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

10:30-16:00 WEDNESDAY 30 SEPTEMBER 2015, WELLCOME TRUST

Present:

Martin Bobrow (chair)
Paul Burton
Ros Eeles (until 2pm)
Mark Elliot
Mark Guyer
Tim Hubbard
Bartha Maria Knoppers
Andrew Morris (until 2pm)
Melanie Wright

Via teleconference:
Mark McCarthy (until 1pm)
James Banks

Attending for morning session:
Beth Thompson (Wellcome Trust)
Niklas Blomberg (Elixir)

Funders:
Fiona Reddington (CRUK)
Jamie Enoch (CRUK)
Rebecca Fairbairn (ESRC)
Clare Feary (ESRC)
Geraldine Clement-Stoneham (MRC)
Jon Fistein (MRC) (from 1pm)
Katherine Littler (Wellcome Trust)
Natalie Banner (Wellcome Trust)

Apologies for absence:
Paul Flicek
John Hobcraft
Peter Knight
Onora O’Neill
1.0 Welcome

1.1 MB introduced two new funder members to the group, Jamie Enoch (CRUK) and Rebecca Fairbairn (ESRC).

1.2 The minutes of the seventh meeting of EAGDA, dated 20 April 2015, were ratified and no matters were arising.

1.3 The dates for the 2016 meetings were confirmed as: **Tuesday 26 April** and **Tuesday 4 October**.

1.4 MB announced that he would be stepping down as chair of EAGDA after this meeting, and that James Banks had agreed to take up the role of chairman. On behalf of the group, he was thanked for his leadership of EAGDA over the past three years.

1.5 The Governance of Data Access report was discussed as one of the key recent outputs from the group. Aside from the ‘World View’ piece in Nature which accompanied its launch, *Research Fortnight* have expressed an interest in the report and a draft article has been submitted to the journal outlining its key recommendations – awaiting a response from the editor. It was suggested that further dissemination to a wider audience would be beneficial.

1.6 **ACTION:** PB to explore possibility of submitting Governance of Data Access report to the *International Journal of Epidemiology* for publication.

2.0 Funder updates

2.1 Wellcome Trust

- The project on exploring **public attitudes to commercial access to data** is underway, with the first workshops occurring in September. Ipsos MORI have been commissioned to undertake the project, involving over 200 people including members of the public, patient group, cohort participants and health professionals. The report is due to be released in early 2016.
  - BMK noted that the OECD currently has a working group on the use of health data who would be interested in the project.

2.2

- In June, the Trust held a workshop between **HSCIC staff** and researchers using HES and other HSCIC data, in an attempt to help resolve some of the substantial difficulties the research community has faced in accessing data. Despite positive discussions and a commitment from HSCIC to improve the situation, many issues remain unresolved and there continue to be significant delays in what is a very complex and messy operational and policy landscape. This is having a detrimental effect on research: a concern echoed by other funders.
2.3 Restrictions and delays may reflect the fact that many HEIs do not currently have adequate provisions in place for data security, e.g., through the IG Toolkit. It was suggested that there was an onus on RCUK and other funders to push for the implementation of high security standards, but it should also be acknowledged that the political deadlock over patient data made progress difficult.

2.4 There are plans being discussed to set up a senior academic advisory group, potentially commissioned via OSCHR (Office for the Strategic Coordination of Health Research) to get an overview of the fragmented landscape at a high level and to fix the current system.

2.5 The ADRN wished to raise concerns that the HSCIC is interpreting legislation protecting health data in a very narrow way, focused on clinical care, and did not appear to recognise the need or value of access to data for crossover domains such as those involving the social sciences.

2.6 EAGDA as a group expressed concern about the protracted period through which research across a range of fields and disciplines has been impacted by the difficulties with data access. There is a strong need for a unitary framework for access to data created with buy-in from the highest levels of DH, that is functional, coherent and that avoids unnecessary delays.

2.7 • The Trust is currently working on an Ebola clinical trial data sharing initiative with Peter Horby’s group in Oxford. The initiative aims to make data gathered in clinical trials available for access via an independent access committee, constituted primarily of members from the countries affected. As the clinical trials are small scale, it was felt that the risk of identification of participants meant that access should be controlled. Ideally this platform will provide enable access to other kinds of data beyond clinical trials, but it is at present a first step in making data quickly accessible.

2.8 It was queried whether a formal assessment of re-identification had been undertaken on the clinical trial datasets, and ME offered support to help undertake such an assessment if required.

2.9 • David Carr sits on the RCUK working group for the Concordat on Open Data in research, which closed for consultation recently. The draft principles align well with the principles of good practice set out in EAGDA’s Governance of Data access report. The Concordat text should be finalised by the end of 2015.

2.10 EAGDA noted its approval of this progress.

2.11 CRUK • The Scientific Executive Board has approved plans to require all funded grants to provide data management and long term curation plans, which will be assessed prior to funding being released. For population health
grants, metrics around discoverability will be assessed annually, and on a
three yearly basis for basic and biological research grants.

2.12 • The stipulation that funds for data-sharing cannot be applied for has now
been removed from CRUK policy, in recognition of the dedicated time and
cost resources that are required to properly enable data access.

2.13 • CRUK researchers have developed a ‘workaround’ for the current
problems at HSCIC, with specific case officers from HSCIC working directly
with researchers and holding weekly teleconferences to assist with the
application process. This is a resource-heavy, non-scalable fix but it is
working temporarily whilst the broader policy and operational issues are
worked out.

2.14 • CRUK are in discussion with Cary Gross at the YODA project, who is
developing metrics for clinical trial data sharing.

2.15 • The National Cancer Institute in the US has developed a ‘team science’
toolkit. It includes draft letters for senior management on recognising the
skills and expertise required for team science, and metrics for
acknowledging and crediting collaborations and interdisciplinary working.

2.16 • Following the recent Farr Institute conference, a Frontiers meeting will be
held in December to increase the focus on cancer research in the Farr
network.

2.17 EAGDA applauded CRUK’s progress in driving forward the agenda of improving
data access among its research community.

2.18 EAGDA noted that there are substantial differences between different kinds of
research in terms of impacts and what counts as good data sharing, and these
must be recognised in any metrics that are developed. The social sciences cannot
be judged on the same criteria as the biological sciences and appropriate metrics
are essential.

2.19 Other funders expressed interest in CRUK’s funding approach to data
management and sharing plans: it is not clear how funding committees could be
guided with regard to judging what a good or appropriate data management plan
looks like in different circumstances. There might be a role for EAGDA in helping
develop some guidance (see 4.12)

2.20 ESRC • Funding calls for Centres for Doctoral Training and Doctoral Training
Programmes are now live. Two thematic CDTs are also being
commissioned, one "biosocial" and co-funded with BBSRC. This will
specifically focus on exploiting the breadth of biosocial data; developing
technological and methodological innovations; and exploring the legal and
ethical interfaces exposed by the biosocial agenda both within the UK and
internationally.
2.21 • ESRC are close to releasing a revised **Framework for Research Ethics**. Revisions include encouragement for researchers to think ethically and identify potential ethical issues throughout the lifecycle of a project. The Framework will include an ethics toolkit and illustrative case studies, covering issues such as building trust and protecting identities; data deposit; consent and anonymisation. ESRC noted that this move followed their commitment in the funders’ response to EAGDA on re-identification.

2.22 • The **Administrative Data Research Network** now has around 60 active projects, but they continue to suffer from a lack of clarity on the legal basis for data-sharing and resourcing issues at government departmental and agency level now appear to be a key component in slowing/blocking data release.
   
   o The ADRN is considering seconding staff to specific projects to support and provide data science expertise for researchers.
   
   o It can be very costly to create meaningful and useful variables from the vast datasets, so it is important that derived variables can be returned to the ADRN for use by others to avoid repeated costly analysis. MW acknowledged that the ADRN will need to address the issue of data retention and the return of variables once it is more fully established.

2.23 • The **Secondary Data Analysis Initiative** – which aims to ensure datasets are reused where possible – will become a rolling, open competition which will allow greater flexibility to target particular datasets, and provide more opportunity to leverage co-funding. The annual budget will be around £3m (6 or 7 proposals, 3 times a year), but this may rise with co-funded themed opportunities.

2.24 • Through 2015, ESRC have run **Public Dialogues** specifically focused on the re-use of private sector data for research and on work to enable access to data for research. The report will be published later in the autumn. Emerging findings include:
   
   o Participants understood the value of sharing data for research and public service improvement, and the need to balance individual privacy and social benefit, and were willing to trade off;
   
   o Participants were clear that publicly funded data infrastructure should be independent of both private sector and government interests, and that impartiality on its use for research should be protected;
   
   o Data storage and ownership were the hardest concepts for participants to grasp. Many believe they retain ownership even when there data is given to others. Digital storage is seen to be as (or more) safe as physical.

2.25 **MRC** update deferred to the afternoon session.
2.26 **ACTIONS:**
- **BMK** to provide Secretariat with contact details for OECD working group on uses of health data.
- **Secretariat** to draft a paragraph for EAGDA summarising their concerns about the current situation with HSCIC and access to data, to submit to the funders.
- **KL** to check whether a formal assessment of disclosure risk has been undertaken for the Ebola clinical trial data, and liaise with **ME** if assistance is required for this.

3.0 **Big data and international data access**

3.1 MG gave a presentation outlining the work of the Big Data to Knowledge (BD2K) initiative from the NIH:
- The initiative is focused on five action areas to establish a data science ecosystem across the NIH, namely: leadership, sustainability, policy and process, discovery and innovation, and workforce development and diversity.
- Central to BD2K is the concept of “The Commons”: a virtual shared space to find and catalogue the use of shared digital research objects.
- The NIH data sharing policy, initially developed for GWAS studies, will eventually be extended to be applicable to all data generated by NIH grants.
- The Scientific Data Council is a high-level oversight group for policy development and implementation that will be responsible for this. BD2K is a fixed term funding initiative and may in the long run be incorporated into the National Library of Medicine.
- Between BD2K and the Precision Medicine Initiative, the bridge between research and clinical care must be negotiated.

3.2 The group discussed the innovation of BD2K and queried whether, given its NIH focus, it was possible for it to allow an international focus, which would be important given the global nature of many research consortia. Mechanisms of equivalency for data policies will be needed so that different countries are not bound by the same rules but can nonetheless share data and adhere to common principles.

3.3 With regard to the prospects for international collaboration, if any positive solutions to the challenges of big data sharing across the UK/US and other countries are to be developed, funders need to recognise that resource will need to be devoted to these endeavours.

3.4 Interesting developments are occurring in this space which should be nurtured and shared: for example, creating tools to incentivise data-sharing, through block chain technologies and instruments of ledgering to measure attributions of effort and collaboration. EAGDA agreed that it could explore the development of these
technologies for data sharing in a future project.

3.5 Niklas Blomberg described the work of Elixir, a pan-European initiative to develop a sustainable infrastructure to drive data-sharing for biological research, involving academia and industry:

- Unlike some academic spheres, the life sciences are fragmented and widely distributed, which makes the sharing of large datasets difficult, particularly reference and aggregate datasets.
- There are substantial opportunities to incentivise and reward data sharing through systematic tracking of attribution and use of datasets, if data repositories and archives could be easily linked with publications and journals.
- Careful formatting and detailed metadata are essential to make data usable by others, not just discoverable and available. Co-ordination and supervision is required to enable data to travel, but there are issues of scalability to overcome if solutions are to be pan-European or global rather than workable only for a specific consortium or project.
- Elixir is working with the GA4GH Beacons project to enable data sharing between cohorts, with a pilot starting this month. Nordic countries tend to be forward-thinking about data sharing across their borders and among their cohorts, and have similar legislation. They are developing shared authentication and management tools for data sharing, with mutual recognition of credentials across universities.
- In Sweden, universities invest in human capital infrastructure, with dedicated personnel to support data management and sharing: grant holders (funded by central government) are entitled to two weeks’ worth of free data support to enable sharing. Access to data is based on a ‘Triple A’ approach of:
  - authentication (bona fides; track record of research integrity);
  - attestation (agreement regarding what researcher can and cannot do with data); and
  - authorisation (access enabled for a limited time period).
- The development of standards in academic data sharing could benefit from lessons in industry, e.g., the development of the GSM platform for mobile telecommunications (collaboration at interface; competition at implementation level). A rigorous approach to standardisation in health and biomedical research is needed, with visibility about the activities of different groups. Before standards can be developed, it needs to be clear what the questions are that standards are aiming to resolve. Standards could enable attributions of data usage to be more easily tracked.

3.6 Beth Thompson provided EAGDA with an update on the current status of negotiations on the EU Data Protection Regulation. The Council of Ministers agreed their position in June, which is quite positive for research. Although it does not harmonise data protection provisions across Europe, it delegates research provisions to member states, meaning that current research in different countries should not be damaged. Given Parliament’s position (which would damage
research), a maintenance of the status quo would be considered a success.

3.7 Beth noted that the ‘data saves lives’ campaign is launching a petition next week, supporting the use of data for health research purposes.

3.8 **ACTION:** Secretariat to distribute link to data saves lives petition along with the minutes.

4.0 **Work plan**

4.1 EAGDA discussed the work plan and suggestions for future areas to focus on, drawing on the morning’s discussion.

4.2 **Legal control and responsibility**
Following several meetings and circulation of the draft spec to three experts, the Secretariat considers that the legal control commissioned research should have its remit slightly expanded to cover some of the ethical and cultural barriers to sharing data across borders, and examination of how the law is used as a way to prevent data being shared appropriately.

- The Secretariat has also identified a potential group with legal and clinical expertise who would be in a position to bid for this work, thus indicating that the necessary expertise is feasible to bring together.
- The Wellcome Trust indicated it should be in a position to provide additional funding if the necessary to undertake the project with the right expertise on board.

4.3 EAGDA and the funders agreed with the approach. It was requested that the revised specification include the issue of liability for databases with regard to the clinical validity of data they hold.

4.4 **ACTION:** Secretariat to circulate a revised specification for the project once amended after the meeting.

4.5 **Sanctions**
EAGDA held discussion of sanctions over from the previous meeting. Receipt of the ADRN sanctions policy was gratefully acknowledged by the group. The funders consider that it would not be feasible to harmonise their respective policies on research integrity with regard to sanctions for deliberate misuse of data.

4.6 It was agreed that a set of high level, common principles could be shared by all, outlining the kinds of penalties that would result for individuals, grants and institutions who transgress in particular ways. Typologies of sanctions (contractual, employment, financial etc) by different bodies (funder, institution, journal) could be set out and commonalities found that could form the basis for a short document of key principle. This would allow differences in implementation and integration into different funder policies, whilst providing a clear, united message to the research community and the public.
It was noted that an organisation such as the ADRN would benefit from an agreement among funders that breaches to its data policy would attract sanctions from other funders, to back up its own policies.

Failure to enable data to be shared where appropriate is a separate issue and EAGDA agreed that it would not at present be appropriate to sanction researchers who did not adhere to data sharing policies.

EAGDA agreed that beyond research uses of data, as data becomes more accessible and storage options include cloud providers, sanctions for malicious breaches of data security should be called for. It was noted that in its response to the Science and Technology Committee’s inquiry into ‘The Big Data Dilemma’ the Wellcome Trust called for criminal sanctions for deliberate re-identification of individuals from big data.

**ACTION:** Secretariat to bring together a typology of available sanctions from funders, institutions and other bodies, to establish areas of commonality where harmonisation is feasible.

**Taking Data Seriously**
Following the update from CRUK, EAGDA discussed whether it would be feasible and desirable to develop guidance for funding committees and grant handler on how to assess data management and sharing plans, so that data sharing and access is taken seriously in funding decision. From the CRUK perspective, it is clear that over time the population health committee has developed its perspective on how research should be done, and that data sharing is an integral part of good research practice (rather than being a discrete activity).

EAGDA discussed whether it could devise practical guidance that could usefully cover a range of different disciplines and grant types. This could go together with recommendations for how funders could monitor whether commitments are being honoured and assess the impact of data sharing activities by their grant holders. The CRUK population health committee could be requested to provide a sense check on any documentation developed.

**ACTION:** Secretariat to draft initial outline of evidence document on how funding committees take data sharing into account, for circulation to the group.

**UK Biobank EGC Review**
The recent review of the UK Biobank Ethics and Governance Council posed a question that the MRC and Wellcome Trust wished to pose to EAGDA. The Review recommended that an overarching EGC for UK cohorts should be established, and it was proposed that EAGDA could potentially be in a position to undertake this role, if the expertise on the group were slightly adjusted. EAGDA discussed the suggestion. It was agreed that UK Biobank should be incorporated into broader discussions about UK cohorts, but unanimously agreed that EAGDA did not have the right constitution or remit to undertake an ethics and governance role for cohort studies. EAGDA could support recommendations from such a grouping there could be a useful overlap between them to ensure that issues could be fed between them for discussion.
4.15 **Other areas**
The work plan discussion emphasised EAGDA’s interest in identifying how any of the initiatives or recommendations it suggests do, if put into practice, have any impact on academic research practice and data access. Initiatives need to be evaluated and metrics of evaluation should be built in at the outset. However, it is not yet clear what metrics would be of value, as usage of data does not necessarily reflect scientific utility or value. In light of the postponed Hefce consultation on the future REF, EAGDA considered this an important topic to consider for discussion.

4.16 **Trustworthy Research Environments/MRC update**
JF introduced EAGDA to the work he has been undertaking on behalf of the Ministerial Industry Strategy Group in relation to the development of Trustworthy Research Environments (TREs):
- The governance landscape for health data is currently fragmented, and a principles-based approach should enable the problems that have been faced to be fixed for clinical care, academic research, industry and commissioning purposes.
- Both the research community and data providers need a clear, coherent framework and to learn lessons from the problems of the past two years. Trustworthiness of data access systems needs to be developed among different stakeholders, and this applies to all health care data (not only for research).
- A feedback loop of data is going to become more central to both health care and research, with them feeding into each other – it is therefore imperative that both the health care system and researchers using data subscribed to the same principles of data access and management. Scotland and Scandinavian countries have already adopted this approach, arguing that the efficient operation of the whole healthcare system requires data sharing, with research integral to this.

Subject to some clarifications and comments from individual members, EAGDA endorsed the principles-based approach outlined in the TREs paper circulated by JF to the funders.

4.17 **ACTIONS:** EAGDA members to respond to JF directly with comments and queries on the TREs paper.

5.0 **EAGDA member updates**

5.1 **UKAN Framework**
ME introduced the UK Anonymisation Network’s draft Anonymisation Framework, circulated to the group prior to the meeting. The network is a cross-sector, cross-disciplinary group that has sought a wide range of inputs to develop a plain English guide to anonymisation. It is intended as a training tool for thinking about different ‘data situations’, data transfers and circumstances in which anonymisation issues arise. The framework can also be used to help develop institutional policy. It will be published as an open access book and is being
reviewed by a range of people, including critics. Feedback from EAGDA as a ‘critical friend’ would be welcome.

- EAGDA congratulated ME on the development of a comprehensible guide to issues that are extremely complex.
- It would be useful if the terms used were concordant with those used elsewhere to ensure there is clarity in terminology, e.g., aligning with terms in the Global Alliance for Genomics and Health Privacy and Security Policy.

5.2 **ACTIONS:** ME to circulate feedback form to EAGDA members. Secretariat to explore possibility of including link to UKAN resource once published from the EAGDA webpage.

5.3 **Prostate cancer consortium**
Via the Secretariat, RE informed EAGDA that the prostate cancer consortium in which she works had successfully created a data sharing agreement with over 100 signatories, which is a significant development. The process and framework used is going to be written up into a paper with the intention that other groups could use it as a toolkit or guide for data sharing in international consortia.

- EAGDA welcomed this development and passed congratulations on to RE for this achievement.

6.0 **AOB**

6.1 JF drew EAGDA’s attention to the recently-announced role for Dame Fiona Caldicott in devising the form and wording of patient opt-outs from having data from their medical records used beyond their direct care.

CLOSE