



Government
Office for Science

Rigour and Openness in 21st Century Science

Sir Mark Walport, Chief Scientific Advisor to HM Government



Government Chief Scientific Adviser

Knowledge translated to economic

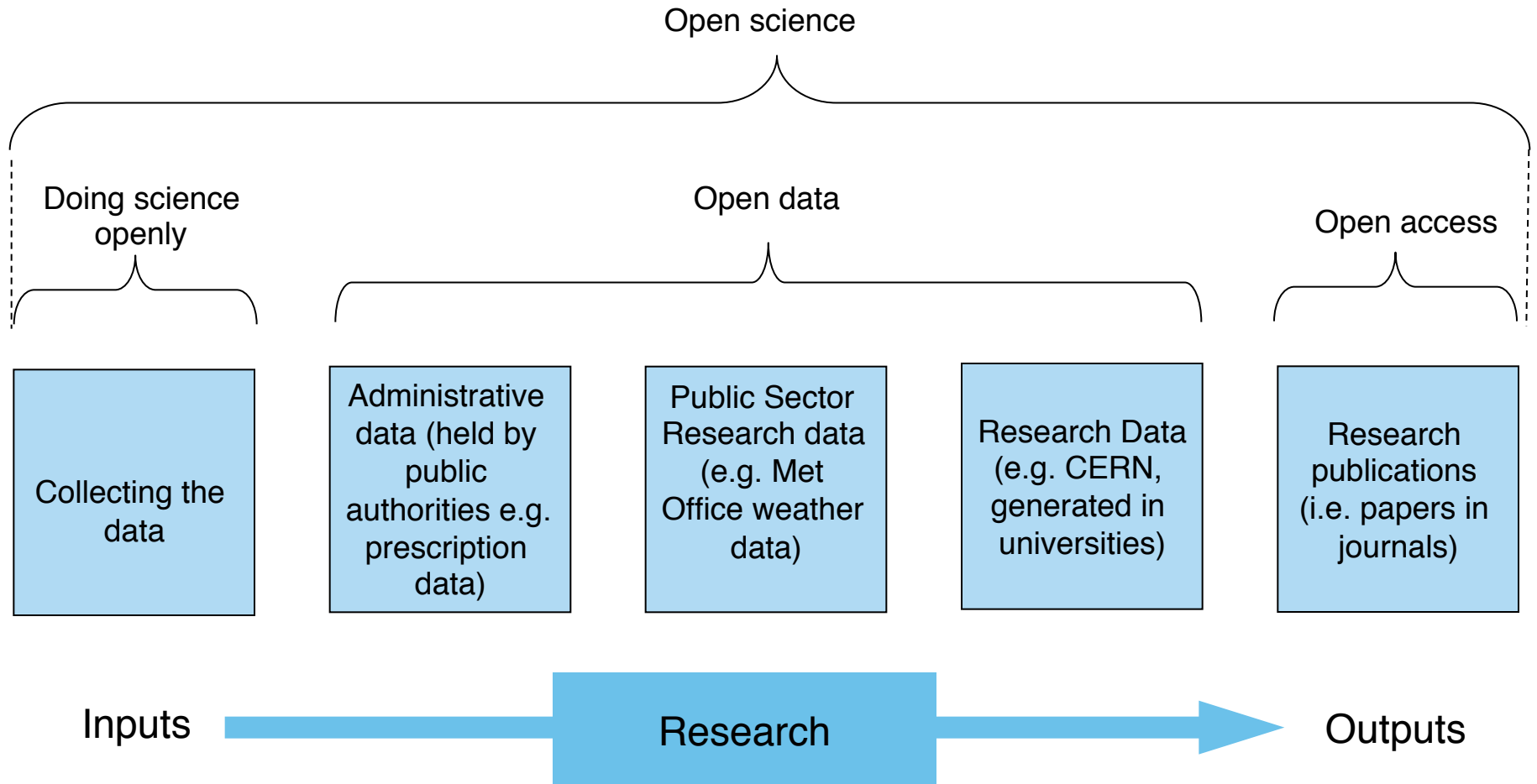
advantage: Promoting the contribution of science, engineering, technology and the social sciences to economic growth by linking industry, academia and government

- **Infrastructure resilience:** Developing the capabilities that are vital to the infrastructure that underpins our security, well-being and resilience
- **The right science for emergencies:** Providing the best scientific advice in the case of emergencies
- **Underpinning policy with evidence:** Ensuring the best use of quantitative and qualitative analysis across government
- **Advocacy and leadership for science:** Providing advocacy and strong leadership for science inside and outside government





A taxonomy of openness





Collecting the Data: professional community working in a different way

Examples:

- Genome project
- Citizen science:
 - AshTag
 - Galaxy Zoo

Benefits:

- Collaboration
- Scale
- Statistical power



Unraveling the genome...



- Human Genome Project
- SNP consortium
- Hapmap
- Cancer Genome Project
- Copy Number Variation
- WT Trust Case Control Consortium
- 1000 Genomes
- Encode
- UK 10,000 Genomes
- Deciphering Developmental Disease
- H3 Africa

eEnsembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Human (GRCh37) Location: 3:175,651,696-198,022,430

Location-based displays

- Whole genome
- Chromosome summary**
- Region overview
- Region in detail
- Scrollable region (BETA)
- Comparative Genomics
 - Alignments (image) (64)
 - Alignments (text) (64)
 - Region Comparison (65)
 - Synteny (15)
- Genetic Variation
 - Resequencing (20)
 - Linkage Data
- Markers
- Other genome browsers
 - UCSC
 - NCBI
 - Vega

Configure this page

Manage your data

Export data

Bookmark this page

Chromosome 3: 175,651,696-198,022,430

Assembly exceptions for chromosome 3: HSCRB3_1_CTG1, H6325_PATCH, H6186_PATCH, H6391_PATCH, H6380_PATCH, H61091_PATCH.

Chromosome summary

Chromosome 3 summary tracks: Protein Coding Genes, Non Coding Genes, Pseudogenes, % GC Repeats, Variations.

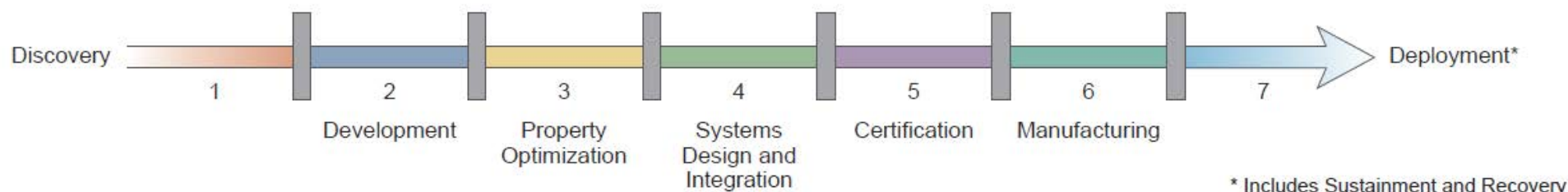
Common denominators...

- collaboration
- funding partnership
- public/private
- openness



Materials Genome Initiative

- Infrastructure to accelerate advanced materials discovery
- Time frame for incorporating new classes of materials into applications is typically about 10 to 20 years from initial research to first use.
- \$100M initiative proposes: open innovation; advances in modelling algorithms; a data exchange system.
- Aims to shorten cycle by 50%.



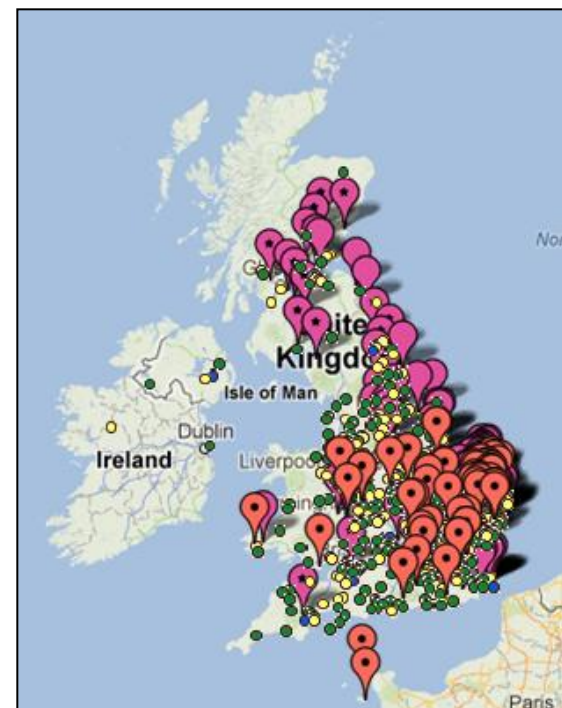
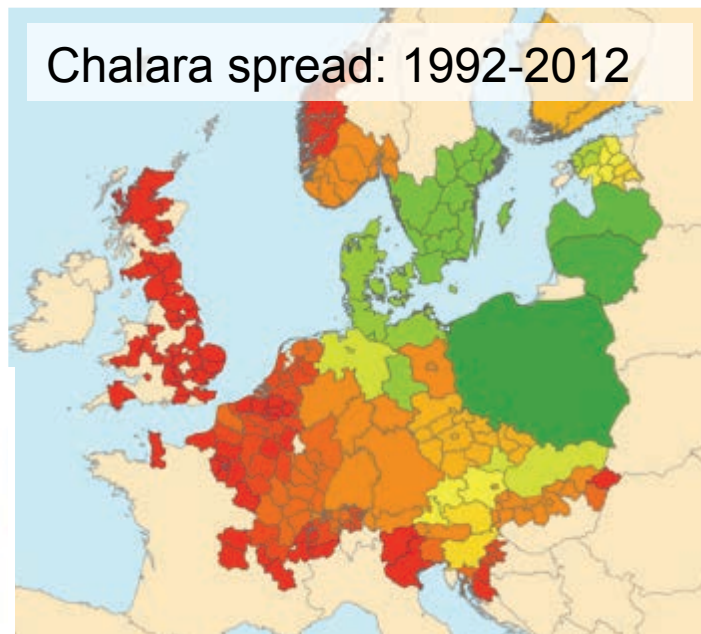


- Openly collected science is already helping policy makers.
- AshTag app allows users to submit photos and locations of sightings to a team who will refer them or to the Forestry Commission, which is leading efforts to stop the disease's spread with the Department for Environment, Food and Rural Affairs (Defra).

Citizen Science



Chalara spread: 1992-2012





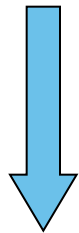
Massively Collaborative Mathematics

- Novel communication technologies are changing the social dynamics of science
- Massively Collaborative Mathematics: Tim Gowers and the Polymath Project.
- A blog serving as an open forum for contributors to work on a complex unsolved mathematical problem:
 - “A new combinatorial proof to the density version of the Hales-Jewett theorem”
- 27 people made more than 800 comments
- In just over a month, the problem was solved

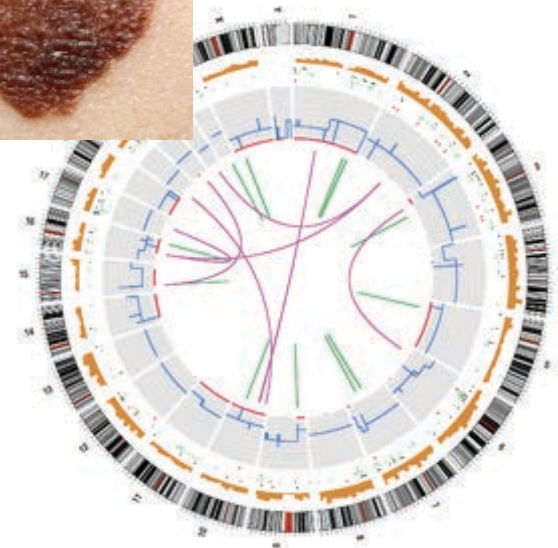
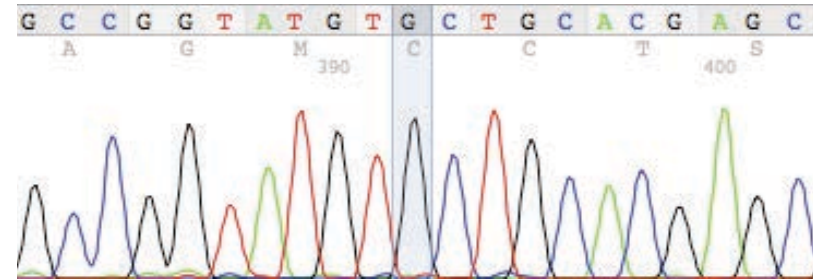


From data to knowledge to society

- Data
- Information
- Knowledge
- Application

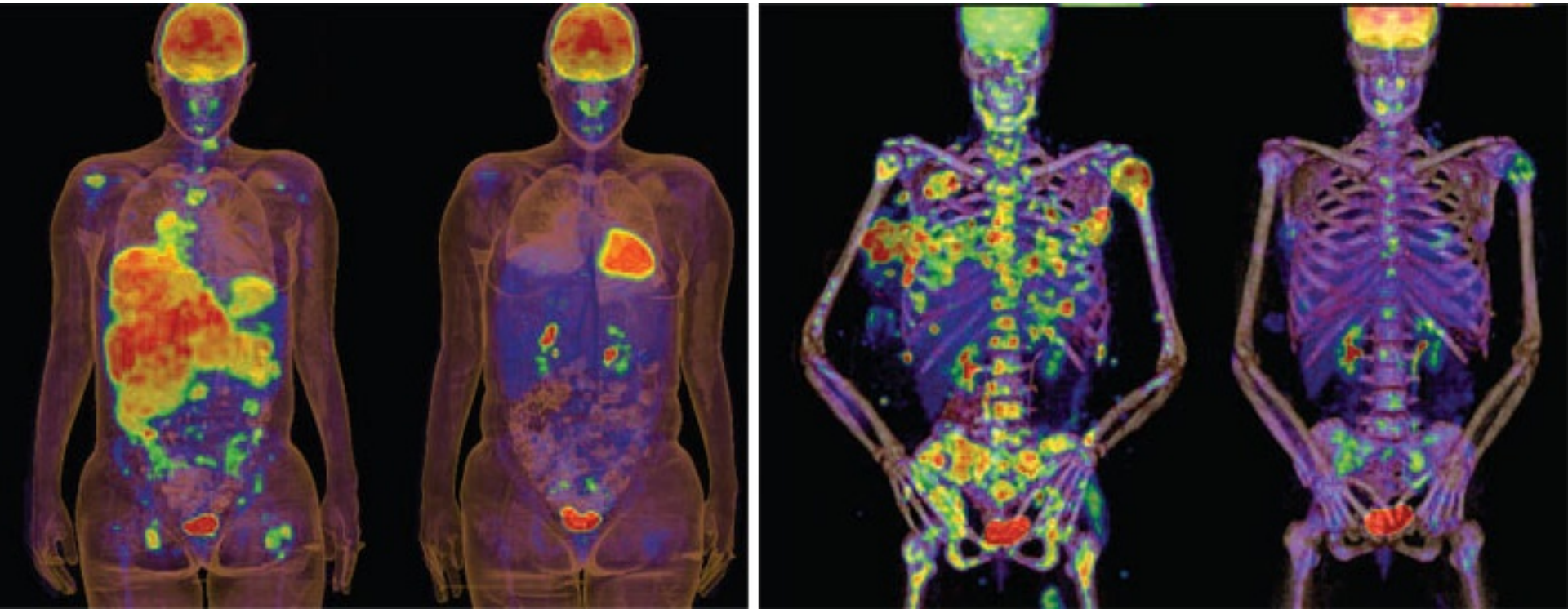


Societal benefit





From data to knowledge to society



Vemurafenib – oral targeted therapy for treatment of metastatic malignant melanoma in patients whose tumors carry the BRAFV600E mutation. Approximately 60% of melanoma patients have tumors that carry this mutation



Openness not an unalloyed good

A challenge to:

- Commerce
- Privacy
- Security



...Need intelligent openness

...Some dilemmas



Privately collected data often can be open...

Structural Genomics Consortium

- Not-for-profit public-private partnership to conduct basic science.
- Determine 3D protein structures which may be targets for drug discovery.
- Once such targets are discovered, they are placed in the public domain.

...but not always

- and what about issues of public safety:
 - Clinical trials,
 - Aircraft safety





Developed for one context, used for others

- Civilian \longleftrightarrow Military use
- Science subverted to do harm
- Civilian: developed in one domain – used in others





Dual use science

- Flu: Spanish flu sequencing 2005, H5N1
- Avian flu controversy, Netherlands and US teams introducing genetic changes to understand H5N1 transmissibility between species
- Publication in Nature and Science put on hold 2012
- International actors explored the issues:
 - Research is not risk free
 - Need to get framework right (legislative, ethical, physical)
 - Tools to address risks: risk assessment, ethical standards, checks and balances, eg peer review
- Self governance and the role of scientists – research funders, science press: putting in place standards and guidance to balance benefits of publication vs potential harm

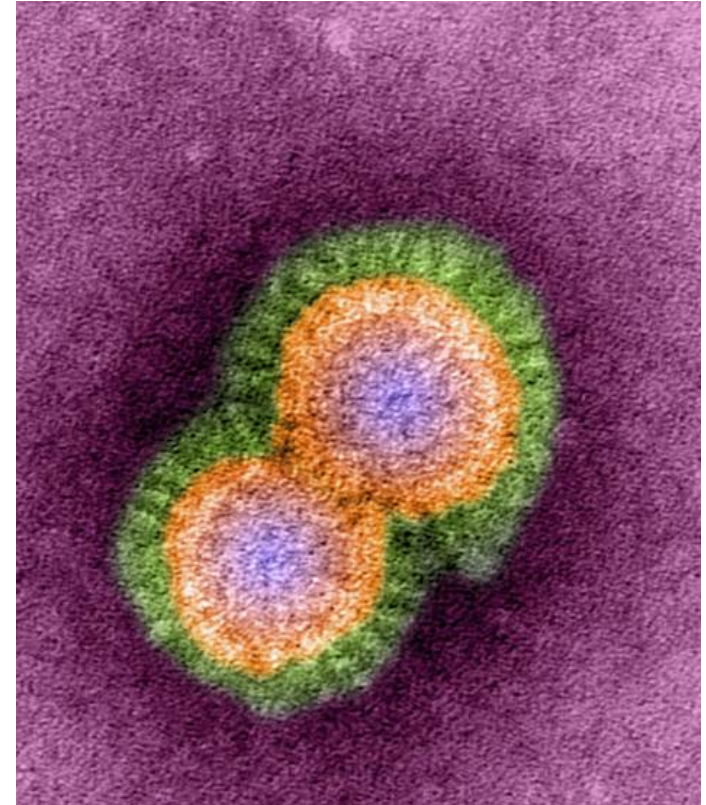
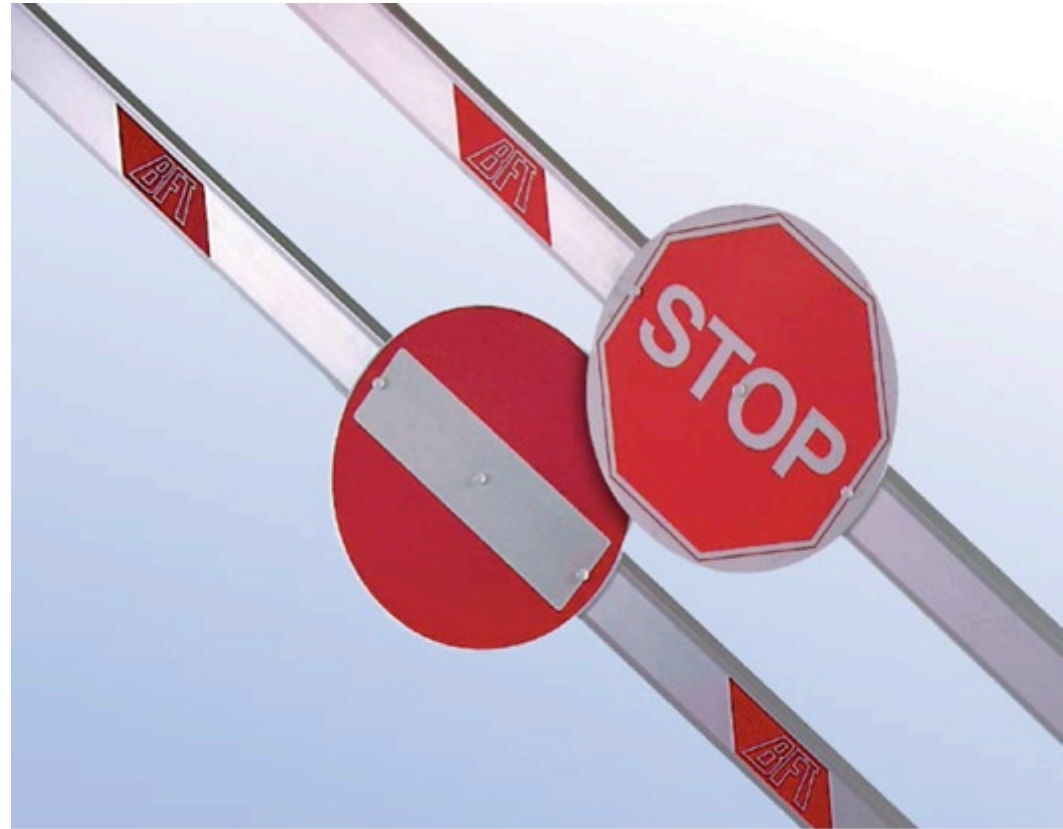


Image: Science Photo Library



The default should be openness

- Where it has worked
- Where lack of openness has caused problems
- Where there are genuine difficulties





Gastro-intestinal infection in Hamburg

- E-coli outbreak spread through several countries affecting 4000 people
- Strain analysed and genome released under an open data license.
- Two dozen reports in a week with interest from 4 continents
- Crucial information about strain's virulence and resistance





Climate change transparency

THE INDEPENDENT CLIMATE CHANGE

News Opinion **Environment** Sport Life & Style Arts & Entertainment

Climate Change Green Living Nature

Home > Environment > Climate Change

Climategate scientist 'hid flaws in data', say sceptics

Professor in leaked email scandal tried to hide fact that numbers he used were wrong

guardian.co.uk

Phil Jones survives MPs' grilling over climate emails

TIMESONLINE

UN climate change expert: there could be more errors in report





Developing world - Indonesian flu

- There are difficulties in ensuring access to data from developing countries.
- Some uneasy at the prospect that those with greater scientific resources will benefit overseas interests, to the *detriment* of home researchers.
- Indonesia ceased providing access to their flu samples in 2007
- This policy was reversed only after the World Health Organisation put in place protocols for equitable access to vaccines and medicines in future pandemics.





- Research isn't finished until it is published
- Maximise impact of research by maximising distribution
- Publication is a cost of research
- IT revolution enables revolution in methods in disseminating the findings of research



e
LIFE



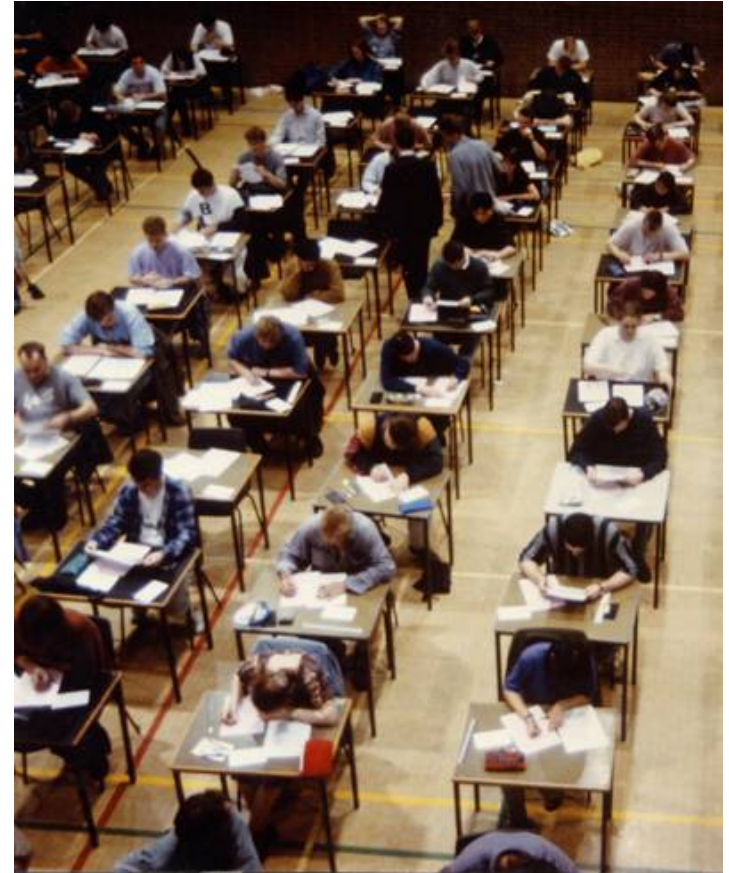
Open access offers much more than just the 'paper'

- Massive data sets and metadata
- Data display
- Text mining
- Community annotation and feedback
- Other linkages





- How we can improve process:
 - Scale, collaboration
 - Scrutiny
 - Enabling reproducibility
- Honest error vs fraud
- The best way forward is through openness

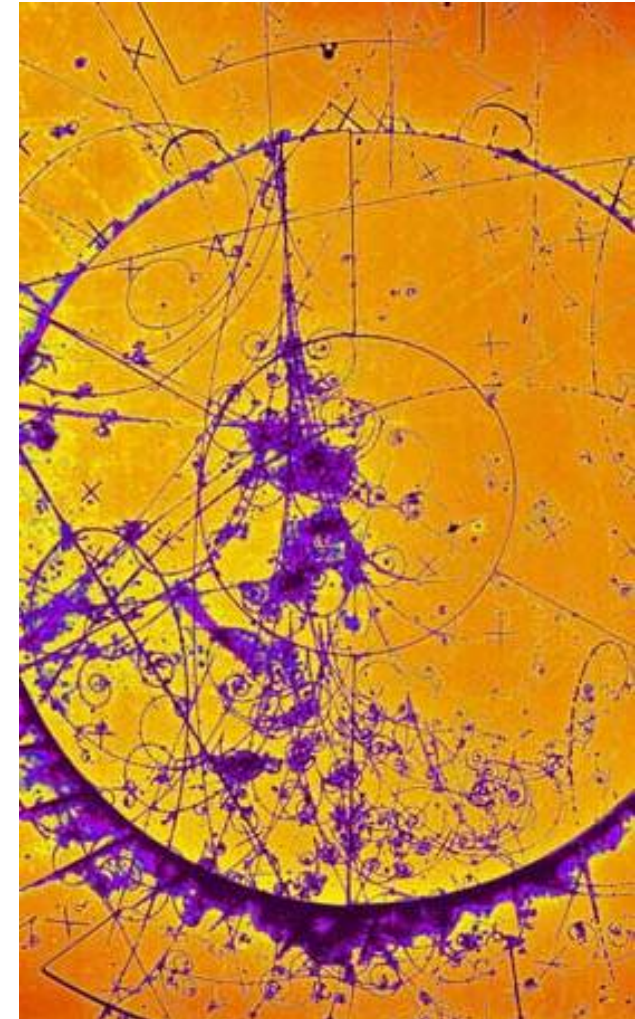




'Faster than light' neutrinos

Image: Science Photo Library

- Muon neutrinos fired 730 km from CERN to the Gran Sasso National Laboratory, central Italy
- Neutrinos seemed to travel faster than the speed of light
- CERN opened the result to broader scrutiny
- More than 200 papers appeared on arXiv.org attempting to debunk or explain the effect
- Team later reported two flaws in equipment set-up: a fiber optic cable attached improperly & a clock oscillator ticking too fast





The right open data enables:

- Measurement
- Accountability
- Better execution
- Innovation





Harnessing ICT: A national diabetes system for Scotland

Total Scottish Population 5.2M

People with diabetes : 251,132 (4.9%)

People with Type 1 DM : ~27,000
(0.5%)

All patients nationally are registered
onto a single register; the SCI-DC
register

SCI-DC used in all 38 hospitals

Nightly capture of data from all 1043
primary care practices across Scotland

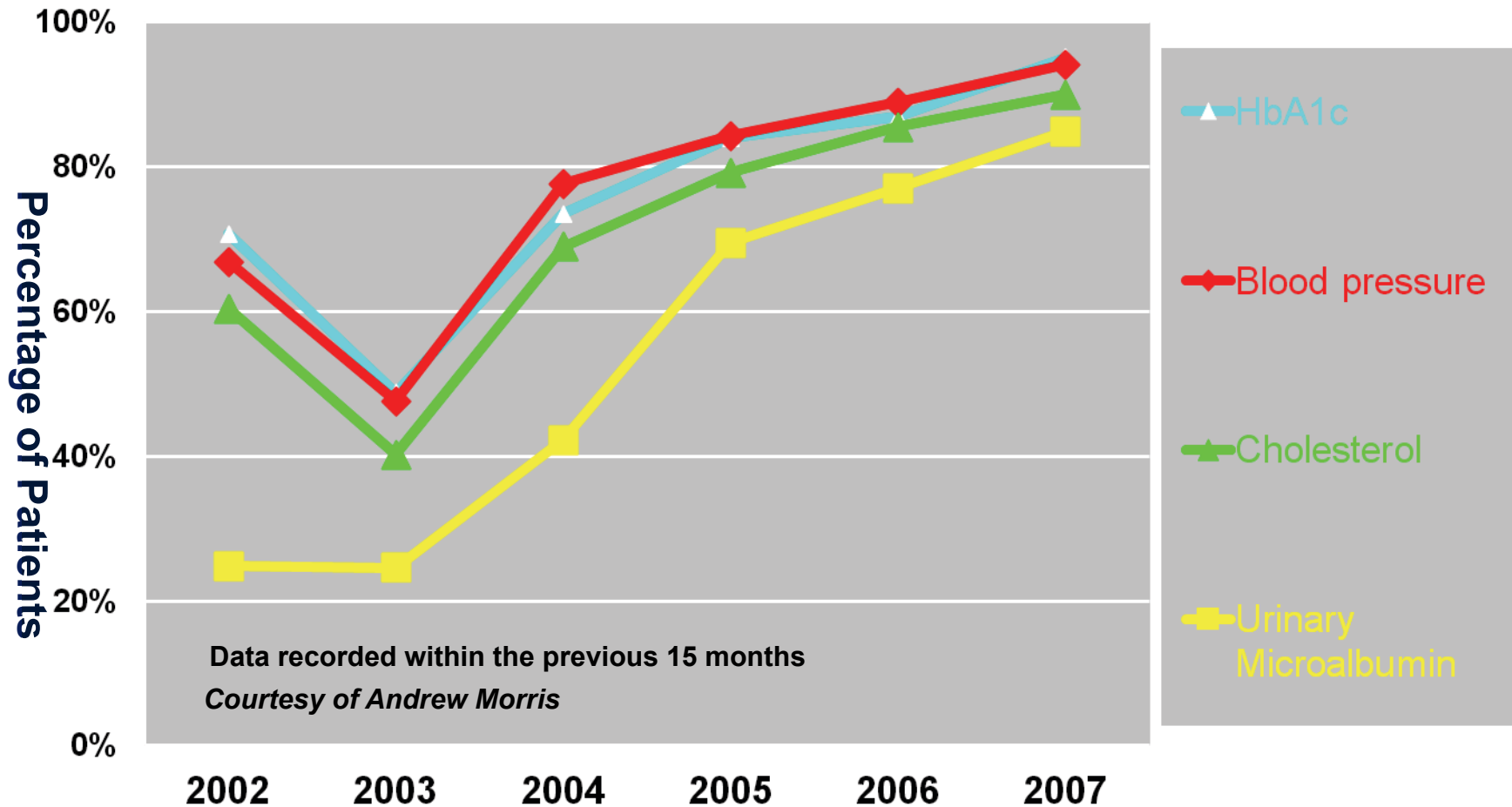
Courtesy of Andrew Morris





Scottish Diabetes Survey – over 90% capture of key variables since 2007

Recording of Key Biomedical Markers



<http://www.diabetesinScotland.org.uk/Publications/SDS%202010.pdf>



Improved clinical outcomes

Original Article: Complications

Decreasing amputation rates in patients with diabetes—a population-based study

C. J. Schofield, N. Yu*, A. S. Jain and G. P. Leese

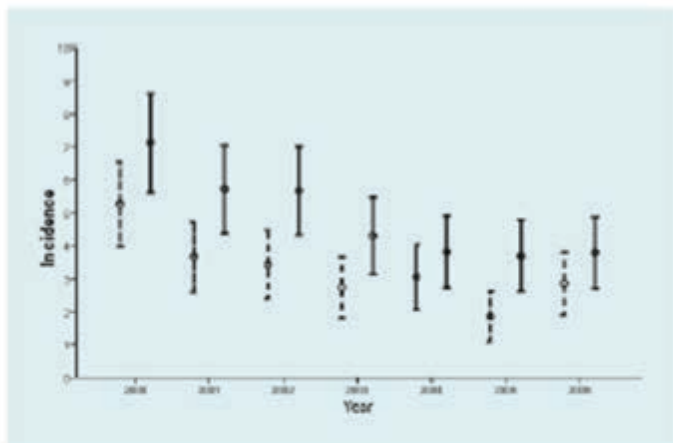
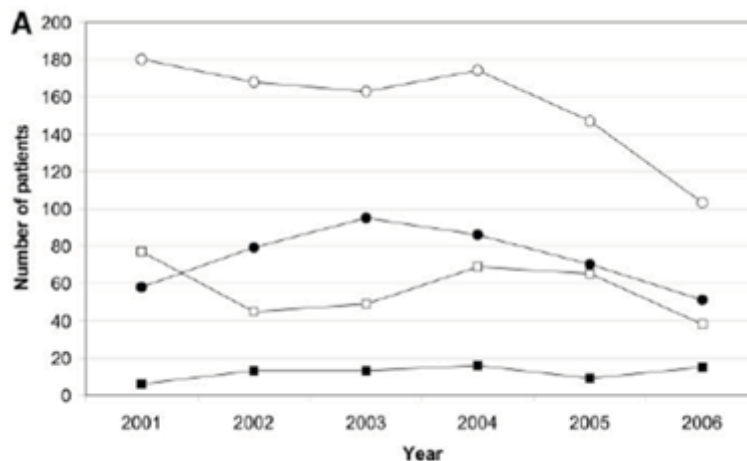


FIGURE 1 Incidence of adjusted total amputation with 95% confidence intervals (solid lines) and incidence of adjusted major amputation with 95% confidence intervals (dashed lines) per 1000 patients with diabetes. Incidence adjusted for age and sex.

Diabetic Medicine 2009

Diabetic Retinopathy: More Patients, Less Laser

A longitudinal population-based study in Tayside, Scotland



Diabetes Care 2008

Courtesy of Andrew Morris



Privacy and confidentiality

- Personal information is individual & precious to each one of us – it's vital that we treat it properly
- A balancing act...between the right to privacy and the necessity to hold and share data
- A framework is needed:
 - to protect individuals
 - build & maintain confidence
 - facilitate research





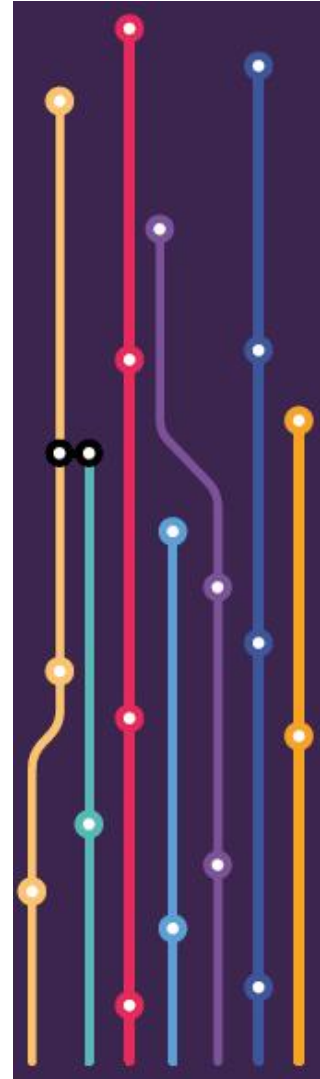
Good policy depends on the best data

Challenge to bring together data sets:

- across government
- and from the private sector

Administrative Data Taskforce

- Formed in December 2011, reported in December 2012.
- Examined the best procedures and mechanisms to make administrative data available for research safely
- Aim allow research that is already technically feasible to be undertaken with integrity in a much more consistent, reliable and efficient manner
- Value to academic research, but also research and policy evaluation within government departments



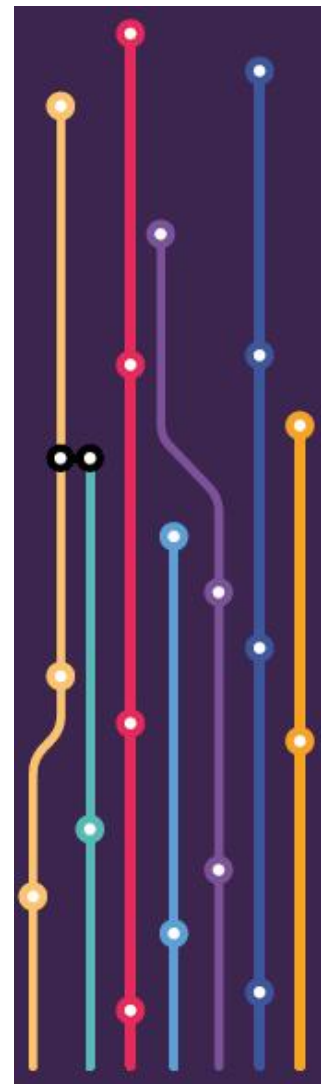


Administrative Data Taskforce

5 key recommendations:

- An Administrative Data Research Centre (ADRC) should be established in each of the four countries in the UK
- Legislation should be enacted to facilitate research access to administrative data and to allow data linkage between departments to take place more efficiently
- A single UK-wide researcher accreditation process, built on best national and international practice, should be established
- The centres should put in place plans for engaging with the public
- Funding should be provided

Government to respond by Summer 2013





Open science requires:

- Collecting the data in the best way
- Sharing the data openly
- Making the products of research available openly

But open data per se not an unalloyed good. Needs to be:

- Intelligible
- Accessible
- Assessable
- Useable

When we get that right we are able to ensure:

Data → Information → Knowledge → Application → Societal benefit