The Academy of Medical Sciences

Response to the Royal Society consultation: Science as a public enterprise

August 2011

1. The Academy of Medical Sciences welcomes the opportunity to respond to the Royal Society's '*Call for evidence: science as a public enterprise'*.[1] The Academy promotes advances in medical science and campaigns to ensure these are translated into healthcare benefits for society. Our Fellowship includes leading medical scientists from hospitals, academia, industry and the public service.

2. The use of scientific information and how it should be managed to support innovative and productive research are important areas for debate. The Academy recognises the potential advantages of information sharing across all sectors of science. In medical science, the sharing of data (e.g. raw or partly processed data; clinical, statistical and patient data; summaries of research findings; and research protocols) can enable efficient research, accelerate scientific progress and build trust and transparency in the scientific enterprise. However, we recognise that there are challenges involved in sharing certain information and situations where sharing data or findings may be impractical, inappropriate or unethical. As well as focusing on the benefits and challenges of sharing data in relation to medical science, this response highlights the importance of sharing data across sectors and explores specific issues around access to patient data. We would be pleased to expand on any of the points raised in this submission.

Overview and principles

3. In this response we have tried to distinguish between the different types of scientific information in medical sciences, including:
   - **Raw or partially processed data**. This might include results from studies involving animals or humans, as well as biochemical research or findings generated in the process of developing new equipment (e.g. sensors). These data are most likely to be of interest to other scientists, regulators and some special interest groups. The use of patient data is essential to research that underpins our knowledge of disease, the development of diagnostic and therapeutic interventions, and the delivery of health services.[2] In our submission we highlight the particular issues around the provision and use of patient data.
   - **Research findings based on analysis or interpretation of data**. Ideally, these findings will be made openly available, following peer review, in a scientific publication with suitable mechanisms in place to communicate the implications to decision makers and the public.
   - **Information on research protocols**. Although not addressed in detail in this submission, sharing information on research methodology is vital to enable replication of an experiment, to allow analysis of the data by other researchers (a basic principle of the scientific method) and to understand the limitations of the findings. These protocols should be published alongside the research findings and data.

[1] Royal Society consultation (2011). *Call for evidence: science as a public enterprise.* http://royalsociety.org/policy/sape/

[2] Academy of Medical Sciences (2006). *Personal data for public good: using health information in medical research*. http://www.acmedsci.ac.uk/p99puid62.html

4. In the biomedical sciences, many national and international stakeholders have an interest in scientific information generated in the UK, as producers/providers, consumers or both. These stakeholders include organisations and individuals in academia, industry, medical research charities, the NHS, Government, regulators, media and publics (including special interest groups). New relationships between these stakeholders, which involve sharing of information (e.g. between industry, academia and the NHS, or between scientists and the public) provide opportunities and challenges that we outline below.

5. The Academy is committed to supporting medical science and ensuring that its findings are translated into health and wealth benefits for society. As outlined in this submission, effective and responsible sharing of scientific information can contribute to these goals by facilitating efficient and cost-effective research, aiding translation of scientific findings and engaging stakeholders in the scientific process. Because of these benefits, the Academy believes that, in principle, the findings of research, the protocols used and the data generated should be disseminated in an accessible form as quickly and as widely as possible. However, we recognise that there are exceptions, including obligations to protect the confidentiality of research participants, the high cost/benefit ratio of making certain data available in an appropriate format, the need to incentivise research by allowing a reasonable time for researchers and funders to exploit their research, and the potential for misuse of scientific information.

Benefits of data sharing

6. The benefits of sharing data can be categorised under three broad (and overlapping) headings: facilitating efficient research; aiding the translation of research and evidence-based policy making; and increasing trust and transparency.

Facilitating efficient research

7. Most scientific progress is incremental, where advances are made by testing and building on previous research. Access to the findings, protocols and sometimes data generated by previous research is vital to the rapid progress of science (and ultimately the translation of these findings into societal benefits). In an era of limited funds, it is particularly important that research resources are used effectively. Sharing findings and data across the public, private and charitable sectors can improve cost-effectiveness by avoiding unintended replication of research, optimising the utility of findings, and identifying new research avenues.
8. Genome-wide association (GWA) studies provide an example of one area where progress is dependent on responsible sharing of research data amongst researchers.[3] GWA studies are a powerful tool for deciphering the role of genetics in human biology and common disease. By analysing hundreds of thousands of genetic variants, and comparing individuals with a specific disease against carefully selected controls, the approach is already providing novel insights, with the promise of identifying new biological pathways and drug targets. Given the small effect sizes of the associated genetic variants, increasing statistical power through data sharing and meta-analysis of studies has been a major feature of progress. Success will be dependent on responsible data sharing amongst researchers and the development of effective and safe mechanisms to share genotype and phenotype data. The Wellcome Trust Case-Control Consortium (WTCCC) is an example of efforts to collect and share this information.[4]

[3] Academy of Medical Sciences (2009). *Genome-wide association studies: understanding the genetics of common disease.* http://www.acmedsci.ac.uk/p101.html
[4] The Wellcome Trust *(Case control consortium)* (2007). http://www.wtccc.org.uk/

9. There are often inherent delays in publishing findings and data in scientific journals, and we welcome steps taken by researchers and funders to publicise research being undertaken. It can be particularly important for researchers crossing disciplines or moving into new areas to be aware of ongoing research where the results may not yet be published. We also welcome initiatives to make publics and patients aware of ongoing research, such as the National Institute for Health Research (NIHR) Clinical Trials Gateway,[5] which provides information to help patients and clinicians locate relevant trials in which they may wish to become involved.

10. The ethical issues relating to data sharing vary with the types of data and its use. However, there is clearly an ethical case for maximising the use of data generated through research using animals (especially more sentient species) and studies involving humans (both patients and healthy volunteers), and preventing unnecessary duplication of such research.

Aiding the translation of research and evidence-based policy making

11. The availability of findings, protocols and sometimes data generated by research is vital to aid scientific progress and the translation of findings, and to ensure that public policy is informed by evidence.

12. The sharing of scientific findings and data can speed up the translation of research into health and wealth benefits. To capitalise on novel insights into the biology of disease a new process of drug development is necessary, which draws on strengths from across sectors and requires organisations to share data (and expertise, skills and resources). An increased focus on academic-industry collaborations is contributing to a new pre-competitive 'front end' to drug development, which is characterised by greater sharing of early-stage research data. Strong public-private partnerships founded on high quality, open and accessible data have the potential to benefit all groups engaged in drug discovery by increasing our understanding of biological processes and helping to decrease the attrition of clinical programmes.

13. The National Academies play an important role in promoting and facilitating engagement between scientists and policymakers, for example in providing an evidence base for public health initiatives. The global scientific community's response to the threat of pandemic flu is an example where medical scientists and policymakers have shared data to protect health. In our report on pandemic influenza, the Academy and Royal Society recommended that Government should lead efforts to coordinate data sharing in a pandemic.[6] The sharing of data can also aid capacity building efforts and address healthcare needs in lower and middle income countries. One example is the sequencing of tropical parasite genomes (such as *Plasmodium falciparum*, a causative agent of malaria). These data, whilst generated in the UK, have direct relevance to the work of researchers in Africa.[7]

Increasing trust and transparency

14. Science is increasingly part of public culture. As accountability and engagement with the public increases, sharing data is likely to become part of the public's general expectations of the role of a scientist. We have anecdotal evidence that Freedom of Information (FOI) requests to universities and research funders are increasing. This demonstrates an increase in interest from different sectors of the public that needs to be considered. Sharing findings (and sometimes raw data) with the public and other stakeholders outside

[5] NIHR UK Clinical Trials Gateway (2011). http://www.ukctg.nihr.ac.uk/default.aspx
[6] The Royal Society and Academy of Medical Sciences (2006). *Pandemic influenza- science to policy.* http://www.acmedsci.ac.uk/p99puid89.html
[7] GeneDB: hosted by the Sanger Institute (2011). http://www.genedb.org/Homepage

the scientific and policy community has many benefits. It can increase public engagement with science and it may boost rates of clinical trial enrolment. Increasingly, the availability of data allows scientists to harness the creativity of 'citizen scientists'. An open and transparent attitude by the scientific community can foster an interest in medical science and can enhance the evidence base for medical practice, which can in turn improve the way that future findings are interpreted by the public and the media. However, we emphasise the importance of robust peer review in raising the standards of information available; work that is released into the public domain without some level of quality assurance could lead to situations where imperfect or incorrect science is used by the media and others. Ultimately this could be detrimental to the public's overall trust in research.

The challenges involved in sharing scientific information

15. While there is a strong case in favour of data sharing, care needs to be taken when releasing raw or partially processed data. Datasets need to be accompanied by methodological information that enables accurate interpretation and reuse of the raw data. This should make clear the possibility of confounding factors, or the use of necessary manipulations. Consideration also needs to be given to: the cost of making quality-assured raw data available in an accessible form; and ensuring information is provided in a standardised format via appropriate open-access software. While there are advantages of ensuring raw data are subject to 'peer-review' in advance of release to ensure quality and consistency, consideration also needs to be given to the additional expense and delays this may incur.

16. Consideration of appropriate data sharing and, for example, the time period for which the data are available, must account for the important incentive provided by intellectual property (IP) rights. While IP should not be used to unnecessarily restrict access to data, it is an important mechanism for encouraging private investment and the translation of research into health and wealth benefits. We recognise that, in many cases, researchers will want a certain amount of time to exploit and harness datasets, for example, in relation to large cohort studies where time may also be required for researchers to develop the protocols for analysing the data.

17. Sharing data has implications for researchers, even at the most junior level - many PhD theses are published online as a requirement for graduation. Additional time may be needed before making these works publically available, as students may wish to publish their work in journals after their viva. Publishing their theses before students have had time to write papers leaves these students vulnerable to being 'scooped' by other researchers.

18. Progress in science is dependent on maintaining a firm foundation of information, on which future scientists can base their research. Peer review of research and data is essential in securing this firm foundation, but we have previously highlighted how high rejection rates from journals, and the requirement for unnecessary revisions before publication, can delay access to the outcomes of funded research, waste the time of researchers and ultimately slow the progression of science.[8] We welcome initiatives by established open access journals such as the Public Library of Science (PLoS) and new journals, such as that being established by the Wellcome Trust and partners, which seek to increase the speed of

[8] Academy of Medical Sciences (2011). *Response to House of Commons Science & Technology Inquiry into Peer Review.* http://www.acmedsci.ac.uk/p100puid217.html

publication.[9] There is a particular challenge in getting negative results published, despite the fact that they can be as informative as positive ones, although the *Journal of Negative Results in Biomedicine* is seeking to address this.[10]

19. Careful consideration should be taken before sharing data that could have national security implications (for example, bioterrorism) or sharing data that could have unintended consequences. These issues have previously been discussed by the Royal Society.[11]

A case study: patient data

20. Patient data are essential to research that underpins our knowledge of disease, the development of diagnostic and therapeutic interventions, and the delivery of services. This information can include both health data (e.g. cholesterol levels or cancer diagnoses) and non-health data (e.g. postcode, ethnic group or occupation). Accurate and timely sharing of personal data is essential for functions beyond those connected to individual treatment, which ensure that the delivery of health care is high quality, cost-effective, efficient and evidence based. Such secondary uses of data include: medical research; clinical and financial audits; health service planning; resource management; teaching and training; national statistics; public health surveillance and drug safety monitoring. We have previously highlighted the Million Women Study as a prime example where patient data have been used to test hypotheses that would be impossible in any other context.[12]

21. The respect and protection of personal information is one of the foremost responsibilities of health professionals. However, in a modern health care setting, the normal processes of care require ever more frequent judgements to be made about when, how and with whom sensitive personal data can be shared. The manner in which data are shared must reflect the obligations and expectations of confidentiality and effective procedures should be in place to prevent the unintentional disclosure of sensitive data; data should only be used for properly authorised purposes; and those handling sensitive data must understand and respect patients' interests.[13]

22. The medical record has also evolved in recent years and is an essential method of sharing information between healthcare professionals. A wide range of personal data are now included in medical records, such as information on lifestyle and family history, clinical and social factors, as well as diagnostic and other test results. In most cases, patient data that can be used in health research are collected by the NHS. These include records at GPs' surgeries or hospitals, collected as a routine part of patient care. It is important to note that these data are used extensively within the NHS to underpin all aspects of service delivery and, as such, are routinely shared in a secure and confidential manner with members of clinical care teams. Data are also shared within organisations undertaking clinical audit or to evaluate compliance to NHS standards.

23. With the development of electronic records across the NHS there is a real opportunity to maximise the potential of patient records in evaluating interventions, in epidemiological studies and in surveillance of infectious and non-communicable diseases. The UK has the

[9] Wellcome Trust (2011). *Open access Journal.* http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Open-access/Journal/index.htm
[10] Journal of Negative Results in Biomedicine (2011). http://www.jnrbm.com/
[11] The Royal Society (2009). *New approaches to biological risk assessment*. http://royalsociety.org/New-approaches-to-biological-risk-assessment/
[12] The Million women study (2011). *A confidential national study of women's health.* http://www.millionwomenstudy.org/
[13] Academy of Medical Sciences (2006). *Personal data for public good: using health information in medical research.* http://www.acmedsci.ac.uk/index.php?pid=48&prid=5

potential to lead the way in this field but concerted action is needed to maximise our assets, particularly the advantages of having a single national healthcare system.

24. Broadly speaking, the public is supportive of health research and willing to share personal information providing appropriate measures are in place.[14] However this support will always be conditional and there is a need to better communicate to the public and patients what is meant by the use of patient data in research, and to improve public engagement in discussions relating to policy decisions in this area.

25. Patient data can be accessed for use in research in several forms:[15]
    - Identifiable data. These include information in patient records such as patients' names, addresses, postcodes, dates of birth, dates of death and NHS numbers. There are also aspects of health data that could become identifying when they relate to a diagnosis of a rare condition, or when combined with other data.
    - Key-coded data (also called pseudonymised data). These cannot directly identify an individual, but a 'key' is available that enables the patient's identity to be re-linked to the data by a person or technology with access to the 'key'.
    - Anonymised data. There is no way of linking the data with the original patient record.

26. As highlighted in the Academy's review of the regulation and governance of health research,[16] it is essential that access to patient data both enables high-quality research and is done under conditions that enjoy public and professional support. It is an important aspect of research studies using patient data that different sources of data can be brought together and linked. This is usually initially at the level of the individual data subject, even if the datasets are subsequently only made available to researchers in anonymised or key-coded form. Technological and methodological advances in approaches to linkage that preserve confidentiality are a priority of many recent data initiatives (e.g. the Scottish Health Informatics Programme). It is essential that efforts to link data reliably are not undermined (or impeded) by the processes of regulation and governance, and that linkage across different health sectors, government departments and geographical areas is possible.

27. There is a need for mechanisms that allow approved researchers to access patient records in confidence, so as to be able to identify eligible patients for specific research studies. The development of what are known as 'safe havens' (or honest brokers) has become a well established concept around the use of data in research in recent years. Safe havens are secure environments for coding and handling data and have three key characteristics (as outlined in the Data Sharing Review[17]):
    - They provide a secure environment for processing identifiable personal data.
    - Only 'approved researchers' can gain access to the data.
    - There should be penalties for anyone who abuses personal data.

28. As recommended, in the Academy's report '*A new pathway for the regulation and governance of health research'* (2011), the health departments should continue their work to establish safe havens through the Research Capability Programme and its equivalents in the devolved nations. In particular, the Research Capability Programme should roll out the

[14] Academy of Medical Sciences (2011). *A new pathway for the regulation of health research.* http://www.acmedsci.ac.uk/index.php?pid=47&prid=88 (Section 6.4.5 & Box 6.6)
[15] Academy of Medical Sciences (2006). *Personal data for public good: using health information in medical research.* http://www.acmedsci.ac.uk/index.php?pid=48&prid=5
[16] Academy of Medical Sciences (2011). *A new pathway for the regulation of health research.* http://www.acmedsci.ac.uk/index.php?pid=47&prid=88
[17] Thomas R & Walport M (2008). *Data sharing review report.* http://www.justice.gov.uk/reviews/docs/data-sharing-review- report.pdf

full system as soon as possible, incorporating lessons learnt from the pilot, to ensure the UK is maximizing opportunities in this area. If necessary, legislation should be introduced that would enable safe havens to operate as laid out in the Data Sharing Review, so that researchers have access to secure data and that patient safeguards are fully met.

29. The Academy has also previously recommended that:
    - The forthcoming revision of the EU Data Protection Directive 95/46/EC87 provides an opportunity for clearer interpretation of the UK Data Protection Act, in relation to the use of patient data in research, as well as a chance to introduce further clarity into its text. Recent discussions indicate that there continues to be a need for clarification, as interpretation of the Act varies widely among researchers, R&D offices and ethics committees. The key aspects that should be considered are: definitions relating to consent requirements and the associated processes; how the Data Protection Act fits with the rest of the regulation pathway in relation to access to patient data for use in research for benefit to patients; the proportionality of the Data Protection Act; and clarity on roles and responsibilities for data controllers and data processors focusing on the impact in NHS R&D offices.
    - The definition of a clinical care team should be clarified so that approved members of research teams are considered members of the clinical care team and therefore have the same contractual obligations (i.e. the same sanctions for any breach of confidentiality).

The role of stakeholders in increasing access to information

30. There are several challenges and barriers to sharing data and findings, and a coordinated effort from all stakeholders is needed to overcome them. All stakeholders have a key role to play in encouraging the sharing of data. Importantly, scientists are central to creating a data sharing culture; scientists who share data should be given recognition for their work, and scientists who make use of others' data should do so responsibly and with appropriate acknowledgement. Much could be done to increase scientists' awareness of the benefits of data sharing, and they should receive training on the best ways to share their data. Some institutions provide repositories of data that benefit the whole scientific community. As previously noted, some institutes and funders publicise their current (unpublished) research, which reduces duplication of effort and will raise awareness of data sharing opportunities around different projects.

31. We welcome steps taken by research funders such as the Wellcome Trust and the Medical Research Council to increase access to scientific information and data. This includes mandating publication of the work that they fund in open access journals, requiring data sharing plans in grant applications and providing toolkits to support researchers.[18] We hope others will follow suit in ensuring efficient and cost-effective data sharing mechanisms. In the medical sciences, specific online repositories of digitised data have been created, including PubMed,[19] NCBI,[20] and EMBL-EBI.[21]

[18] Medical Research Council (MRC) digital curation centre (2011). *Position statement.* http://www.dcc.ac.uk/resources/policy-and-legal/research-funding-policies/mrc and the Wellcome Trust (2011). *The Wellcome Trust's open access policy.* http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTD018855.htm.
[19] PubMed, US national library of Medicine and the National Institutes of Health (2011). http://www.ncbi.nlm.nih.gov/pubmed
[20] The National Centre for Biotechnology Information (NCBI) (2011). http://www.ncbi.nlm.nih.gov/
[21] The European Molecular Biology Laboratory, European Bioinformatics Institute: *EMBL Nucleotide sequence database* (2011). http://www.ebi.ac.uk/embl/

32. The Academy welcomes the increase in sharing of data (e.g. research findings and datasets) between industry, the NHS and academia. Such partnerships will enable the UK to remain a world-class leader in medical research. The benefits of cross-sector partnerships are shared across the sectors and examples of such collaborations include: Pfizer and the University College London Institute of Ophthalmology; the Manchester Cancer Research Centre and Astra Zeneca; and the unique collaboration between the University of Dundee's Division of Signal Transduction Therapy and five of the world's leading pharmaceutical companies.[22] Opportunities for flexible collaboration across sectors need to be seized and made possible by ensuring a world-class biomedical workforce with the skills to move between and bridge sectors. These partnerships must be underpinned by suitable agreements that balance risk across partners and facilitate innovation through the appropriate and realistic handling of background and emergent intellectual property.

33. We have previously highlighted the significant role that journals play in enabling the sharing of data and findings. The advent of open access journals has improved access. We welcome commitments by journals such as PLoS to making data and other supporting information available and in collaborating with a number of subject-specific initiatives (e.g. relating to micro arrays and genes sequences) in order to develop relevant policies.[23]

34. Ultimately, all stakeholders have an obligation to promote data sharing, and a united effort is required to embed a culture of data sharing in science. Effective data sharing requires good feedback and communication among all parties. Such feedback could feed into the creation of 'best practice guidelines' on how researchers share data. Stakeholders should be involved in the development and dissemination of these guidelines and fully commit to their enforcement.

[22] Academy of Medical sciences (2010). *Academia, Industry and the NHS: Collaboration and innovation.* http://www.acmedsci.ac.uk/p101puid202.html

[23] PLoS One Journal (2011). *Sharing of Materials, methods and data.* http://www.plosone.org/static/policies.action;jsessionid=23BE4B67C04BB059A4ECCA32834E0C5C.ambra02#sharing

This response was prepared by Harriet Dickinson (Policy Intern) and informed by the Academy's Fellows and our previous reports and statements. We are also grateful for discussions with colleagues and the Wellcome Trust and GlaxoSmithKline. This document was approved on behalf of the Academy by the Foreign Secretary Professor Robert Souhami FMedSci. For enquiries concerning this response, please contract Dr Rachel Quinn (rachel.quinn@acmedsci.ac.uk).

The Academy of Medical Sciences
The Academy of Medical Sciences promotes advances in medical science and campaigns to ensure these are converted into healthcare benefits for society. Our Fellows are the UK's leading medical scientists from hospitals and general practice, academia, industry and the public service.

The Academy seeks to play a pivotal role in determining the future of medical science in the UK, and the benefits that society will enjoy in years to come. We champion the UK's strengths in medical science, promote careers and capacity building, encourage the implementation of new ideas and solutions – often through novel partnerships – and help to remove barriers to progress.

Academy of Medical Sciences,
41 Portland Place,
London,
W1B 1QH.
+44(0)20 7969 5288
info@acmedsci.ac.uk
www.acmedsci.ac.uk
Registered Charity No. 1070618
Registered Company No. 35202.