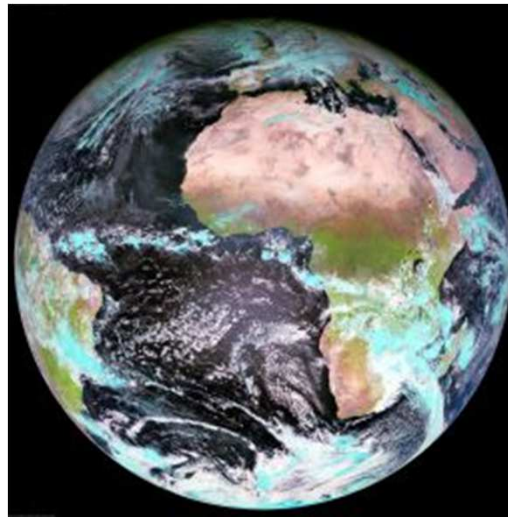
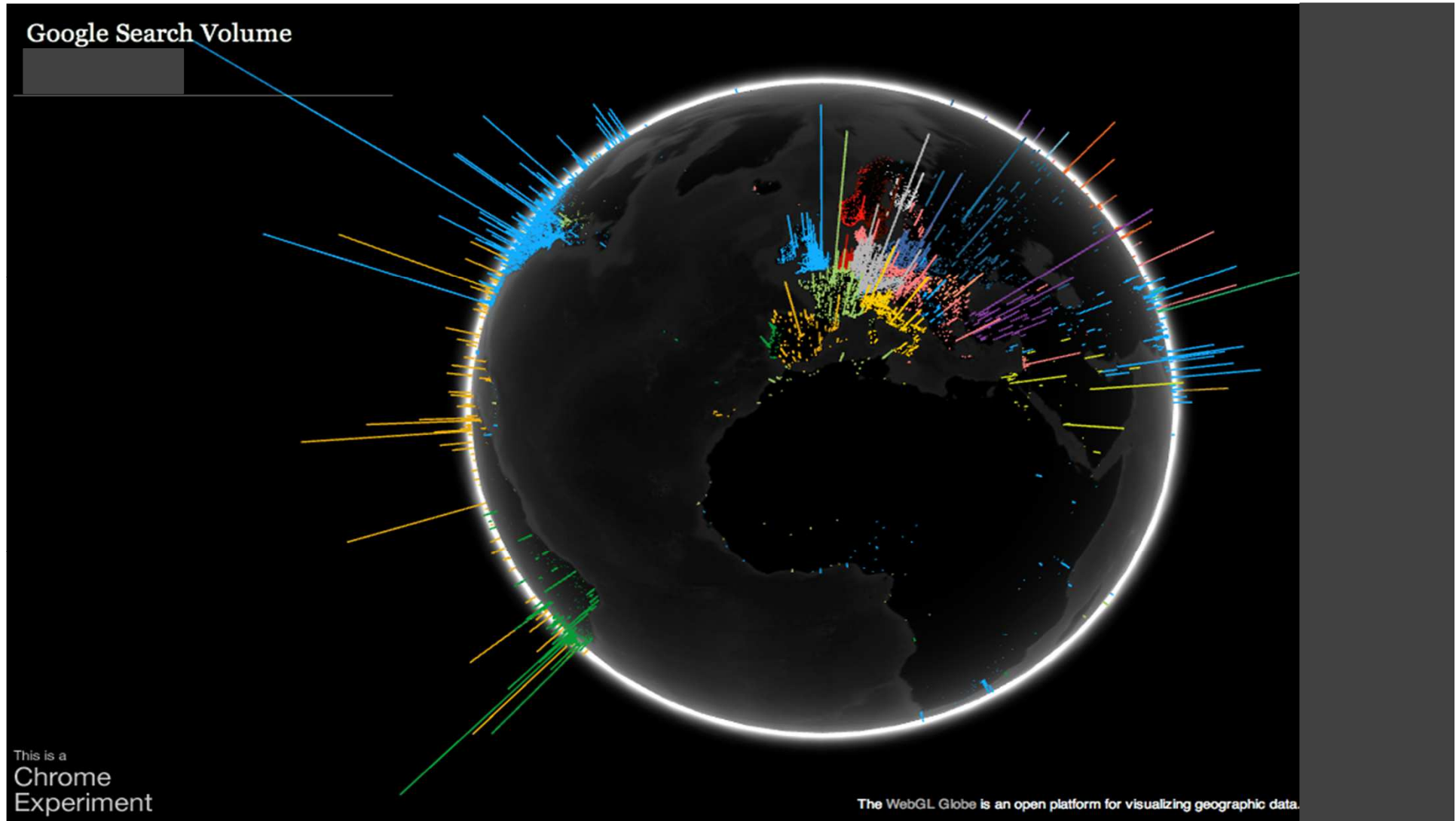


Data Matters: a view from a research funder



David Lynn
Wellcome Trust
d.lynn@wellcome.ac.uk

The world today:
knowledge sharing



The world today:
communications



The Wellcome Trust

- a global charitable foundation dedicated to achieving extraordinary improvements in human and animal health
- support the brightest minds in biomedical research and the medical humanities
- Sir Henry Wellcome, was a business man, philanthropist and collector, dedicated to enhancing global health and preserving and sharing his diverse collections
- in our work today, we have retained a long-standing commitment to maximising the long-term value of research outputs (including research papers and datasets)



Sharing research data: a long track record...

- building key data resources for research community:
 - lead role in major international projects – including Human Genome Project
 - funding key databases (e.g. EBI resources)
- brokering policy consensus on data sharing issues:
 - Bermuda principles (1996),
 - Fort Lauderdale principles (2003)
- leading advocate of open access publishing – requirement for all grants since 2006



Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility

Report of a meeting organized by the Wellcome Trust and held on 14-15 January 2003 at Fort Lauderdale, USA.



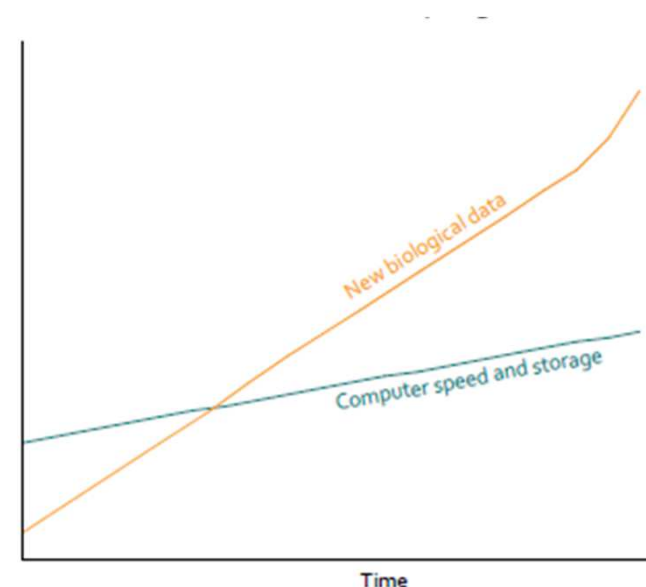
FRONTIERS MEETING
USE OF ELECTRONIC
PATIENT RECORDS
FOR RESEARCH
AND HEALTH BENEFIT
24-25 May 2007



wellcome^{trust}

Rising volume and complexity of research data

- according to the EBI, the volume of new biological data is **doubling every 5 months**
- the growth is particularly rapid in genomics:
 - we can now analyse the equivalent of a human genome every 14 minutes at a cost of \$5,000 - **400 times quicker** than when the draft human genome was first published in 2000.
 - expected that sequencers will become **1000 to 1,000,00 times more productive** over the next decade



Data management and sharing policy

- Our data management and sharing policy was published in Jan 2007:
 - requires **all** funded researchers to maximise access to research data with as few restrictions as possible
 - requires **data management & sharing plans** for proposals generating data that could be shared for added value
 - commits to meet costs for data sharing activities outlined in the plans
- The policy was reviewed in 2010:
 - we produced new guidance to clarify what should be covered in data sharing plans
 - we will track adoption of plans through ongoing liaison with researchers



An example – the Wellcome Trust Case Control Consortium (WTCCC)

- researchers recognised the massive potential of genome-wide association studies to characterise the genetic factors underlying diseases
- in order to achieve required power, research groups needed to join forces to study effects in large numbers of patients and healthy volunteers
- WTCCC (launched in 2005) – brought together over 20 leading UK groups to look at 500,000 points of genetic variation in 17,000 samples
- initiative has had two further phases – now involves over 50 groups, and has identified over 90 new variants in diseases analysed
- Trust has worked with WTCCC to establish its data access mechanisms – data rapidly available via application to data access committee



Data sharing – the key challenges

Infrastructural:

sustaining the infrastructure required for long-term data storage and curation

Cultural:

incentives and recognition for researchers who share their data

Technical:

developing data standards, metadata, platforms needed for inter-operability

Professional:

training and career development of data specialists and bioinformaticians

Ethical:

protecting the confidentiality of research participants



Working together to address the challenges

We are working in partnership in the UK, Europe and internationally:

- working with other global health funders to promote sharing of research data to improve public health (joint statement in Jan 2011)
- partnering in the ELIXIR initiative to build sustainable funding model for life science data infrastructure in Europe
- strengthening data access & governance procedures for genetics & cohort studies with MRC, ESRC and Cancer Research UK
- contributing to key policy discussions (e.g. Royal Society project on Science and Public Enterprise; National Science Board)

Comment

Sharing research data to improve public health

The purpose of medical research is to analyse and understand health and disease. A key and expensive element is the study of populations to explore how interactions between behaviour and environment, in the context of genetic diversity, determine variation and variation in health and disease. As funders of public health research, we need to ensure that research outputs are used to maximize knowledge and potential health benefits. In turn, the populations who participate in research, and the taxpayers who foot the bill, have the right to expect that every last ounce of knowledge will be wrong from the research.

Ensuring data are made widely available to the research community accelerates the pace of discovery and enhances the efficiency of the research enterprise. In many research fields—from genetics and molecular biology to the social sciences—data sharing is already ingrained in how researchers work. In genetics and genomics, the pooling of studies of different populations has led to an explosion of knowledge on the genetic determinants of human variation in health and disease. Well-established repositories and tools enable researchers to access and interrogate shared data resources, and build on one another's work.

By contrast, this culture has yet to be widely embraced by the public health research community. Much of the infrastructure, technical standards, and incentives that are needed to support data sharing are lacking, and these data can hold particular sensitivities. And some researchers are reluctant to share data. Too often, data are treated as the private property of investigators who aim to maximise their publication record at the expense of the widest possible use of the data. This situation threatens to limit both the

Purpose (joint statement of purpose—vision, principles, and goals)

Vision
We intend to work together to increase the availability to the scientific community of the research data we fund, and to collect from populations for the purpose of health research, and to promote the efficient use of those data to accelerate improvements in public health.

Principles
Funders agree to promote greater access to and use of data in ways that are:

- **Equitable**: It should recognize and balance the needs of researchers who generate and use data, other analysts who might want to reuse those data, and communities and funders who expect health benefits to arise from research.
- **Ethical**: It should protect the privacy of individuals and the dignity of communities, while simultaneously respecting the imperative to improve public health through the most productive use of data.
- **Efficient**: It should improve the quality and value of research, and increase its contribution to improving public health, approaches should be proportionate and build on existing practice and reduce unnecessary duplication and competition.

Immediate goals

- Standards of data management are developed, promoted, and embedded so that research data can be shared routinely and research effectively.

(Continues in next column)

(continued from previous column)

- Funders and employees of researchers recognize data management and sharing of well-managed datasets as an important professional indicator of success in research.
- Researchers creating datasets for secondary analysis from shared primary data are expected to share those datasets and to work transparently and in research good practice, giving due acknowledgment to the generators of the original data.

Longer term objectives

- Data collected for health research are made available to the scientific community for analysis which adds value to existing knowledge and which leads to improvements in health.
- The research community, particularly those collecting data in developing countries, develop the capacity to manage and analyse those datasets locally, as well as contributing to international analysis efforts.
- To the extent possible, datasets underpinning research papers in peer-reviewed journals are archived and made available to other researchers in a clear and transparent manner.
- The human and technical resources and infrastructures needed to support data management, archiving, and access are developed and supported for long-term sustainability.

The full statement is online: <http://www.wellcome.ac.uk/publichealthdata>. Other funding organisations are invited to join as signatories and partners in this work.

www.nature.com. Published online January 2, 2011. DOI: 10.1038/10144-473610127361

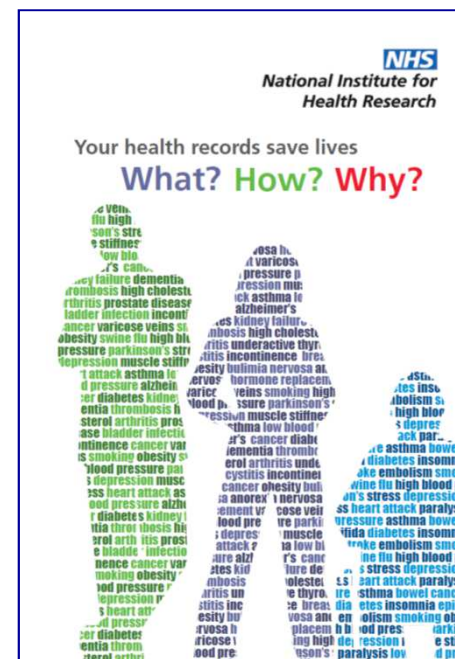
Promoting research access to public data - opportunities

- increasing research access to **administrative data** offers immense potential for health research and other fields
- we believe that use of **electronic patient records** will deliver major benefits for patient care, planning and service delivery – experience in Scotland demonstrates this
- there are also tremendous opportunities to utilise data from patient records in research in order to enhance:
 - pharmacovigilance
 - recruitment to clinical trials
 - epidemiology and cohort studies



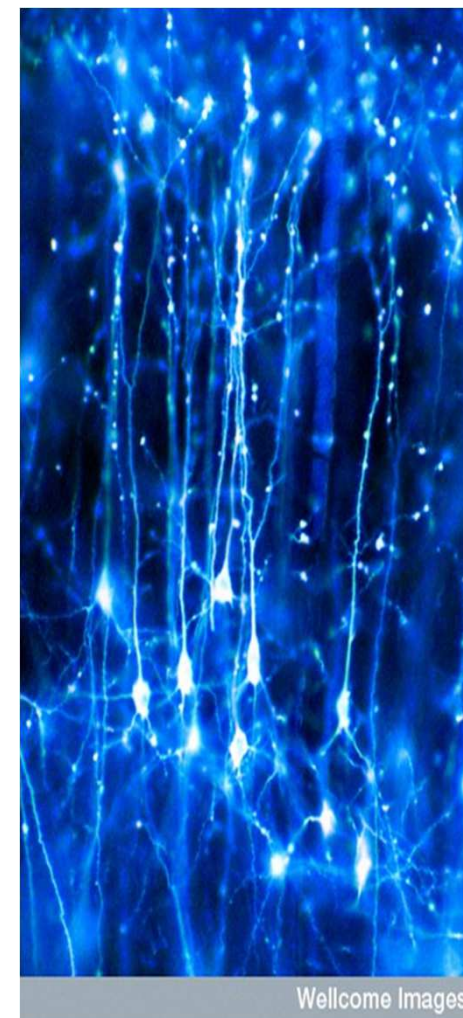
Realising the opportunities

- we are engaging actively in discussions on open data:
 - discussions with Cabinet Office and Treasury on the open data strategy and Plan for Growth
 - member of the new administrative data task force
- we are taking forward a broad range of activities to develop e-health
 - funding major research e-health initiatives in partnership with UK research councils
 - convening pilot programmes to inform the roll out of e-health systems in the NHS
 - engaging key stakeholders – patients and GPs
 - working with Government on implementation of the Clinical Practice Research Data Link



Building an enabling environment

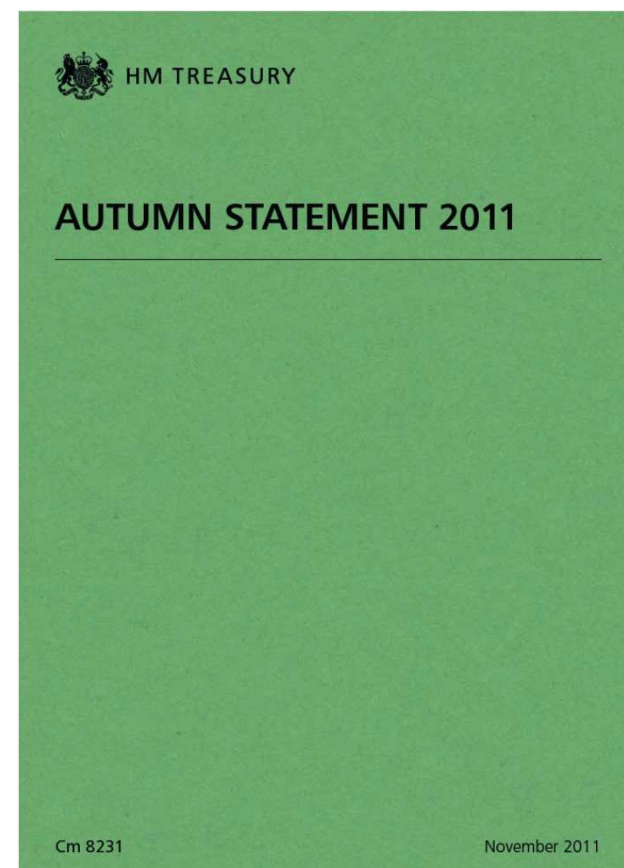
- the regulatory framework governing research access to data must achieve an appropriate balance:
 - ensuring proportionate safeguards to protect the privacy of data subjects...
 - ...whilst enabling legitimate research uses to maximise public benefit
- to achieve this, there is a need to:
 - reduce regulatory complexity and ensure consistent interpretation
 - ensure revisions to data protection legislation at EU level, and its UK implementation, recognise and support research use
 - build the systems ('safe havens') needed to enable data to be used safely and securely



Encouraging developments

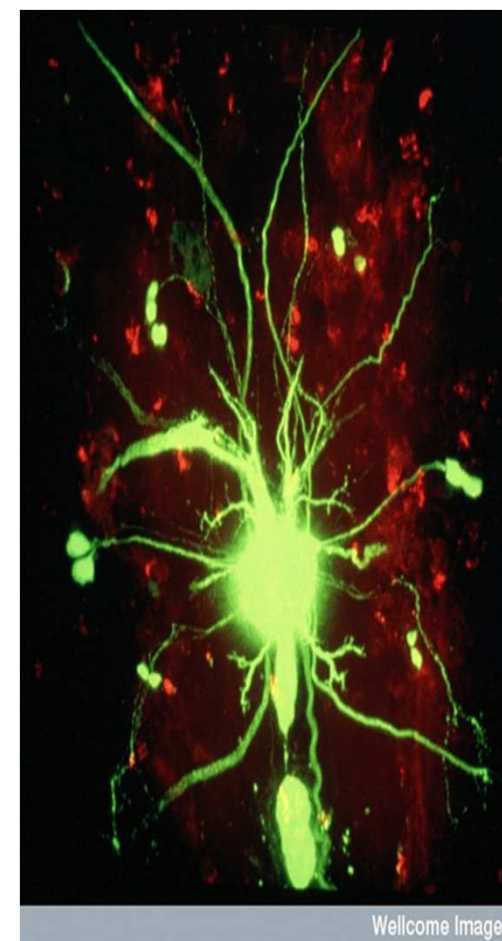
We welcome Government's recent commitments to:

- establish a new 'Open Data Institute' to innovate and exploit opportunities provided by the open data strategy
- link primary and secondary healthcare datasets, including a complementary secure service for research
- provide on-line access to health records for all patients in the NHS



Conclusion – our role as a funder

- As a research funder we are dedicated to ensuring that:
 - research data are managed, preserved and shared in a way that maximises public benefit
 - researchers can access data of value in a safe and secure manner
- We achieve this through:
 - **funding** major data resources and data sharing as an integral component of research
 - **policy leadership** to develop and implement policies to maximise value of data
 - **working in partnership** to build the resources and culture needed for data sharing
 - **advocacy** – influencing Government and contributing to key policy discussions



Further information

<http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/index.htm>

wellcome^{trust}

Our vision | Funding | Managing a grant | Education resources | News and features | Investments | **About us**

eGrants | [About](#) | [Log in](#)

[Publications](#) | [Jobs](#) | **[Policy](#)** | [Strategy](#) | [Organisation](#) | [History](#) | [Timeline](#) | [Contact us](#)

Data sharing

We aim to ensure that the data generated by the research we support is managed and shared in a way that maximises the benefit to the public.

We work actively with the research community and other relevant stakeholders to shape and apply good practice with regard to the management of research data.

In addition to implementing our policy, which applies across the full breadth of the research we support, we are taking forward targeted activities in partnership with others to promote data sharing in specific research fields.

Spotlight issues

- Counterfeit medicines
- Health impacts of climate change
- Influenza
- Personal information

Data sharing

- Data management and sharing
 - Guidance for researchers
 - Public health and epidemiology
 - Large-scale genetics research
- Open access
- Harmful misuse of research
- Human Fertilisation and Embryology Act

Data management and sharing policy

Our position statement on the management and sharing of data that arises from the research we fund, and associated FAQs.

Guidance for researchers

Resources to assist researchers in developing data management and sharing plans, including information on data repositories.

Public health and epidemiology research

We are working with others to develop agreed principles for data sharing among public health and epidemiological

Large-scale genetics research

We have a long track record of maximising access to genetic datasets and are working to address associated ethical

REPORT
Ethical, legal and social issues arising from the use of GWAS in medical research

[Download PDF](#)

Email: d.lynn@wellcome.ac.uk