



Data sharing in policy and practice with Cancer Research UK

Paola Quattroni
CRUK Data Research Funding Manager

University of Cambridge 11.05.17

WHAT TO EXPECT FROM THIS WORKSHOP

- ✓ Cancer Research UK's ambition and strategy
- ✓ Data sharing policy and implementation
- ✓ Cancer Research UK activities
- ✓ Incentives to increase data sharing
- ✓ How can we work together?

WHAT TO EXPECT FROM THIS WORKSHOP

- ✓ Cancer Research UK's ambition and strategy
- ✓ Data sharing policy and implementation
- ✓ Cancer Research UK activities
- ✓ Incentives to increase data sharing
- ✓ How can we work together?

CANCER RESEARCH UK'S AMBITION

Accelerate progress and see three quarters of people surviving cancer within the next 20 years

- We support the highest quality research then translated to patient benefit
- The research that we fund is a public good and should be made available to everyone



TO BRING FORWARD THE DAY ALL CANCERS ARE CURED



FUNDING THE BEST SCIENCE

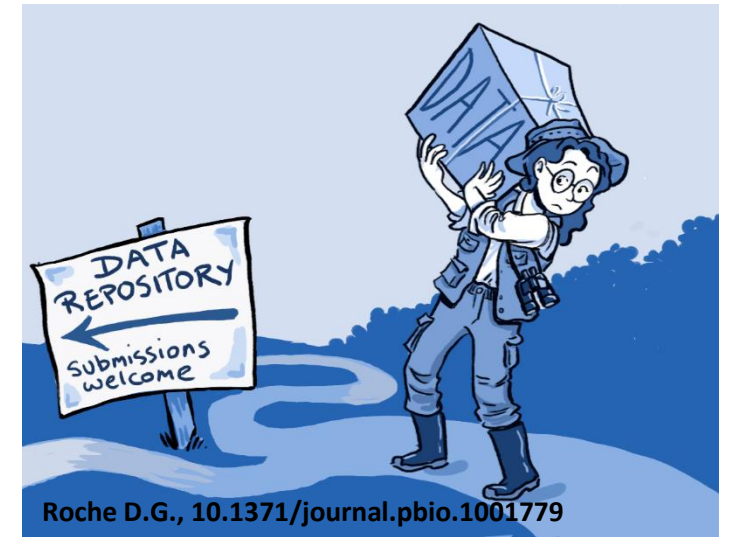
RESEARCH INTEGRITY
IS KEY

WHAT WE ARE ASKING

- Work to the highest standards following best practices
- Ensure that results are disseminated as quickly as possible
- Ensure transparency, integrity and reproducibility of research results
- Agree to the principles of the Research Integrity Concordat
- Commit to transparent trial results (even negative results must be published in a timely manner)

DATA SHARING

- Data sharing is a key part of the open science movement
- Data sharing means making raw research available in open data repositories and via controlled access
- Data is easily accessible, intelligible, reproducible, replicable, verifiable
- Sharing data is key to integrity, transparency and reproducibility



Data sharing can help Cancer Research UK achieve its ambition

AVOID RESEARCH WASTE

- Historically it is not common practice to share data
- This has hindered scientific progress
- Outdated research practices
- Duplication of efforts
- Inaccuracy and poor quality of experiments

Why Most Published Research Findings Are False

John P.A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for

factors that influence this problem and some corollaries thereof.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a p -value less than 0.05. Research is not most appropriately represented and summarized by p -values, but, unfortunately, there is a widespread notion that medical research articles

It can be proven that most claimed research findings are false.

is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is $R/(R+1)$. The probability of a study finding a true relationship reflects the power $1 - \beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate, α . Assuming that c relationships are being probed in the field, the expected values of the 2×2 table are given in Table 1. After a research

SUBSCRIBE

SCIENTIFIC
AMERICAN

English ▾ Cart  Sign In | Register

THE SCIENCES MIND HEALTH TECH SUSTAINABILITY EDUCATION VIDEO PODCASTS BLOGS STORE Q

POLICY & ETHICS

Is There a Reproducibility Crisis in Science?

By Nature Video on May 28, 2016



BBC

 Sign in



News

Sport

Weather

iPlayer

TV

Radio

NEWS

Home

UK

World

Business

Politics

Tech

Science

Health

Education

Entertainment

Science & Environment


Most scientists 'can't replicate studies by their peers'

‘Today researchers realize that they can't advance cancer progress alone... What's required extends beyond any individual or any individual discipline, beyond medicine itself... It requires somewhat of a change in mindset. It requires a lot more openness - open data, open collaboration and above all, open minds.’

Joe Biden, Former US Vice President, ASCO 2016

THE SPECIFIC RELEVANCE OF DATA SHARING TO CANCER RESEARCH


HUFFPOST COLLEGE


Edition: US ▾ 


FRONT PAGE POLITICS ENTERTAINMENT WHAT'S WORKING HEALTHY LIVING WORLDPOST HIGHLINE HUFFPOST LIVE

Leave your number to win a dream trip from London Gatwick. **Enter by 24 April**

T&Cs apply



AdChoices 

 Sign in  News Sport Weather iPlayer TV Rac

NEWS

Home UK World Business Politics Tech Science Health Education Enterta


Business Your Money Market Data Markets Companies Economy

THE BLOG

Can “Moon Shot” for Cancer Cure Overcome Data Sharing Problems?




🕒 04/18/2016 03:54 pm ET

Bloomberg the Company & Its Products ▾ | Bloomberg Anywhere Remote Login | Bloomberg Terminal Demo Request



 Markets Tech Pursuits Politics Opinion Businessweek

One-Third of Clinical Trial Results Never Disclosed, Study Finds

by Rebecca Spalding

February 17, 2016 – 11:30 PM GMT   

- ▶ Only 29% published in journal in two years, 13% on registry
- ▶ Dissemination varies widely among 51 academic medical centers

One-third of clinical trials conducted at 51 major U.S. universities and academic hospitals were never published in a peer-reviewed journal or in a government registry online, according to a new study in the BMJ, formerly the British Medical Journal.

POLITICS

Biden urges cancer researchers to work together toward ‘moonshot’ goals



MARK HUMPHREY/AP

Vice President Joe Biden says too many barriers remain to cancer research.

Could cures for cancer lie hidden in the cloud?

 The New York Times 

Politics SUBSCRIBE | LOG IN

President Weighs In on Data From Genes



President Obama speaking Thursday at a White House forum on a major biomedical research initiative he began last year, the Precision Medicine Initiative.

ZACH GIBSON / THE NEW YORK TIMES

By JULIE HIRSCHFELD DAVIS
FEBRUARY 25, 2016

WASHINGTON — President Obama on Thursday waded into the complex and high-stakes debate over whether patients own their genetic information, saying that he believes that his tissues and any discoveries that stem from his DNA belong to him.

 **CANCER RESEARCH UK**

WHAT TO EXPECT FROM THIS WORKSHOP

- ✓ Cancer Research UK's ambition and strategy
- ✓ Data sharing policy and implementation
- ✓ Cancer Research UK activities
- ✓ Incentives to increase data sharing
- ✓ How can we work together?

DRIVERS FOR CRUK'S DATA SHARING POLICY

Conceptual case

- Transparency, integrity and reproducibility
- Reducing duplication/waste, and getting maximum use from data

Real world drivers

- Demand from patients and public to better showcase what we fund
- Funding panel members' interest in data sharing
- Demand from researchers for CRUK data
- Journals requesting data to be shared with publications

Alignment with other funders

- Close working with MRC; EPSRC; ESRC; Wellcome
- Expert Advisory Group on Data Access
- Association of Medical Research Charities

DRIVERS FOR RESEARCHERS TO SHARE DATA

Markowetz *Genome Biology* (2015) 16:274
DOI 10.1186/s13059-015-0850-7



COMMENT

Open Access

Five selfish reasons to work reproducibly



Florian Markowetz

1. 'Helps to avoid disaster'
2. 'Makes it easier to write papers'
3. 'Helps reviewers to see it your way'
4. 'Enables continuity of work'
5. 'Helps to build reputation'

OUR POLICY ON DATA SHARING AND PRESERVATION

Issues	CRUK's position
Data outputs	All research data, including unpublished data, -ve results, code...
Costs	Appropriately justified costs can be included in grant proposals
Metadata	Community standards/information encouraged
Preservation	Data to be retained for at least 5 years after grant end
Discoverability	Highlight available data and communicate dataset content/outline modes of sharing
Timeframes	Limited period of exclusive use permitted where justified
Acknowledgement	Secondary users should credit original researcher and CRUK
Europe PMC	Electronic copies deposited in PubMed not later than 6 months after publication
Publishing routes	Encourage 'open' publication routes
License	Where an APC is paid, papers must have CC-BY licence

A FLEXIBLE APPROACH TO DATA SHARING

A number of factors may enable or limit the amount of data which can be shared

- ✓ Scientific discipline
- ✓ Infrastructure
- ✓ Value of data for other researchers
- ✓ Likely impact of the data on policy and practice
- ✓ Ethical and legal issues
 - ✓ Patient consent and confidentiality
 - ✓ Patents and protection of intellectual property
- ✓ Cultural challenges
 - Concerns around future use, and perceived loss of control
 - Limited formal academic credit mechanisms

WHAT TO EXPECT FROM THIS WORKSHOP

- ✓ Cancer Research UK's ambition and strategy
- ✓ Data sharing policy and implementation
- ✓ **Cancer Research UK activities**
- ✓ Incentives to increase data sharing
- ✓ How can we work together?

CONCORDAT ON OPEN RESEARCH DATA

- ✓ Open access to research data
- ✓ Good data management
- ✓ Data should be accessible discoverable and usable
- ✓ Restrictions to data sharing should be justifiable and justified
- ✓ Use of other data should conform to legal, ethical and regulatory frameworks
- ✓ Data supporting publication should be available and citable



CURRENT EFFORTS TO IMPROVE DATA SHARING

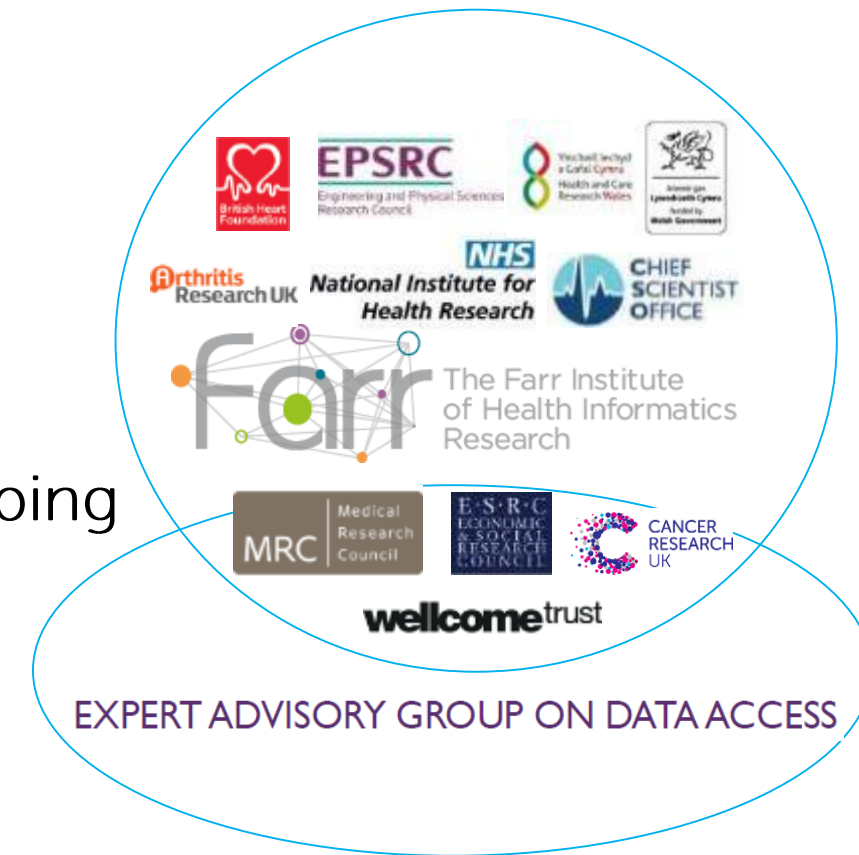
- ✓ Better data sharing plans from our main response-mode committee grantees
- ✓ Monitoring data sharing through grant review mechanisms
- ✓ Ongoing discussions around clinical trial data sharing
- ✓ Practical guidance on website to assist with data management and sharing

UPDATING OUR APPLICATION PROCESS

- Data management plans requirement
- Ask researchers to share all research outputs and not only publications – datasets, software, codes, protocols
- Allowing researchers to cite preprints in their grant applications
- Educate reviewers and committee members to assess research considering impact of all research outputs

COLLABORATIONS AND PARTNERSHIPS

- ✓ Convergence of policies with other funders around:
 - Data sharing
 - Data science
- ✓ Working with publishers on policies and impact metrics
- ✓ Working with institutions and researchers (developing working frameworks and incentives)
- ✓ Working with data/infrastructure providers (NHS Digital, repositories, PHE)



CURRENT CRUK DATA SHARING INITIATIVES

We participate in national and international initiatives to support data sharing in the biomedical science



EPIC study



THE MILLION WOMEN STUDY

WHAT TO EXPECT FROM THIS WORKSHOP

- ✓ Cancer Research UK's ambition and strategy
- ✓ Data sharing policy and implementation
- ✓ Incentives to increase data sharing
- ✓ How can we work together?

SAN FRANCISCO DECLARATION OF RESEARCH ASSESSMENT (DORA)



To address the need to improve the ways in which the output of scientific research is evaluated by funding agencies, academic institutions, and other parties

PRACTICES WE ARE ENDORSING

- Discourage the use of journal-based metrics as surrogate measures of the quality of individual articles, to assess an individual scientist contribution, or in hiring, promotion and funding decisions
- We will be more explicit about the criteria used in evaluating the scientific productivity of a grant
- Scientific content of a paper is what counts
- We consider the value and impact of all research outputs

POSSIBLE INCENTIVES AND REWARDS CURRENTLY DISCUSSED

- Consider a broad range of impact measures (e.g. qualitative indicators such as influence on policy and practice)
- Focus on a small number of publications and explain their significance
- Changing citation practices (awarding authorship credit to researchers who share their data)
- Develop a badging system to support 'open behaviour' (Open Access badge, Open Data badge, etc.)

WHAT TO EXPECT FROM THIS WORKSHOP

- ✓ Cancer Research UK's ambition and strategy
- ✓ Data sharing policy and implementation
- ✓ Incentives to increase data sharing
- ✓ How can we work together?

WORKING IN COLLABORATION

- Understand researchers' needs
- Feedback from researchers is important
- Stimulate dialogue with researchers and institutions

FUTURE PLANS

- ✓ Work closely with researchers and data managers – to understand the challenges and work through these collaboratively
- ✓ Monitor and improve compliance with our data sharing and open access policies.
- ✓ Increase transparency of CRUK-funded research outputs.
- ✓ Practical steps to support and reward data sharing.
- ✓ Improve awareness of how data can be used and how data sharing will benefit patients.
- ✓ Support skills development and training
- ✓ Increase international collaborations to improve data sharing practices.

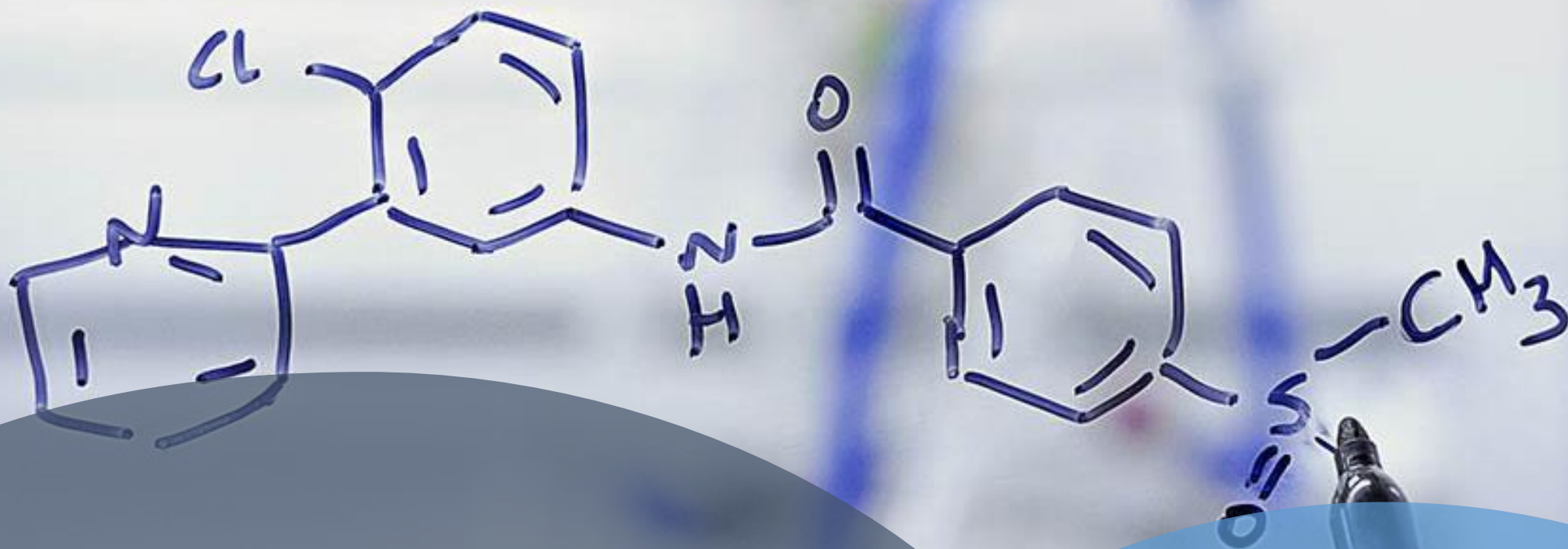
DRIVE A DATA SHARING CULTURE IN THE CANCER RESEARCH COMMUNITY

Paola Quattroni
Data Research Funding Manager
Cancer Research UK

Paola.Quattroni@cancer.org.uk

@PaolaQuattroni 





DATA MANAGEMENT PLANS

DATA MANAGEMENT PLANS

We require submission of data sharing plans at application stage

We provide a data management plan (template) for applicants to the Population Research Committee

We provide guidance to data sharing, links to repositories, and resources

PRACTICAL GUIDANCE

Practical guidance for researchers on writing data sharing plans



Download subject-specific guidance


- › [Discovery Research](#)
- › [Clinical Research](#)
- › [Population Research](#)

Share this page

 Facebook

 Google+

 Twitter

 Other

When you apply for a CRUK grant, we ask you to fill in a Data Sharing Plan. These are some of the issues you may wish to consider as you complete this Plan, and then as you seek to implement the planned activities in the course of your research.


This guidance should be read in conjunction with [our policy on data sharing](#), which sets out our high-level expectations with which grantholders must comply.

1. Why is data sharing important? 

2. Why does CRUK require a data sharing plan in applications? 

3. Where and when do I submit a data sharing plan? 

4. What should I discuss in my free text data sharing plan? 

5. Additional resources for data management and sharing 

A GOOD DATA MANAGEMENT PLAN

- Type of data and formats
- What data can be shared
- Management and storage of data
- Metadata Data preservation
- Data security and confidentiality
- Restrictions to data sharing – with explanation
- Whether data can be openly shared or upon request
- Governance and systems in place
- Ethical approvals

OTHER RESOURCES

CRUK Data sharing policy

http://www.cancerresearchuk.org/sites/default/files/data_sharing_policy.pdf

CRUK practical guidance

<http://www.cancerresearchuk.org/funding-for-researchers/applying-for-funding/practical-guidance-for-researchers-on-writing-data-sharing-plans>

CRUK Data sharing FAQs

<http://www.cancerresearchuk.org/funding-for-researchers/applying-for-funding/policies-that-affect-your-grant/submission-of-a-data-sharing-and-preservation-strategy/data-sharing-faqs>

UK Data Archive

<http://data-archive.ac.uk/create-manage/advice-training/training-resources>

Digital Curation Centre

<http://www.dcc.ac.uk/resources/data-management-plans>

Wellcome researchers survey on data sharing

https://figshare.com/articles/Survey_of_Wellcome_researchers_and_their_attitudes_to_open_research/4055448/1

DMPS EXAMPLES

- Science Committee
- Population Research Committee
- Clinical Research Committee



INTERESTED IN A
DMP WORKSHOP?

Thank you!

Paola.Quattroni@cancer.org.uk

@Paola Quattroni 

