DataTerms to submit final draft report to EAGDA in October.

4.0 **Sanctions**

NB provided an overview of the edits made to the papers on sanctions – splitting into an evidence paper and a position paper that sets out a series of issues for funders to consider and take forward.

- 5.1 EAGDA commented on the latest draft:
- The group agreed the paper should include 'Accountability' in the title, to reflect the shift in emphasis towards a broader approach than just punitive sanctions for wrongdoing.
 - Analogies with Health and Safety requirements could be made in the paper, which also covers training of staff and security systems. Existing information governance best practice (e.g. IG Toolkit) should also be mentioned.
 - Other sectors that use self-reporting to good effect could also be mentioned, e.g. air traffic control.
 - The spectrum outlined should include suggestions for audit and compliance, systematically mandated by funders, as tools for improving and incentivising good behaviour. These should be proportionate so they do not become overly burdensome or an end in themselves.
- 5.2 **ACTION:** Secretariat will make amendments to the position paper as suggested, with a view to sign off at the next meeting.

5.0 Data Management Plans

NB provided an overview of the changes made to the Data Management plans paper in response to comments made at the last meeting.

NB also highlighted that the current draft was feeding into the development of Wellcome's new policy on consolidating data management into a broader outputs management plan, covering papers, data, code and software. This should be released before the summer, under the management of the Open Research team at Wellcome.

5.1 EAGDA commented on the latest draft:

- It is helpful to identify clear actions for funders to highlight the need to have data on where they are now and what changes they want to see: Do funders know what they spend on data management and sharing? How much are we spending to resource data sharing? It is only through this type of benchmarking that progress can be tracked and lessons learned.
- Reporting requirements for grantees is key, both to hold them to account for what they commit to and the costs of data sharing and for enabling funders to ensure their funds are spent efficiently, e.g. within individual HEIs where data infrastructures and management could be coordinated.

5.2 **ACTION:** Final version of the paper to be circulated with minutes and formally submitted to the funders over summer.

6.0 EAGDA updates

6.1 Data infrastructures meeting

NB outlined the meeting held at the UK Data Service in October 2016, which brought together people from different research disciplines to discuss opportunities and challenges for data infrastructures and curation.

Participants agreed it was a valuable meeting and it could be worth considering a further one in 18 months/2 years to see how the ideas and initiatives discussed have progressed.

ACTION: Secretariat to put meeting report and list of participants on EAGDA website.

6.2 IJE Editorial

Because of a change in editorial staff, the writing team (led by Paul Burton) need to resubmit the editorial. EAGDA will be updated when we have a publication date.

7.0 Member updates

7.1 Paul Flicek reported that Ewan Birney has been appointed as the new chair of the GA4GH. He will focus on changing the Alliance's ways of working to make it more focused on delivery and practical applications.

7.2 Andrew Morris reported that Health Data Research UK would formally start in August, with a focus on developing the capacity (both people and skills) and methods to accelerate the pace and scale of health and biomedical data science across the UK.

8.0 Funders' updates

8.1 **CRUK:** CRUK will soon enable pre-prints to be submitted for grant applications as part of revisions to applications and review process.

8.2 **MRC:** MRC are flagging concerns through the Research Advisory Group for NHS

Digital about historic consents and their validity for NHS Digital's tightened governance procedures. It is not yet clear how these concerns are going to be addressed but at present the issue is preventing access to some data for historically consented studies.

8.3 Wellcome:

- Wellcome' Open Research team are developing a new policy on outputs management, to ensure that grant applicants consider enabling access to their outputs at the outset of their grant process. This will cover papers (under the Open Access policy) alongside data, software and code.
- Wellcome's Board of Governors recently approved a new Longitudinal Studies strategy which will bring all funding and decision-making about longitudinal studies under one funding mechanism. This will enable decisions about e.g. data governance (such as ETA-DAC) to be considered more strategically.
- The Strategy also has a key aim to improve and enhance the linking of data and use of novel data sources to enhance cohort and longitudinal data.
- Understanding Patient Data (also supported by MRC and ESRC) now has a live website providing tools and resources to help improve conversations about how health information is used.

9.0 ESRC Longitudinal Studies Review

- 9.1
- The ESRC Longitudinal Studies Review is in progress, gathering evidence until July with the panels due to report to the ESRC Council in January 2018.
 - ESRC will share papers with EAGDA on data linkage and harmonisation in due course, for input.
 - One of the key challenges for the Review will be to identify how better access to health data could substitute for expensive data collection directly from cohorts.

10.0 AOB

- 10.1 Onora O'Neill announced that she would step down from EAGDA with immediate effect. On behalf of EAGDA and the funders, JB thanked Onora for her exceptional contributions to EAGDA's work over the past five years.
- 10.2 Dates of any further meetings will be decided once the funders review and plans for the future are complete.
- 10.3 There being no other business the meeting was closed.