

Initiative en matière de
données de recherche
Approche des trois organismes



CIHR IRSC

Canadian Institutes of Health Research
Instituts de recherche en santé du Canada

Matthew Lucas, PhD
Conseil de recherches en sciences humaines

SSHRC  CRSH

le 19 avril 2018

PLAN DE LA PRÉSENTATION

- Raison d'être de la gestion des données de recherche
- Élaboration de la Politique interorganisme sur la gestion des données
- Activités de mobilisation récentes

QU'EST-CE QU'UNE DONNÉE DE RECHERCHE?

Une donnée de recherche, c'est un contenu utilisé comme source principale au soutien d'une recherche, de travaux réalisés dans le cadre d'une bourse, d'une activité artistique ou d'une recherche-crédation.

Aussi, on utilise une donnée de recherche comme élément de preuve dans un processus de recherche; une telle donnée est communément acceptée par les chercheurs comme un instrument essentiel à la validation des conclusions et des résultats d'une recherche.

LA GESTION DES DONNÉES A POUR BUT D'ASSURER :

L'excellence en recherche

La diffusion de la recherche

L'impact de la recherche

Les pratiques exemplaires en recherche

EXCELLENCE EN MATIÈRE DE RECHERCHE

Gestion de projet

Soutien de la
reproductibilité

Évitement du travail en
double

Lignes directrices sur les
revues spécialisées



DIFFUSION DES RÉSULTATS DE RECHERCHE

Mise en commun des
données

Citations

Interdisciplinarité

Sharing Detailed Research Data Is Associated with Increased Citation Rate

Heather A. Piwowar*, Roger S. Day, Douglas B. Fridsma

Department of Biomedical Informatics, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, United States of America

Background. Sharing research data provides benefit to the general scientific community, but the benefit is less obvious for the investigator who makes his or her data available. **Principal Findings.** We examined the citation history of 85 cancer microarray clinical trial publications with respect to the availability of their data. The 48% of trials with publicly available microarray data received 85% of the aggregate citations. Publicly available data was significantly ($p=0.006$) associated with a 69% increase in citations, independently of journal impact factor, date of publication, and author country of origin using linear regression. **Significance.** This correlation between publicly available data and increased literature impact may further motivate investigators to share their detailed research data.

Citation: Piwowar HA, Day RS, Fridsma DB (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. PLoS ONE 2(3): e308. doi:10.1371/journal.pone.0000308

INTRODUCTION

Sharing information facilitates science. Publicly sharing detailed research data—sample attributes, clinical factors, patient outcomes, DNA sequences, raw mRNA microarray measurements—with other researchers allows these valuable resources to contribute far beyond their original analysis[1]. In addition to being used to confirm original results, raw data can be used to explore related or new hypotheses, particularly when combined with other publicly available data sets. Real data is indispensable when investigating and developing study methods, analysis techniques, and software implementations. The larger scientific community also benefits: sharing data encourages multiple perspectives, helps to identify errors, discourages fraud, is useful for training new researchers, and increases efficient use of funding and patient population resources by avoiding duplicate data collection.

Believing that that these benefits outweigh the costs of sharing research data, many initiatives actively encourage investigators to make their data available. Some journals, including the *PLoS* family, require the submission of detailed biomedical data to publicly available databases as a condition of publication[2–4]. Since 2003, the NIH has required a data sharing plan for all large funding grants. The growing open-access publishing movement will perhaps increase peer pressure to share data.

However, while the general research community benefits from shared data, much of the burden for sharing the data falls to the study investigator. Are there benefits for the investigators themselves?

A currency of value to many investigators is the number of times their publications are cited. Although limited as a proxy for the

RESULTS

We studied the citations of 85 cancer microarray clinical trials published between January 1999 and April 2003, as identified in a systematic review by Ntzani and Ioannidis[7] and listed in Supplementary Text S1. We found 41 of the 85 clinical trials (48%) made their microarray data publicly available on the internet. Most data sets were located on lab websites (28), with a few found on publisher websites (4), or within public databases (6 in the Stanford Microarray Database (SMD)[8], 6 in Gene Expression Omnibus (GEO)[9], 2 in ArrayExpress[10], 2 in the NCI GeneExpression Data Portal (GEDP)(gedp.nci.nih.gov); some datasets in more than one location). The internet locations of the datasets are listed in Supplementary Text S2. The majority of datasets were made available concurrently with the trial publication, as illustrated within the WayBackMachine internet archives (www.archive.org/web/web.php) for 25 of the datasets and mention of supplementary data within the trial publication itself for 10 of the remaining 16 datasets. As seen in Table 1, trials published in high impact journals, prior to 2001, or with US authors were more likely to share their data.

The cohort of 83 trials was cited an aggregate of 6239 times in 2004–2005 by 3133 distinct articles (median of 1.0 cohort citation per article, range 1–23). The 48% of trials which shared their data received a total of 5334 citations (85% of aggregate), distributed as shown in Figure 1.

Academic Editor: John Ioannidis, University of Ioannina School of Medicine, Greece

INCIDENCE DE LA RECHERCHE

Sur le domaine

Sur le contexte social

Sur les politiques

Data on bushfire fatalities helping to prevent more deaths

Who did the research?

Risk Frontiers (based at Macquarie University), RMIT University and Bushfire Cooperative Research Centre.

Melanoma study leads to ban on commercial sunbeds

What is the research about?

The **Melbourne Institute** at the University of Melbourne, funded by the Australian Government.

What is the project about?

The **Household, Income and Labour Dynamics in Australia (HILDA) survey** is an annual study of around 8000 households or around 20,000 individuals. It collects information about economic wellbeing, employment, health and family.

HILDA is a longitudinal study, meaning respondents are followed year-on-year to observe changes over time. The data is confidentialised to remove identifying

What is the real-life data impact?

The HILDA Survey is **known** as one of the best panel surveys in the world, collecting a broad range of high quality economic and social research data about Australian lives.

The data collected and analysed through HILDA have had a substantial impact on Australian policy development. Its **findings** have been used to help shape many policies, including:

- proposals in **Australia's Future Tax System Review**
- the Government's 2011 Paid Parental leave policy
- decisions about minimum wage rates



EverGraze project boosts Australian farming

Who did the research?

Charles Sturt University in partnership with **Future Farm Industries Cooperative Research Centre (FFCI CRC)** and state departments in Victoria, New South Wales and Western Australia.

What is the project about?

EverGraze is a collaborative research project testing the innovative use of perennial flora

on farmland in high rainfall zones of southern Australia. It has helped farmers improve soil health (e.g. acidity and salinity), ultimately increasing livestock profitability.

More than 250 scientists, advisers and farmers have contributed to EverGraze since its inception in 2002, including experts in soil science, agronomy, environmental science, hydrology and economics.

With complex modelling across multiple

30

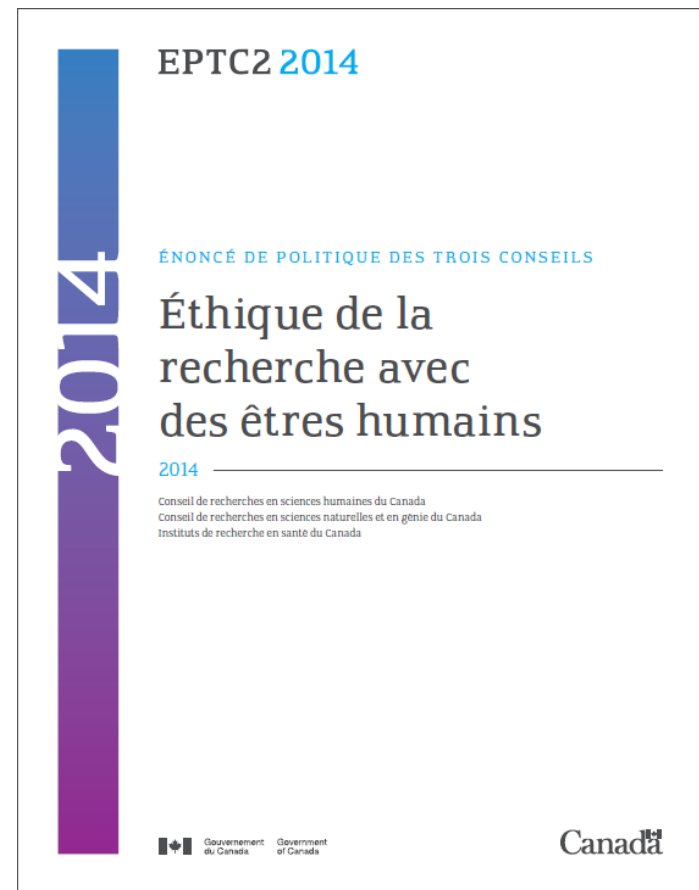
PRATIQUES EXEMPLAIRES DE RECHERCHE

Transparence

Confiance

Éthique

Utilisation responsable
des deniers publics



AVANCÉES À L'ÉCHELLE MONDIALE

Partenaires financiers en recherche

- National Science Foundation
- National Endowment for the Humanities - ODH
- National Institutes of Health
- UK Research Councils
- European Commission
- National Natural Science Foundation of China

Fondations et organismes de bienfaisance

- American Heart Association
- Bill and Melinda Gates Foundation
- Alfred P Sloan Foundation
- Gordon and Betty Moore Foundation
- Open Society Foundations
- Royal Society
- Wellcome Trust
- Cancer Research UK

Établissements de recherche





OCUL Ontario Council of University Libraries



Canada



Universities Canada. Universités Canada.

CAUL CBUA Council of Atlantic University Libraries
Conseil des bibliothèques universitaires de l'Atlantique

BCI BUREAU DE COOPERATION INTERUNIVERSITAIRE



THE LEADERSHIP COUNCIL LE CONSEIL DU LEADERSHIP

compute canada | calcul canada



Bibliothèques

Chercheurs

Études supérieures

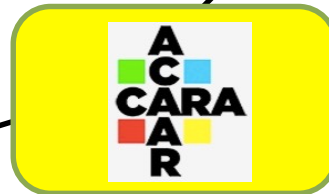
TI

Éthique

Service de la gestion de la recherche

canarie

RDC DRC Research Data Canada - Données de Recherche Canada



casrai

Offert par Chuck Humphrey

CIHR IRSC Canadian Institutes of Health Research / Instituts de recherche en santé du Canada

NSERC CRSNG

SSHRC CRSH

GOUVERNEMENT DU CANADA

DIRECTIVE SUR LE GOUVERNEMENT OUVERT



Table des matières

PRÉAMBULE DU PRÉSIDENT DU CONSEIL DU TRÉSOR DU CANADA 2

I. INTRODUCTION 3

 Se diriger vers un gouvernement ouvert et transparent 3

 Un effort conjoint 4

 Le Partenariat pour un gouvernement ouvert 4

II. RÉALISATIONS À CE JOUR 6

III. ÉLABORATION DU TROISIÈME PLAN BIENNUEL 8

IV. ENGAGEMENTS DU CANADA RELATIFS AU GOUVERNEMENT OUVERT (2016-2018) 9

 A. Ouvert par défaut 11

 Engagement 1 : Améliorer l'accès à l'information Error! Bookmark not defined.

 Engagement 2 : Simplifier les demandes de renseignements personnels 14

 Engagement 3 : Étendre et améliorer les données ouvertes 15

 Engagement 4 : Fournir et préserver l'information ouverte 17

 Engagement 5 : Définir une approche de mesure du rendement pour un gouvernement ouvert 18

 Engagement 6 : Développer les compétences relatives au gouvernement ouvert dans l'ensemble de la fonction publique fédérale 19

 Engagement 7 : Intégrer les exigences en matière de transparence dans la stratégie relative aux services fédéraux 20

 Engagement 8 : Améliorer l'accès aux collections culturelles et patrimoniales 21

 B. Transparence financière 22

 Engagement 9 : Améliorer l'ouverture de l'information sur les dépenses et l'approvisionnement au gouvernement 23

 Engagement 10 : Accroître la transparence des données budgétaires et des analyses économiques et financières 24

 Engagement 11 : Accroître la transparence sur le financement à l'aide de subventions et de contributions 25

 Engagement 12 : Améliorer l'information publique sur les sociétés canadiennes 27

 C. Innovation, prospérité et développement durable 28

 Engagement 13 : Accroître la disponibilité et l'utilité des données géospatiales 29

 Engagement 14 : Accroître l'ouverture des activités scientifiques fédérales (Science ouverte) 30

 Engagement 15 : Améliorer les renseignements sur les investissements en innovation au Canada (IOIK) 32

 Engagement 16 : Harmoniser les données ouvertes à l'échelle du Canada (Données ouvertes Canada) 33

 Engagement 17 : Mettre en œuvre la Loi sur les mesures de transparence du secteur extractif 35

 Engagement 18 : Appuyer les initiatives d'ouverture et de transparence dans le monde entier 36

 D. Engager les Canadiens, les Canadiennes et le monde 37

 Engagement 19 : Mobiliser la société civile sur les enjeux liés au gouvernement ouvert 38

 Engagement 20 : Permettre le dialogue ouvert et l'élaboration de politiques ouverte 39

 Engagement 21 : Promouvoir le gouvernement ouvert à l'échelle mondiale 41

 Engagement 22 : Mobiliser les Canadiens et les Canadiennes pour améliorer les principaux services de l'Agence du revenu du Canada 42

V. CONCLUSION 43

ANNEXE: Aperçu du troisième Plan biennal dans le cadre du Partenariat pour un Gouvernement Ouvert (2016-2018) 45

CONTEXTE DE L'ÉLABORATION DE LA POLITIQUE INTERORGANISME

2013

Tirer profit des données massives : vers un nouveau cadre stratégique pour l'évolution de la recherche numérique au Canada

2016

Déclaration de principe des trois organismes sur la gestion des données numériques

2017-2018

Projet de politique des trois organismes sur la gestion des données de recherche



Impact attendu d'une politique interorganismes de gestion des données

L'objectif des trois organismes subventionnaires est de favoriser la mise sur pied d'une culture de recherche selon laquelle :

- La gestion efficace des données est un élément déterminant d'excellence en recherche reconnu au sein de toutes les disciplines et fait partie du processus habituel d'une recherche;
- Un plus grand nombre d'ensembles de données canadiennes sont cités et valorisés en tant que produits de recherche pendant la période de validité, la promotion et les processus d'examen par les pairs;
- Les chercheurs du Canada sont équipés et prêts à participer à des projets de recherche de collaboration, où les exigences en gestion de données deviennent la norme;
- Les établissements de recherche du Canada sont prêts à soutenir la gestion des données que leurs chercheurs produisent;
- La capacité d'archiver, de trouver et de réutiliser de façon responsable les données de recherche est accrue, ce qui permet de faire de nouvelles découvertes et de créer l'innovation.

ÉBAUCHE DE POLITIQUE DES TROIS ORGANISMES SUR LA GESTION DES DONNÉES NUMÉRIQUES

- Aux fins de consultation
- La rétroaction permettra l'élaboration de la version finale de la politique
- La politique proposée comporte trois exigences éventuelles :
 1. Établissements : stratégie des établissements
 2. Chercheurs : plans de gestion des données
 3. Chercheurs : dépôt de données
- Mise en œuvre progressive

ÉBAUCHE DE POLITIQUE DES TROIS ORGANISMES SUR LA GESTION DES DONNÉES NUMÉRIQUES

1. Établissements : Stratégie des établissements

- Chaque établissement qui gère des fonds des trois organismes pourrait être tenu de créer une stratégie organisationnelle de gestion des données de recherche. La stratégie pourrait énoncer la façon dont l'établissement fournira à ses chercheurs un environnement qui permet et appuie des pratiques de gestion des données de recherche de niveau mondial.
- La stratégie pourrait être publiée et rendue disponible au public sur le site Web de l'établissement, avec les coordonnées d'une personne-ressource pour répondre aux questions posées concernant la stratégie.

ÉBAUCHE DE POLITIQUE DES TROIS ORGANISMES SUR LA GESTION DES DONNÉES NUMÉRIQUES

Pourquoi a-t-on besoin de stratégies institutionnelles?

- Pour reconnaître le rôle des établissements dans le soutien de la gestion des données
- Pour fournir une occasion aux établissements de se pencher sur les lacunes existantes et sur la façon de les éliminer dans une perspective à l'échelle du campus
- Pour peut-être aider les établissements à concevoir une approche qui fonctionne pour eux, tout en favorisant l'harmonisation et la collaboration avec les autres établissements
- Pour peut-être fournir de l'information aux organismes concernant la capacité de gestion des données
- Pour servir de fondation aux exigences qui pourraient suivre

Soutien

Modèle de stratégie institutionnelle du réseau Portage

ÉBAUCHE DE POLITIQUE DES TROIS ORGANISMES SUR LA GESTION DES DONNÉES NUMÉRIQUES

2. Chercheurs : Plans de gestion de données

- Les bénéficiaires de subventions pourraient être tenus de créer des plans de gestion de données (PGD) pour les projets de recherche appuyés totalement ou partiellement par les fonds des trois organismes. Comme condition de déblocage des fonds, les bénéficiaires des subventions pourraient soumettre ces plans au bureau de la recherche de leur établissement.
- Pour les occasions spéciales de financement, les organismes pourraient exiger que des PGD soient soumis à l'organisme approprié au moment de la demande; dans de tels cas, ils peuvent être pris en compte dans le processus de règlement des demandes.

ÉBAUCHE DE POLITIQUE DES TROIS ORGANISMES SUR LA GESTION DES DONNÉES NUMÉRIQUES

Pourquoi a-t-on besoin de plans de gestion des données?

- Parce qu'un tel plan représente une pratique exemplaire émergente à l'échelle mondiale.
- Parce qu'il est une excellente façon de permettre aux chercheurs d'entrevoir des possibilités à exploiter et des défis à relever dans la gestion de leurs données, bien avant que ces possibilités et ces défis ne se manifestent.
- Parce que selon les chercheurs, le processus de mise au point d'un plan de gestion des données les aide à améliorer leurs plans et leurs méthodologies de recherche.
- Parce qu'un plan de gestion des données peut favoriser la conduite responsable d'une recherche et le processus d'approbation de l'éthique de recherche.
- Parce qu'un tel plan aiderait à déterminer et à atténuer les problèmes liés à la propriété des données, à la possibilité d'échanger des données, etc.

Soutien

Assistant PGD du réseau Portage

ÉBAUCHE DE POLITIQUE DES TROIS ORGANISMES SUR LA GESTION DES DONNÉES NUMÉRIQUES

3. Chercheurs : Dépôt de données

- Pour l'ensemble des données de recherche et codes qui appuient les publications de revues, les prépublications et les autres produits de recherche qui découlent de la recherche soutenue par les organismes, les bénéficiaires de subventions pourraient être tenus de déposer ces données et codes dans un dépôt public approprié ou dans une autre plateforme qui assurera un stockage, une préservation et une conservation sûrs, et (s'il y a lieu) un accès aux données.

ÉBAUCHE DE POLITIQUE DES TROIS ORGANISMES SUR LA GESTION DES DONNÉES NUMÉRIQUES

Pourquoi a-t-on besoin d'un dépôt des données?

- Les méthodes, les attentes et la sécurité en ligne sont appelées à changer - l'entreposage des données dans un endroit sécuritaire permet de mieux assurer leur sécurité et de faire en sorte qu'elles pourront ultérieurement être utilisées par des créateurs
- Un dépôt des données aidera à utiliser les fonds publics de façon appropriée, ce qui autrement pourrait être incertain
- Un tel dépôt favorisera la reproductibilité des résultats
- L'échange de données sera favorisé

Soutien

Dépôt fédéré de données de recherche

ACTIVITÉS DE MOBILISATION RÉCENTES

- Entretiens avec un vaste éventail d'intervenants
 - Chercheurs, associations savantes et scientifiques, organisations de promotion et de soutien de la gestion des données, collègues d'organismes subventionnaires partout dans le monde
- Réunions régionales des intervenants
 - Réunions à Vancouver, à Calgary, à Toronto, à Montréal et à Halifax
 - On a constaté que les participants aux réunions étaient enthousiastes et optimistes par rapport aux possibilités qu'offre la gestion des données pour l'excellence en matière de recherche, mais ils ont aussi exprimé des préoccupations par rapport à certaines difficultés, notamment la sensibilisation des chercheurs, la capacité et la question du financement.

LA RÉTROACTION COMMUNAUTAIRE EST ESSENTIELLE

La rétroaction communautaire est essentielle pour orienter la conception finale de la politique et son mode de mise en œuvre.

Les organismes tiennent compte de l'ébauche de politique comme une proposition permettant de faire progresser la discussion avec les nombreux intervenants de la communauté de recherche, avec une politique des trois organismes comme produit final désiré.

Les agences prévoient de lancer une consultation en ligne pour faire progresser la discussion avec la communauté.

Merci!

Si vous avez d'autres questions,
veuillez les envoyer à:

ResearchData-Donneesderecherche@sshrc-crsh.gc.ca