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

10:30-15:30 TUESDAY 4 OCTOBER 2016, WELLCOME TRUST

Present:
James Banks (chair)
Paul Burton
Mark Elliot
Mark Guyer
Paul Flicek
Tim Hubbard
Justin Riordan-Jones
Bartha Maria Knoppers
Andrew Morris
Onora O'Neill

Funders:
Fiona Reddington (CRUK)
Rebecca Fairbairn (ESRC)
Geraldine Clement-Stoneham (MRC)
Rachel Knowles (MRC)
Katherine Littler (Wellcome Trust)
Natalie Banner (Wellcome Trust)

Apologies for absence:
Ros Eeles
John Hobcraft
Mark McCarthy
Melanie Wright

Also attending:
Dave Carr (Wellcome) 10:30-11:30
Alice Jamieson (Wellcome) observing 10:30-15:00
Sara Marshall (Wellcome) 10:30-11:30
Frances Rawle (MRC) 13:15-13:45
Beth Thompson (Wellcome) 12:00-12:30
David Townend, Graeme Laurie, Katarina (on behalf of ‘Data Terms’ team) 13:15-13:45
1.0 Welcome

1.1 JB welcomed the group and relayed apologies for absence. JB welcome Justin Riordan-Jones from the Department of Health, attending the meeting following Peter Knight’s departure from DH.

1.2 The minutes of the ninth meeting of EAGDA, dated 26 April 2016, were ratified.

2.0 Matters arising and EAGDA updates

2.1 Item 2.1: PB updated on the editorial for the International Journal of Epidemiology on the Governance of Data Access report. The editorial was submitted but there has been a change in editorial staff. PB is following up with the new team to gauge their interest in the submission. He will update EAGDA on progress.

2.2 Item 9.3: The data infrastructures meeting will be going ahead on Tuesday 11 October, with several EAGDA members participating. The Secretariat will compose a summary note of the meeting for circulation to EAGDA and funders after the event.

2.3 Item 3.1.1: JB and NB attended a meeting on data governance on behalf of EAGDA, convened by the British Academy and Royal Society. The meeting was the starting point for a joint RS/BA project that aims to map out the current landscape of data governance across multiple sectors and identify existing initiatives and case studies. The initiative has recently launched a call for evidence and will attempt to identify characteristics of a good model of governance for new and emerging data-driven technologies.

EAGDA discussed the need for this wide-ranging project, as many conversations about data governance are being conducted in siloes at present, even though the governance issues may well overlap across different contexts. This has been seen for example with the Investigatory Powers Bill. This primarily concerns issues of data security insofar as they relate to national security, but the governance issues raised and solutions discussed are highly relevant to and interdependent with data in the research sector.

It was agreed that EAGDA should continue to engage with this project and seek to be ‘on the front foot’ to help break down the disciplinary siloes that create fragmentation in data governance (involving language, culture and different perceptions of what good data governance might look like).

3.0 Funders’ updates

3.1 Wellcome Trust:

3.1.1 DC updated EAGDA on the development phase of Wellcome’s the open research priority area, which is looking for opportunities to build on the work Wellcome has done to promote open access and maximise the value of research outputs, including data, software and code. EAGDA’s prior work on establishing incentives for data sharing is feeding into this development. Current activities include:

- Surveying Wellcome and ESRC funded researchers to identify barriers to data sharing across different disciplines. Current practice is highly variable but common ones include fear of losing publication opportunities, lack of...
resources to support data and code sharing activities and lack of academic recognition for such activities.

- Support for pilot initiatives with communities that are ready to shift towards greater openness, to demonstrate the value of making data more widely available.
- A new research platform has been launched with F1000 called ‘Wellcome Open Research’, which enables rapid publishing of any output, including negative findings.
- The Open Science Prize (partnership with the Howard Hughes Medical Institute and NIH) now has six international teams building new tools and prototypes to promote and enable better sharing of open data.
- Wellcome is promoting access to clinical trial data, running the independent review mechanism for the pharma-led clinicalstudydatarequest.com. It is now looking into whether the platform could be opened up to include academic clinical trials as well.

3.1.2 • NB updated on the independent patient data taskforce, being led by Nicola Perrin. Among key activities of interest are:
  - The development of an accessible, meaningful vocabulary to describe different levels of data identifiability, which all stakeholders could adopt to avoid confusion over the meaning of terms such as ‘anonymised’ data.
  - A bank of case studies on how patient data is used/has been used that will be a searchable resource for clinicians, the public, media, etc.
  - Horizon scanning for the impact that new data-driven technologies will have on public conversations about the use of patient data for research.

EAGDA welcomed these developments and asked for DC and NB to report back on these at the next meeting.

3.1.3 • Sara Marshall informed EAGDA of Wellcome’s current work to develop a cohort strategy. Wellcome’s investment in cohorts has generally been response-mode but there is now a need to develop a series of transparent principles and metrics by which to judge the progress of individual cohorts, including how they should be sunsetted when they have come to the end of their useful life. Wellcome will come to EAGDA for their input on some of these issues early in 2017.

EAGDA highlighted the need to identify the distinction between the value from cohort studies and that of big data mining: the inferential strength from cohorts is distinct from what you can extract from other, less controlled, data mining activities.

The National Institute on Ageing’s Behavioural and Social Research division (BSR) have recently reviewed their behavioural and cohort data infrastructure investments and strategy, which could provide some useful pointers.

**ACTION:** EAGDA to be updated as needed. JB to put Sara Marshall in touch with contact at the NIA if requested.

3.2 ESRC:

3.2.1 • RF reported on the ESRC’s current review of longitudinal studies. It is a decade since the last review and the landscape has changed markedly, especially in
light of new and emerging possibilities for data linkage, the development of informatics research and the closure of the Life Study. ESRC wishes to develop a clearer vision for its investment in longitudinal studies, identifying their roles and strengths at different life stages.

3.2.2 EAGDA members agreed to provide recommendations for international contacts who may be good panel members for the review.

It was noted that there may be valuable material from MRC’s background work for its landscape of cohorts published in 2014.

The Maelstrom initiative in Canada may provide a useful comparator for maximising data harmonisation across cohorts.

**ACTION:** Link to review call for evidence to be circulated with the meeting minutes.

3.3 MRC:

3.3.1 GCS reported on the plans for a new interdisciplinary research institute in biomedical informatics. It has secured funding of £37.5m from MRC over five years and will be a partnership between HRA, ESRC, EPSRC and other charities and funders. The search for an international director will commence imminently.

3.3.2 EAGDA welcomed the development and commented on the challenges of providing platforms through which to access data. Government and academic siloes continue to exist; data linkage and informatics research is hampered when there are multiple data owners with different requirements, restrictions and processes. To capitalise on informatics potential there has to be willingness, infrastructure and resource both from data producers/controllers and from those wishing to use the data.

EAGDA further discussed where commercial interests might sit in relation to this plan. Commercial actors could be users of data about to increasingly third party suppliers of platforms, cloud service and software, acting as data processors.

3.4 During the discussion it was queried whether funders or researcher knew whether ISO standards were being developed or updated in relation to data access and consent; and if so whether these have any bearing on the current legal framework for data use or codes of ethical practice.

4.0 EAGDA member updates

4.1 Mark Elliot (on behalf of Melanie Wright):

- ADRN has faced significant barriers in gaining access to government-held administrative data for research purposes. Progress has been made recently, for example with DWP to enable access.
  - It was noted that there are parallel issues with the Farr institute: data sit in government-held departments/NDBs and although research structures are being set up to enable better use of the data, the efforts are undermined if the data controllers do not make substantial efforts to open up the pipeline of access.

- The ADRN is hosting a ‘Talk Big Data’ speaker series: information can be found here: [https://adrn.ac.uk/news-training/courses-seminars-workshops/](https://adrn.ac.uk/news-training/courses-seminars-workshops/)
4.2 Mark Elliot:
- The UKAN Anonymisation Decision-making Framework has been well-received and UKAN have requests from a number of jurisdictions to adapt it to their contexts, including Australia, Turkey, the US and the EU in light of the GDPR. It is also consulting with the EMA regarding their anonymisation policy.
- The international networking UKAN is doing is revealing many duplicated conversations that could benefit from siloes being broken down across disciplines and borders.

5.0 Data Protection Regulation
5.1 Beth Thompson gave an update on the implementation of the EU General Data Protection Regulation:
- The Regulation will take effect from May 2018. Powers will largely be delegated to member states meaning that the status quo will be maintained and a fragmentary set of laws across Europe is likely.
- The focus of Wellcome’s joint advocacy work has been on ensuring robust implementation across the EU, with research organisations feeding into discussions. It has produced two resources: an analysis of key articles relating to research and their implications; and a joint statement, published by Science Europe with predominantly EU signatories, asking member states to take research into account and harmonise across countries where possible.
- It is very likely that the Regulation will be implemented before Brexit is complete and it is as yet unclear how it will be handled. Government has indicated it would transpose all EU law into UK law in the first instance to avoid a legal vacuum.
- DCMS is keen to get input from the research community on how to handle this significant piece of legislation that cuts across many sectors.

5.2 EAGDA thanked Beth for the update and her continued leadership on the issue. Further comments were:
- A GA4GH task team is developing text on the biomedical research implications of the GDPR which will be complete by the end of October.
- It is unclear whether or not the GDPR uses terms consistently across the regulation, which will generate problems for implementation. Interpretations (for example of ‘pseudonymisation’ vary across Europe and the Regulation does not help clarify or build on previously existing workable definitions.
- The GDPR will, however, create opportunities to create more consistency between the Data Protection Act and common law in research (e.g., s.251 exemptions), which would clarify and simplify the legal landscape for researchers.

6.0 Sanctions for data misuse
6.1 NB introduced the evidence paper on sanctions. This paper updated the initial draft presented at the previous EAGDA meeting, incorporating an extensive summary of the legislative and regulatory framework around sanctions for personal data misuse, and an analysis of the differences in behaviours that funders or other bodies might seek to punish – both instances of data breaches and ‘procedural’ breaches that indicate poor practice or a failure to adhere to established principles or protocols for data handling.

6.2 EAGDA discussed the report and considered whether and how to take it forward. It is striking that despite the existence of legal safeguards, there is a lack of consistency in approach across funders and other organisations with authority
over research behaviour.

- The majority of accidents or security incidents involve human error and would not be adequately addressed by imposing sanctions on researchers or institutions.
- Funders could consider an approach such as that adopted by the Health and Safety Executive, which has extensive guidance to make clear where obligations and responsibilities for health and safety lie.
- If sanctions were couched within a clear, broad framework of regulation that covers behaviour right across the spectrum from malicious breaches and misuse through to good practice, in a supportive environment that aims towards improving behaviour, trustworthiness in the system could be improved.
- The majority of violations of data management are not malicious, they are due to failures in proper curation and maintenance over time. This is where actions from the funders could be best directed: to push researchers to process data properly and safely.
- The current situation is uneven across the sector, with punishments looking disproportionate if they are for procedural failures that do not result in harms. Disproportionate or heavy handed responses will not help encourage a culture of good practice or allow for the possibility of learning over time.
- Proportionality is key, both for the nature of sanctions and also the manner in which data misuse could be policed. The DPA does mitigate against criminal activity with data (and the NDG recently recommended the introduction of criminal sanctions for misuse of patient data), but the monitoring of bad practice would take significant resource.
- Failure to enable access to data where appropriate should also be considered bad practice.
- The research sector does have a strong reputation in protecting and correctly handling research participant data, but it is almost inevitable that a breach will arise at some point. The sector and funders in particular should be proactive in this space and embrace the opportunity to develop good research practice for data use alongside sanctions for misuse.
- Under the GDPR, signing up to an approved code of conduct could enable a consortium to be considered ‘adequate’ to allow data processing. Sanctions could be incorporated into such a code.
- ME informed EAGDA of the SURE (safe use of research environments) consortium, which is developing training to deliver the ONS approved researcher accreditation. The next stage of this is to develop a ‘researcher passport’ that could transfer accredited status across different contexts – with the implication that sanctions could also be portable (i.e. if a researcher misuses data in one context, this could be flagged when they seek to apply for data in a different context).

6.2.1 It was agreed that there is a demand to produce recommendations on this issue, covering the spectrum of good practice on data, training and data conduct as well sanctions, as an output that is for the broader community and not only funders. This could outline principles that could form the basis of a code of conduct for the research community, across disciplines.

6.5 **ACTION:** Secretariat to separate paper into standalone evidence paper and a short agenda-setting piece on where sanctions could/should fit in range of tools for funders. Latter piece prepared for next meeting.

7.0 **PACE tribunal**

7.1 Frances Rawle (MRC) outlines the implications and remaining questions for
governance of research data following the PACE tribunal arising from an FoI request to access individual-level research data. The Medical Schools Council is seeking legal advice on behalf of the sector to consider questions that weren’t addressed in the tribunal ruling (e.g. terms of patient consent).

7.2 EAGDA discussed the tribunal ruling and considered broader implications for research governance, participant trust and data access.
- Data requested under FoI has no conditions attached and has to be treated as public domain once released. This bypasses all governance controls on research data and it would be concerning if this ruling established the FoI route as a legitimate means to access research data.
- The issue may have been avoided if data were hosted in a repository such as the UKDS with transparent access conditions. Concerns about onward use, dissemination, risk of re-identification etc could be managed better this way, avoiding breakdowns of trust in the process.
- The case highlights the need to make the case for enabling access to data from research in a proportionate, timely manner, which would allow reanalyses and mitigate the risk of data requestors resorting to FoIA for access. This would allow more expert, nuanced conversations about re-identification risk to be had.

8.0 ‘DataTerms’ project
8.1 David Townend updated EAGDA on progress with the commissioned research examining how concepts relating to ownership and property are used in practice in relation to data access and governance.

8.2 EAGDA discussed the project, its methodology and its aims:
- The project coheres well and looks as though it will provide a very useful insight into research in practice and how legal terms are being used.
- It may be valuable to try and get the clinical community’s perspective in addition: they will often be reluctant to allow data sharing as they consider data ‘belongs’ to patients.
- It will be important to identify the right ‘ecosystems’ of data producers, managers and users to target for the questionnaire. Funders should help with this process.

8.3 ACTION: Funders to work with DataTerms team to identify additional constituencies to invite for the survey phase. Team to report back at next EAGDA meeting in April 2017.

9.0 Data sharing and management plans
9.1 NB introduced the recommendations paper which outlines two types of recommendations to funders: specific actions should be taken jointly, irrespective of their current processes and internal committee and staffing set-ups; and further issues that they should consider and determine precisely how to take forward and resolve.

9.2 EAGDA discussed the paper and raised the following points:
- As they stand the recommendations are a missed opportunity to push funders to work in a more joined up, coherent way together. Conversations about data planning and management are still happening in disciplinary silos and the paper should strongly promote cross-funder working on this issue.
- The paper should draw on the provisions in the Protection of Freedoms Act
(2012)\(^1\) that articulates the rationale for datasets being held by a public authority to be made publicly available. This establishes the mandate to enable secondary use of data being generated by academic researchers within a university and a timetable for when it should be released/made available where appropriate.

- The paper needs to articulate that expectations should be managed across borders. The UK has rich data resources but cohort data management does need to be carefully managed. This approach may not always be understood and funders should make clear what they consider to be good practice.
- The paper is an opportunity to set out the case for generating enthusiasm about data science and data sharing, and the impact they could have on real world health problems and science challenges (e.g. data sharing in public health emergencies). The potential to capitalise on the UK’s data resources should be seen as an opportunity to be proactive, innovative and excited about the science that could result from using these resources more productively.
- The notion of reciprocity should be built into the recommendations on data plans: researchers should be able to mandate secondary data users to return derived, cleaned and/or enriched variables to the original resource.
- It was agreed that data plans should be dynamic and evolve over the course of a grant, acting as a resource for potential data users. They should be publicly available so that clear expectations can be set about when data will be released/accessible. This may also address some of the resource concerns about funders monitoring adherence to plans.
- In the long term, data plans should be machine readable: NIH are undertaking work on this at the moment.

9.3 **ACTION:** Paper to be amended in light of EAGDA discussion. Final draft to be circulated to EAGDA for sign-off before end 2016, for publication early 2017.

10.0 **DeepMind/Royal Free collaboration**

10.1 Tim Hubbard introduced the issue that received media attention earlier this year, when a clinical and research collaboration between the Royal Free NHS Trust and Google DeepMind was publicised. The agreement came about as the hospital IT infrastructure did not allow for adequate linking of patient data and systems to quickly and efficiently implement the algorithm for Acute Kidney Injury.

The case created questions about: the role of the private sector as a contracted third party in care delivery; the meaning of ‘direct care’; fair processing requirements under the Data Protection Act; ethical and governance oversight of data access within and beyond the NHS; and the balance of authority between local and national-level decision-making on levels of governance and control.

10.2 EAGDA discussed the case and made several comments:

- This raises the issue of the extent to which third party contractors will be are being used to host very rich, identifiable data, outside the NHS, to allow processing for care delivery. While the data controller (the hospital) is entitled to do this, there are significant consequences for public confidence if anything goes wrong with this approach, which extend across the whole health and care system including research.
- The distribution of data requires certain governance mechanisms and it is unclear whether local NHS Trusts are adequately equipped to robustly ensure

these are in place, with sufficient ethical and governance oversight.

- In contrast with these local, ad hoc arrangements, researchers go through an extensive, rigorous set of checks and balances in order to access data. There could be knock-on effects on research of poorly managed commercial contracts that damage public confidence.
- Serious consideration needs to be given to how the NHS, and researchers, engage with industry as data-driven science becomes more embedded in health care and research.
- EAGDA discussed the idea of writing a letter to the funders to express concern about this case and to highlight the potential knock on effects on trust and confidence in research using patient data. There is a need to be responsive to the increasing role of the commercial sector in this kind of research. However, it was agreed that as there was not a specific ask or recommendation arising from this issue it would not be an effective action to take forward.

11.0 Work plan

11.1 EAGDA discussed the current work plan and the key themes that had emerged during the meeting (breaking siloes; what limits innovation; consent terms; commercial involvement in research; dissemination and research outputs). It was agreed that the next meeting should provide for a more broad-ranging, open discussion about what the key issues for the group should be.

It was agreed that the Secretariat will send out question prompts well in advance of the next meeting for all members to consider regarding what EAGDA should work on, and how it should do this. This will include some reflection on how the landscape of data access and management has changed since EAGDA was set up, and where EAGDA should it in this landscape.

11.2 **ACTION:** Secretariat to circulate key questions to EAGDA for consideration in advance of next meeting (by end March 2017).

10.0 AOB

10.1 The EAGDA funding term will expire in September 2017. The Secretariat will undertake a review of EAGDA with input from members, funders and a broader group of stakeholders, to decide on future funding by June 2017.

10.2 The date of the next meeting was confirmed as Tuesday 18 April 2017. However, as this falls within Easter week the Secretariat will seek to identify if a different date is preferable and inform the group as needed.

10.3 There being no other business the meeting was closed.