

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

10:30-15:30 TUESDAY 26 APRIL 2016, WELLCOME TRUST

Present:

James Banks (chair)
Paul Burton
Mark Elliot
Mark Guyer (via telephone 11:30-12:30)
Paul Flicek
John Hobcraft
Tim Hubbard
Peter Knight (from 12:30)
Bartha Maria Knoppers
Andrew Morris
Onora O'Neill (until 1:00pm)

Funders:

Fiona Reddington (CRUK)
Jamie Enoch (CRUK)
Rebecca Fairbairn (ESRC)
Clare Feary (ESRC)
Geraldine Clement-Stoneham (MRC)
Katherine Littler (Wellcome Trust)
Natalie Banner (Wellcome Trust)
Dave Carr (Wellcome Trust)

Apologies for absence:

Melanie Wright
Mark McCarthy
Ros Eeles

Also attending:

Jennifer O'Callaghan (Wellcome Trust) observing 10:30-12:30
David Townend (on behalf of 'Data Terms' team) 11:30-13:30
Katarina
Sophie Laurie (NERC) (via telephone) 13:30-13:50

1.0 Welcome

1.1 JB welcomed the group and relayed apologies for absence.

1.2 The minutes of the eighth meeting of EAGDA, dated 30 September 2015, were ratified.

2.0 Matters arising

2.1 Item 1.6: EAGDA agreed to draft an editorial for the *International Journal of Epidemiology* on the Governance of Data Access report, highlighting the key recommendations and considerations for the international epidemiology community. This has been drafted with the help of PB and is currently with JB for editing. A full draft will be circulated to EAGDA for comment in due course.

2.2 Item 2.8: A query was raised by ME over whether a formal assessment of disclosure risk had been undertaken on the Ebola clinical trial data for which a data sharing platform is being set up and co-ordinated by the Wellcome Trust. KL agreed to follow up with Prof. Peter Horby to ascertain this.

3.0 Funders' updates

3.1 Wellcome Trust:

- 3.1.1
- NB reported on the commercial access to health data project. The report was released in March, with a panel discussion (approx. 70 participants) held last week. The report has been well received as an important contribution to discussions about the appropriate use and governance of patient data. There are tentative plans to help develop best practice in ensuring clarity of language about anonymisation.

EAGDA commended the report and highlighted:

- The importance of getting away from terms such as anonymised and pseudonymised, both for the public and for researchers;
- The need to ensure conversations do not occur in silos: many parallel discussions are happening with regard to data in the fields of security and intelligence, finance, etc. There should be a link between these and the ethical and privacy concerns about personal data arising in the biomedical and social sciences, and between the underlying legal frameworks;
- The Royal Society science policy group and the Royal Statistical Society are potentially important conveners on this issue.

It was agreed that EAGDA should keep a watching brief on this issue and that there may be a role for EAGDA in emphasising to funders the need for a holistic response or the brokering of wider relationships as these conversations develop.

- 3.1.2
- DC reported on Wellcome's recent plans to bring together its activities on open science and data, building on the previous work of establishing open access models to incorporate the sharing of research data and other outputs. DC is leading on this with others and the team is at an early stage of developing a roadmap over the next nine months.
 - DC also reported on the Open Science Prize: a collaborative initiative with the

National Institutes of Health and the Howard Hughes Medical Institute, to advocate and promote the innovative development of data sharing tools and platforms. Six winners will be awarded money to develop their prototypes; these will be announced soon.

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

- 3.1.3
- KL reported on Wellcome's recent work to support data sharing in global public health emergencies such as the Zika outbreak in Latin America. There is considerable overlap with EAGDA's work but with an international scope, and several of the challenges and issues emerging are familiar in the UK context (e.g. principles for access agreements; benefit sharing; data ownership; rights of access). Wellcome is developing a work plan to take forward the joint statement on Zika data sharing, to ensure high quality data can be shared quickly. It will look at shifting researcher expectations, incentives and rewards, infrastructure and ethical, legal and governance issues.

EAGDA welcomed this work and asked to be kept updated with developments.

- 3.1.4 **ACTION: NB to circulate slide decks from commercial access report to EAGDA.**

3.2 Cancer Research UK:

- 3.2.1
- FR reported on a recent cancer plenary session at which 10 large scale projects on somatic variation, led by the AACR flagship project GENIE, agreed to develop a harmonised approach to their data standards. It was noted that EAGDA was perceived as a useful point of interaction for the international community and there is enthusiasm from international collaborators to engage with the group.

EAGDA welcomed this news and considered that further international engagement would be important for the group.

- 3.2.2
- FR discussed planned engagement with the Farr Institute and also indicated that contact between the Turing Institute and the biomedical and health research field would be helpful as they consider issues concerning 'big data'.
 - JE informed EAGDA of a big data analytics conference to be held on 29 June, at which OO'N will be speaking. Issues to be addressed will include the ethics of data sharing and data linkage. The conference aims to convene stakeholders from across the health sciences and other sectors such as business and technology to discuss some of the common ethical challenges posed by big data.
- 3.2.3
- JE reported on CRUK's progress with their in-house data sharing processes. CRUK are working towards collecting more thorough data sharing plans from applicants, requested once an award has been confirmed but before funding commences. Grantees have so far been receptive to these changes and have been thinking creatively about how to make data accessible while handling restrictions due to confidentiality and IP. It is clear that studies often collect several types of data and that a tick-box approach to plans would not be feasible owing to the different approaches that might be needed of different

data types.

EAGDA commended CRUK on this excellent progress and requested to be kept up to date with the rollout of the new approach, including any difficulties experienced by CRUK or reported by their researchers.

- 3.2.4
- JE also informed EAGDA about CRUK's joint work with MacMillan on the impact of the implementation of Type 2 objections (patients objecting to information about them being shared beyond HSCIC) on cancer registration datasets. This review will report in three months' time.

3.2.5 **ACTION: NB to circulate link to big data conference with the minutes.**

3.3 Medical Research Council:

- 3.3.1
- GCS reported on progress with the draft concordat on open data, which follows from the Royal Society report on 'Science as an Open Enterprise'. There has been discussion in the drafting process about use of the term 'restriction' of data rather than 'management' or 'control' when describing data that may not be made openly available.

- 3.3.2
- GCS also reported on an incident concerning clinical trial data sharing: there is an appeal with the ICO at present about the release of clinical trial data that was requested under the Freedom of Information Act. The university concerned is appealing against the demand to release the data.

3.3.3 **ACTION: GCS to circulate to EAGDA the ICO ruling when released and also the final version of the Concordat when available.**

3.4 Economic and Social Research Council:

- 3.4.1
- RF updated EAGDA on the ESRC's delivery plan, to be released this week. It focuses on positioning ESRC's ambitions to be world-leading in the development of data infrastructure, exploring new forms and linkages of data and exploiting existing investments.

- 3.4.2
- RF flagged to EAGDA the recent Cabinet Office consultation on 'Better Use of Data' which set the groundwork for legislation to improve data sharing between government departments and also enable government data to be used for research purposes. ESRC responded and Vanessa Cuthill is also on their quality assurance committee to look at how the responses are handled.

- 3.4.3
- ESRC are holding an event with the Nuffield Foundation and another with CLOSER and the Royal Society in July, about the potential of linking administrative data with survey data.

- 3.4.4
- ESRC are undertaking reviews of a series of major data investments this year, including: UKDA, ADRN, big data centres, CLOSER and the longitudinal studies. A review of the lessons learned from the Life Study project is also

available on the ESRC website.

3.4.5 **ACTION: NB to circulate details of administrative data event with the minutes.**

4.0 EAGDA member updates

4.1 Paul Flicek:

- Tom Hudson, a member of the Global Alliance for Genomics in Health (GA4GH) Steering Committee, is leaving to take up a role in industry.
- A collaboration between EBI, Elixir and the Sanger Institute has been awarded a substantial grant for the GA4GH by the Wellcome Trust.

4.2 Bartha Maria Knoppers:

- The GA4GH has developed a data sharing lexicon to encourage shared terminology internationally for creating standard data access agreements.
- A paper has also recently been published¹ in PLoS Genetics on consent codes, to allow clarity across studies on what can and cannot be done with data.
- Work is also underway in conjunction with Elixir to develop a 'control-light' category of data access for data intensive science that doesn't involve confidential data, involving registered access and a single point of authentication.

4.3 Andrew Morris:

- Scotland has developed a system of accreditation for 'safe haven' research environments. Five are currently going through the process.
- Since January, there has been a single point for the Caldicott sign off process, based on Graeme Laurie's proportionate governance model. This should make the approvals process less complex and reduce delays.
- Work is also being undertaken on data interoperability in Scotland, involving data standards, shared authentication, management tools and a common costing model.

4.4 Mark Elliot:

- The ADRN has achieved harmonisation of training across seven data services, for researcher certification. It is aiming to create a form of researcher 'passport' as an accreditation to use data across different services.
- The UK Anonymisation Framework received a large volume of comments and feedback, and has been substantially revised in light of these. The new framework should be published in the next 4-6 weeks.

4.5 Peter Knight: (update provided later in the day)

- HSCIC will change its name to NHS Digital in July. The HSCIC has recently undertaken a lot of work to improve its relationship with the research community.
- A new research advisory board is being created with input from senior researchers and funders, to improve HSCIC's processes and access issues and to provide advice on emerging issues.
- Type 2 opt outs are being implemented from 29 April. It is anticipated that the

¹ <http://dx.doi.org/10.1371/journal.pgen.1005772>

Caldicott review will seek to move away from opt-outs based on information flows and more towards purposes of use.

- The Office for Strategic Coordination of Health Research is bringing together a group, under the leadership of Alex Markham, to look at the bioinformatics environment in the UK and strategic ways to improve data access and use in this field.

4.6 **ACTION: ME to circulate the framework when available.**

ACTION: NB to look into updating the EAGDA website to provide links to this and other resources as appropriate, when new Wellcome website is launched.

5.0 **Legal control and responsibility for data**

- 5.1 David Townsend gave a presentation on the plans of the 'DataTerms' team to undertake the work that EAGDA commissioned late last year, about the way the terminology of ownership and related terms impacts on data sharing practices internationally.

The team will primarily focus on the legal analysis of terms relating to ownership or rights over data in different jurisdictions. As part of this work they will also seek to ascertain how stakeholders use these in practice, especially when negotiating data transfer or sharing internationally. The team hypothesise that overlapping contexts generate 'myths' about what the law does and doesn't say with regard to issues such as data protection and the rights to access or restrict data.

- 5.2 EAGDA members highlighted the value of focusing on the legal analysis phase of the research, in part because this would be useful in its own right and in part because they raised concerns about the ambition of the project given the primarily legal expertise of the team and the timescales involved.

KL clarified that the Secretariat had requested that the project include the applied, practical aspect of how stakeholders perceive and use terminology, as this is of considerable value to the funders as they look to develop policy and shape best practice in data sharing. It has been noted that practical issues of ownership and participant rights with regard to data are coming to the fore especially in LMIC contexts at the moment and practical analysis of how terms are used would be valuable. The team has been structured to include social science expertise to help with this aspect of the project. However, KL conceded that the social science aspects of the project could be refocused and that the team should be allowed more time to achieve this aspect of the project.

DT emphasised that the terminological problem was the central component of the project: legal analysis is part of this, but it is the practice on the ground that generates problems, especially when there are no legal frameworks in place. Understanding these challenges does require additional exploration of how terms are used in practice, beyond a legal analysis.

- 5.3 It was noted that data access committees need to be included in the stakeholder groups. Also participants should be considered, and 'researchers' as a category comprise those generating, managing and using data, so may form distinct groups.

- 5.4 EAGDA concluded that there would be benefits from gaining clarity over the legal terms and that as a first step, this in itself would be a coherent, insightful contribution to

the literature. Exploration of stakeholder opinions and perspectives, could form a second phase of the research. This could subsequently be expanded into further jurisdictions if the funders agree this would be beneficial at a later stage.

5.5 **ACTION: Secretariat to take discussions forward with DT and team about the project plans.**

UPDATE SEPTEMBER 2016: Following the EAGDA meeting, the Secretariat discussed the proposal and EAGDA's feedback extensively with the project team. EAGDA members BMK, OO'N, PB and JB were consulted on the revisions and the following plan was agreed:

- Project to include literature review of existing legal terms relating to 'ownership' and 'property' as they relate to data in different jurisdictions as a first stage
- Survey and focus group design to be scaled back to three jurisdictions and informed by social science expertise.
- Survey to focus on consulting a small group of stakeholders who are known to be considering this issue, to identify how property/ownership concepts are used in practice (not law).
- Range of interpretations that are in play with different stakeholders to be mapped out, together with the practical fixes that they have produced to get around terminological differences.
- *Subject to later consideration*: potentially scale up social science methodology to other jurisdictions, to explore the impact of colloquial uses of property/ownership terms on international data sharing.

6.0 **Sanctions for data misuse**

6.1 NB introduced the draft overview of sanctions for data misuse that had been identified through desk research for a range of organisations, including funders, repositories, publishers, regulatory bodies and research management services. Key points from the typology developed were that:

- Data misuse is often subsumed into broader policies concerning research integrity.
- There are few examples of transparent processes for monitoring data misuse and investigating breaches.
- There is little evidence of co-ordination on accountability for data misuse between different stakeholders in the research process.

6.2 EAGDA discussed the report and considered whether and how to take it forward.

- Misuse can be accidental or deliberate. We should encourage a culture of reporting accidental misuse or near misses so that lessons can be learned quickly. GeL is adopting this 'health and safety' approach to data use.
- The background paper mirrors quite closely the findings that led to the development of the GA4GH accountability policy.
- Often the main source of risk is data users seeking to get around policies and processes to make their research quicker and easier: no security breach necessarily occurs here but rather a breach of process that carries with it a risk. This reflects a broader issue of accountability that extends beyond cases of misuse per se.
- A side benefit of implementing an accreditation process such as that developed by ADRN would be that sanctions/penalties could be applied across different data services. This would bolster the power of sanctions as a deterrent.
- The Royal Society's work on machine learning is also relevant, as data

breaches could be committed by machines and it will be important to consider how this would fit into any sanctions regime.

- There is a sliding scale between research misconduct broadly construed and data misuse, but often little confidence in the systems in place for investigating misconduct (because of the inherent conflict of interest of the institution undertaking the investigation being invested in the research).
 - Funders have signed up to the Concordat on Research Integrity, but this does not discuss implementation of policy or processes for dealing with instances of data misuse.
- 6.3
- The paper requires additional context, including the legal framework for data protection, freedom of information and the common law of confidentiality.
 - A common framework for sanctions would be helpful, particularly in light of the need to implement the EU Data Protection Regulation.
 - It would be helpful to set out a code of conduct from the funders, identifying what they will and won't do when a data breach is suspected. This is a policy decision that is complementary to but separate from the underlying legal framework.
- 6.4
- EAGDA agreed that the paper could be developed, in two-stages:
- Short-term paper setting out the current legal, policy and institutional landscape on accountability and sanctions for data misuse, expanding on the raw data from the current paper.
 - Broader longer-term paper identifying elements or principles of a common framework for codes of conduct for data use, with consideration of how to engage on this issue across funders, repositories, institutions and other parts of the scientific community.
- 6.5
- ACTION: Secretariat to take these suggestions forward and update EAGDA with a specific plan of action at the next meeting. If appropriate, a draft short-term paper to be ready for the next EAGDA meeting.**
- 7.0
- Data sharing at the Natural Environment Research Council (NERC)**
- 7.1
- Sophie Laurie, Head of Innovation at NERC, gave a presentation via teleconference on NERC's recent and planned activities to build data infrastructure and promote data sharing. She noted:
- NERC data centres have a broad UK and global coverage, with a huge and comprehensive range of environmental datasets. Datasets are not always interoperable, however.
 - Data centres have tended to be supply driven by demand led; NERC are working to change this to ensure data is made more accessible to users.
 - NERC has a renewed emphasis on innovation and trying to maximise value from data. An enhanced innovation programme is planned, with a programme board comprising data owners and a range of companies.
 - Research with oil and gas data can be controversial and sensitive, so work is being done to set out clearly the partners NERC works with and what their expectations of conduct are.
 - NERC are keen not to operate in silos or duplicate work that is being undertaken across other areas. Linkage between environmental and health data is one important area of cross-disciplinary work needed.
- 7.2
- EAGDA passed on thanks to Sophie for the presentation and considered how the work of NERC in a different field of data science compared to and reflected the biomedical

and social science fields. EAGDA members commented:

- In terms of technical challenges, the environmental data science field faces similar issues to the biomedical field: petabytes of data to handle; lack of consistent standards; access governance; the practicalities of linkage.
- It is noteworthy that NERC have a more of a business focus than the biomedical and social science fields, with commercial involvement considered as standard practice.
- Natural sciences don't have the same ethical issues associated with human subject research as biomedicine and the social sciences – but geospatial data can be disclosive if linked to other data. Cross-disciplinary infrastructure discussions are therefore needed about how to manage these disclosure risks when linking across very different types of data, e.g. through managed access procedures.
- It is evident that thinking about data management and infrastructure is still occurring in disciplinary silos, even though many of the underlying challenges and questions are the same across scientific areas.
- Scientists working in vastly different disciplines do not necessarily understand the data generated in other fields and the associated ethical, technical and conceptual difficulties of linking the data.
- Some small scale collaborations (e.g. in Scotland) are beginning to bring these diverse academic communities together around particular science-led issues.

8.0 Data sharing and management plans

8.1 NB introduced the draft overview of the circulated paper on how funders and their committees consider data sharing and management plans. Key points were:

- There is little consistency either within or between funders about their approaches to evaluating data sharing and management plans.
- Some committee members expressed reservations about the funders' interest in data plans, preferring to focus their committee discussions solely on the quality of science.
- There is significant variation across and within disciplines about the value of making data available for secondary use.
- Overall, committee members would value a steer from funders about whether and how they should take data plans into account in their decision-making.

8.2 EAGDA discussed the paper and the issues it raised:

- The need to promote good quality science shouldn't undermine the necessity of thinking about data sharing and management in a grant application, whatever the branch of science.
- Planning for data management shouldn't require every project to have a specialist data scientist involved, but considering what to do with the data generated should be an integral part of a researcher's plans.
- Many researchers do not support the principle of data sharing, but will shift in practice over time if funders work to shape the policy landscape and actively support data sharing activities.
- It is recognised that planning for the costs of data access and management can be difficult at the outset.
- Data sharing considerations should not be separate from broader discussions about the value and quality of science. This is reminiscent of conversations about 'impact' a decade ago.
- There is little consideration by the funders of monitoring and checking compliance to data policies.

- Several ideas in the paper could be developed: e.g. a data ombudsman for data users to refer to in instances of data hoarding or unfair access restrictions; making data management and sharing plans public,
- It is unlikely that a one-size-fits-all model for funders' data sharing plan requirements would be appropriate. There will necessarily be significant variation in what kind of plan is appropriate for different data types, and for different type of grants. Nevertheless, there could still be a generic set of considerations that would apply across projects and which would help funders in assessing plans on a case by case basis.
- The Sanger Institute created its own data sharing committee to provide specific expertise on data sharing to aid those committees evaluating a proposal, starting from the default view that most data should be accessible unless there were good reasons for it not to be. Funders could consider a similar body or panel to assess projects' plans and check implementation.
- Funders could consider whether this is an important enough issue to be an institutional responsibility rather than expecting the PI to develop and sustain all aspects of the data management and sharing. It is important to be clear on the locus of responsibility.
- Data management plans could be made public and researcher held to account for delivering on these.

8. EAGDA discussed ways forward for this paper:

- CRUK's approach of asking applicants to develop more detailed plans once a decision has been made to fund an award was considered appropriate.
- It is clear no one size fits all for data management plans, but there are some mechanisms that could be common to all funders: how and when plans are taken into account; how costs are assessed; what responsibility lies with the institution and what with the PI; how the implementation of plans can be monitored.
- From the suggestions made in the paper and ensuing discussion, it was agreed there is scope to map out the landscape and develop a series of recommendations to funders.

8. **ACTION: MRC to supply data from a sweep of their funding committees' views to feed in to the paper.**

ACTION: Secretariat to develop paper into a series of concrete recommendations for funders, for circulation and comment.

9.0 Data infrastructures meeting

9.1 DC introduced the outline proposals for the data infrastructures roundtable meeting, which is being planned following the recommendations of the *Incentives* report. The report stated the need to lower the barriers for data-sharing, in part through linking up expertise in existing major repositories and developing best practice for repositories across disciplines.

The plan is for a day long roundtable discussion for representatives from UK repositories to discuss the challenges for data infrastructures of managing data from cohort and longitudinal studies.

9.2 EAGDA welcomed the plans and provided the following comments:

- Experienced data users and major study representatives need to be included in the participant list.
- Institutions with big data activities should also be considered.
- It will be particularly useful to discuss how to coordinate infrastructures particularly to leverage cross-disciplinary research potential.
- The meeting will be timely but must have a clear, tight focus
- Clarity over 'whose meeting this is' will be important
- Central question for the meeting could be "Why is data being underused and how can data infrastructures help?"
- If there are several key issues that cannot all be addressed, the meeting could be treated as the first of a series, to build up collaborative relationships and networks on these issues.

Several suggestions for additional stakeholders to be invited were suggested.

9.3 **ACTION: EAGDA members to pass suggestions for additional invitees to Secretariat. Small working group including PF and MW to convene and plan meeting.**

10.0 AOB

10.1 EU Data Protection Regulation:

NB provided an update on behalf of Beth Thompson, the Wellcome lead on the EU General Data Protection Regulation:

- The final text was formally adopted by the Council of Ministers and European Parliament in April 2016. This text has been through legal linguistic checks and in the key research clauses there are no significant changes to the text informally agreed in December.
- The Regulation is expected to be published in the Official Journal of the EU in early June 2016. It will become effective two years after this, in 2018.
- The final outcome is a rejection of amendments proposed by the European Parliament that would have had a severely damaging impact on research.
- There are still challenges ahead in implementation, for example the Regulation will not lead to further harmonisation in the EU for research.
- In the UK Wellcome have had promising discussions with the Department for Culture, Media and Sport (DCMS). They want to continue to collaborate on research issues as they implement the Regulation in the UK.

PK reiterated that DCMS are keen to engage with the research community on the issues of implementation and want to adopt a pragmatic approach. It is unclear yet whether current data protection law will be revoked, amended or repealed. A range of issues need to be considered with regard to their impact on research, including: the 'right to be forgotten'; definitions of pseudonymised, anonymised and identifiable data; and sanctions for data breaches.

10.2 EAGDA noted its gratitude to Beth for her leadership on this issue. It was agreed that Beth should be invited to present an update on the implementation process at the next EAGDA meeting in October. EAGDA can then ascertain if it can play a useful role in the formative stages of implementation plans for the UK.

10.3 The date of the next meeting was confirmed as Tuesday 4 October 2016.

10.4 There being no other business the meeting was closed.