



Bordeaux ● Limoges ● Montpellier ● Nîmes ● Toulouse

Le partage des données cliniques: un partage encouragé ?

Dr. Anne Cambon-Thomsen, DR CNRS
UMR 1027, Inserm et Université de
Toulouse III Paul Sabatier
& plateforme « Genotoul societal »

Colloque du club SMAC – Cancéropôle GSO « Statistiques et mathématiques appliquées au cancer »
Modélisation et simulation d'essais cliniques - Toulouse, 9-10 avril 2015

Outline

- Why is data sharing important in biomedical research?
- The changes in biomedical research regarding data
- History of data sharing policies (*ex: human genome, biobanks*)
 - Principles
 - Implementation
- Gaps and tools
- Conclusion

Outline

- Why is data sharing important in biomedical research?
- The changes in biomedical research regarding data
- History of data sharing policies (*ex: human genome, biobanks*)
 - Principles
 - Implementation
- Gaps and tools
- Conclusion

Why is sharing samples and data important?

- Enable research
- Build on existing resources
- Construct communities
- Optimise fund investment in research
- Ethical imperative:
 - true for research data, is it also true for clinical data?
- Policies in place (in light of open access)
- But... sharing is poorly done in practice!

Les fondements de la régulation

- Grandes classes: Données anonymes ou non
- Pour données identifiantes/sensibles (CNIL)
- Une législation européenne (Directive **95/46/CE, sur la protection des données personnelles**)
- en cours de modification pour devenir un règlement (texte initial en janvier 2012, plusieurs étapes passées, final prévu en 2015)

• Concerns about targeted research activities including personal sensitive data processing (health data, genetic, biometric, ethnic...):



- Scientific, statistical, historical health-related researches (*e.g. fundamental or applied public/private researches; clinical trials; retrospective researches...*)
- Biobanking (*human samples / DNA and RNA*)
- Registries (*systematic and continuous processing of medico-economical data e.g cancer, HIV, CVD*) and archived data
- Health-related ICTs

• Some topics of concerns:

- Informed consent (*systematicity / broadness*), and reuses of data/databases
- High public interest (*actionable waiver of data subjects' rights*)
- Balance between data subjects' rights and scientific legitimate processing (*e.g right to erasure / to be forgotten; portability etc.*)
- Measures/Tools for alleviating administrative burdens

Les données à caractère personnel / CNIL

- **Qu'est-ce qu'un fichier ?**
 - Tout ensemble structuré et stable de données à caractère personnel accessibles selon des critères déterminés.
- **Qu'est-ce qu'un traitement?**
 - Toute opération portant sur des données «quel que soit le procédé technique utilisé», tel que la collecte, l'enregistrement, l'organisation, la conservation, l'adaptation ou la modification, l'extraction, la consultation, l'utilisation, la transmission, l'interconnexion de données à caractère personnel,... »

Qu'est-ce qu'une donnée à caractère personnel ?

- *Toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres. »*

5 règles générales à respecter

- Respecter le principe de finalité et de proportionnalité
- Conserver les données pendant une durée limitée
- Assurer la sécurité et la confidentialité des données
- Informer la personne du traitement dont elle fait l'objet et lui permettre d'exercer ses droits d'accès, de modification, de suppression et d'opposition
- Déclarer le traitement auprès de la CNIL
- Si cadre de protocole de recherche biomédicale passant par un CPP (Comité de protection des personnes), procédure simplifiée (méthodologie de référence)

Conditions propres aux données sensibles: autorisation

- Les données sensibles sont celles qui font apparaître, directement ou indirectement, les origines raciales ou ethniques, les opinions politiques, philosophiques ou religieuses ou l'appartenance syndicale des personnes *ou qui sont relatives à la santé ou à la vie sexuelle de celles-ci*
- **Si le traitement comprend** des données génétiques, des données pénales ou comporte des appréciations sur les difficultés sociales des personnes, dans tous ces cas, le traitement doit faire l'objet d'une autorisation de la CNIL.

Pas de solution standard

- Adaptation nécessaire
- Revenir aux fondamentaux des droits des personnes
- Ne pas hésiter à demander de l'aide
- Ne pas sous-estimer le temps nécessaire à l'obtention des avis et autorisations dans son calendrier de recherche

Actualité: L'ouverture des données de santé

- **Le Point.fr (15/3/2014 - Quelles sont ces "données de santé" qui intéressent tant les défenseurs de la transparence et les acteurs de la sécurité sanitaire ?**
- **Anne Cousin** : La loi ne définit pas les données de santé. Mais les données dont nous parlons sont des données anonymes (donc non nominatives) rassemblées dans des bases de données comme le SNIIRAM (le Système national d'information inter-régimes de l'Assurance maladie). Il s'agit de l'âge, du sexe, des prestations remboursées et non remboursées, des nuits d'hôpital, des modes d'exercice des praticiens, etc. Leur accès est très restreint,....
- L'**Institut des données de santé (IDS)** préconise que des instituts de recherche tels que l'**Inserm** aient accès à l'intégralité de la base **SNIIRAM** et aussi que des industriels du secteur de la santé qui réalisent des études de santé publique validées par les pouvoirs publics aient accès à des extractions d'échantillons de données. Le but est de promouvoir la recherche et de mieux gérer le risque maladie.

Vote en cours de la loi de santé

- <http://www.assemblee-nationale.fr/14/dossiers/sante.asp>
- Extrait Compte-rendu Conseil des ministres du 15/10/14 (présentation M Touraine, Loi de santé)
- « *3. Innover pour conforter l'excellence du système de santé*

La politique de santé doit être innovante pour porter des réformes en profondeur. Parce que les enjeux de santé sont en constante évolution, le projet de loi organise un système national des données de santé. Il permettra l'ouverture (open data) des données publiques et un accès compatible avec le secret des données personnelles pour des recherches, projets d'étude et d'évaluation d'intérêt public.

Ce projet de loi rend concrètes les valeurs de justice et d'efficacité dans un projet politique innovant pour la santé des Français au quotidien. Il sera débattu à l'Assemblée nationale début 2015 et les premières mesures entreront en vigueur dès la promulgation de la loi. »

Why are such resource sharing important.... and poorly done?

- Much biomedical/epidemiological research is based on using such data (organised resources)
- Their access to all relevant researchers is essential
 - Promoting their **sharing** is crucial, but does not mean « just » putting files on the web!
 - It requires work.... Poorly recognised
 - There are today principles but few tools and ~ no incentive to that.

Outline

- Why is data sharing important in biomedical research?
- The changes in biomedical research regarding data
- History of data sharing policies (*ex: human genome, biobanks*)
 - Principles
 - Implementation
- Gaps and tools
- Conclusion

Take home message

- We have principles that encourage sharing
 - Instruments (declarations)
- We have explicit sharing policies in some institutions or funding agencies but mainly outside France
- There are gaps between policies and their implementation (*cf the translational valley of death*¹)
 - Lack of tools, incentives, recognition, conflicting regulations and values
- Some initiatives and a large movement
- Data sharing has become a topic of research too

¹MESLIN E., BLASIMME A., CAMBON-THOMSEN A. **Mapping the Translational Science Policy 'Valley of Death'**, *Clin Transl Med.* 2013 Jul 27;2(1):14.

International collaboration

“...The full benefits for which the subjects gave their samples will be realized through maximizing collaborative high quality research. Therefore there is an ethical imperative to promote access and exchange of information.”

Data Storage and DNA Banking for Biomedical Research: Technical, Social and Ethical Issues, ESHG, 2001, art. 17.

What does prevent sharing of (bio)resources ?

Technical aspects?

Protection of individuals?

Intellectual / institutional protection?

Lack of rewarding mechanisms?

An OECD text of reference

- In 2004, the ministries of science and technology asked OECD to define guidelines based on agreed principles in order to facilitate access to digitalised data issued from publicly funded research.
- **OECD Principles and guidelines for access to data from publicly funded research (2007)**

<http://www.oecd.org/dataoecd/9/60/38500823.pdf>

Retenir : ce que l'on produit en recherche est destiné à être partagé

Enjeu : combiner la protection des personnes et la dynamique de recherche

“OECD Recommendations set out collective and precise standards or objectives which the member countries are expected to implement.

A “Recommendation” is a legal instrument of the OECD that is not legally binding but through a long standing practice of the member countries, is considered to have a great moral force.

Recommendations of the OECD are adopted when member governments are prepared to make a political commitment to implement the principles (and/or guidelines) set out therein.

This type of instrument is often referred to as “soft law”.”

Principles

- *A. Openness*
- *B. Flexibility*
- *C. Transparency*
- *D. Legal conformity*
- *E. Protection of intellectual property*
- *F. Formal responsibility*
- *G. Professionalism*
- *H. Interoperability*
- *I. Quality*
- *J. Security*
- *K. Efficiency*
- *L. Accountability*
- *M. Sustainability*

Le domaine de la génomique pilote

- De 1996 à 2014, ce domaine « tire » vers le partage
- Déclarations de principes, Chartes
- Droits et devoirs des différentes parties prenantes

The landscape of data sharing

- 2011: Sharing research data to improve public health: joint statement of purpose

<http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-and-epidemiology/WTDV030690.htm>

(+ Comment in Lancet, 10/1/11) – **Signé par Inserm**

- 2011: A data sharing Code of Conduct for international genomic research
- 2012: **The tension between data sharing and the protection of privacy in genomics research.**
Annu Rev Genomics Hum Genet 13, 415 (2012)
- 2014, dans le cadre de Global alliance
<http://genomicsandhealth.org/>
 - **Framework for Responsible Sharing of Genomic and Health-Related Data**

Sharing research data to improve public health : the statement (extracts)

- **Vision**

We, as funders of health research, intend to work together to increase the availability to the scientific community of the research data we fund that is collected 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:** Any approach to the sharing of data should recognise 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:** All data sharing 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:** Any approach to data sharing 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.

Core Elements for Responsible Data Sharing

- Transparency
- Accountability
- Engagement
- Data Quality and Security
- Privacy, Data Protection and Confidentiality
- Risk-Benefit Analysis
- Recognition and Attribution
- Sustainability
- Education and Training
- Accessibility and Dissemination

Implementation

- Has occurred in the « genome world »
 - Strong commitment of scientific community
 - Scientific necessity
 - The « history » of gene patenting
 - Large public investment
 - A paradox?
- More difficult in other domain of health data

Different models of data access policy

- Strong restriction of access : who decides?
- Open access : question of pressure, quality?
- Sharing, collaboration : what incentives?

Evolution :

“My data, samples, families”....

“My very used collection”

“The collection I collaborated to...that is widely used”

Large scale pushes to collaboration; but what pushes to data sharing?

Some examples

- [http://www.nlm.nih.gov/NIHbmic/nih data sharing policies.html](http://www.nlm.nih.gov/NIHbmic/nih_data_sharing_policies.html)
- <http://www.bbsrc.ac.uk/datasharing>
- <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/>
- <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-and-epidemiology/WTDV030690.htm>

Outline

- Why is data sharing important in biomedical research?
- The changes in biomedical research regarding data
- History of data sharing policies (*ex: human genome, biobanks*)
 - Principles
 - Implementation
- Difficulties, gaps and tools
- Conclusion

What are the blocking factors of the shared use of resources

- Technical ones (quality etc.)
- Institutional ones
 - No exchange, sharing or access policy
 - No practical help
 - Cost
- Intellectual property ones: organised data, databases
- Information ones
 - No easily available information on content
 - Restricted use to pre-defined professional circles
- Ethical/legal restrictions (data protection)
- No obvious positive spin off : incentives?

Some examples of issues encountered

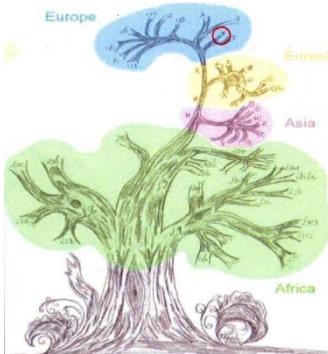
- Papers demonstrating the « virtuality » of anonymization
- The buzz of identifiability

• *Homer et al., PLoS Genet 2008 4(8): e1000167.*
doi:10.1371/journal.pgen.1000167 : “**Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays.**”

Science 18 January 2013: Vol. 339 no. 6117 pp. 321-324 - *Report + Interview + 39 p Supp. Material.*

Identifying Personal Genomes by Surname Inference

Melissa Gymrek^{1,2,3,4}, Amy L. McGuire⁵, David Golan⁶, Eran Halperin^{7,8,9}, Yaniv Erlich¹



Identifying Personal Genomes by Surname Inference

Melissa Gymrek,^{1,2,3,4} Amy L. McGuire,⁵ David Golan,⁶ Eran Halperin,^{7,8,9} Yaniv Erlich^{1*}

Sharing sequencing data sets without identifiers has become a common practice in genomics. Here, we report that surnames can be recovered from personal genomes by profiling short tandem repeats on the Y chromosome (Y-STRs) and querying recreational genetic genealogy databases. We show that a combination of a surname with other types of metadata, such as age and state, can be used to triangulate the identity of the target. A key feature of this technique is that it entirely relies on free, publicly accessible Internet resources. We quantitatively analyze the probability of identification for U.S. males. We further demonstrate the feasibility of this technique by tracing back with high probability the identities of multiple participants in public sequencing projects.

Incentives?

- If one wants to promote sharing it is necessary to recognise and reward the work necessary to make the resources/data « sharable »
- A lot of work not taken into account in research(er) evaluation
- Good intentions, ~no tools

An example of initiative on rewarding mechanisms

- BRIF : Bioresource research impact factor

CONSTAT:

Alors que les bioressources sont des éléments clés de la Recherche biomédicale, leur rôle est sous-estimé et l'effort fourni pour leur mise en place et pour leur mise à disposition mal reconnu.

Pourquoi?

BR pas assez visibles; difficile d'évaluer leur utilité; mal remerciées/citées

Cambon-Thomsen et al. *Nat Genet* 2003, 34:25–26

BIORESOURCES (BR)

- Biobanques
- Bases de données (sans échantillons biologiques)
- Autres collections d'outils biomoléculaires ou bioinformatiques ...
- ...

- Manque d'indicateurs accessibles décrivant un usage et un management efficaces
- Manque de système d'identification pour permettre leur suivi de façon fiable et automatisée
- Manque de standards de citation dans la littérature scientifique



En 2014 – 2015:
La recommandation CoBRA
(Citation of BioResources
in journal Articles)

<http://www.biomedcentral.com/1741-7015/13/33>

Anne Cambon-Thomsen (*dir.*), DREM CNRS

Laurence Mabile (*BRIF manager*) , IR UPS

UMR 1027, Inserm, Université Toulouse III - Paul Sabatier

Sous-Groupes de travail

‘BRIF & Digital Identifiers’

co-chaired by **G. A. Thorisson**, University of Leicester, UK and **P.A. Gourraud**, University of California SF, USA
Pierreantoine.Gourraud@ucsf.edu

‘BRIF Parameters’

chaired by **B. Parodi**, National Inst. Cancer Res. Genoa, IT

barbara.parodi@istge.it

‘BRIF and Journal Editors’

co-chaired by **A. Cambon-Thomsen**, Inserm UMR1027, Tlse, FR and
E. Bravo, Istituto Superiore di Sanità, Rome, IT

anne.cambon-thomsen@univ-tlse3.fr, elena.bravo@iss.it

‘BRIF in Access & Sharing Policies’

co-chaired by **E. Rial-Sebbag**, Inserm UMR1027, Toulouse, FR and **J. Harris**, Norwegian Institute of Public Health, Oslo, Norway
emmanuelle.rial@univ-tlse3.fr, Jennifer.Harris@fhi.no

‘BRIF dissemination’

chaired by **L. Mabile**, Inserm UMR1027, Tlse, FR

[laurence.mobile@univ-tlse3.fr](mailto:laurence.mabile@univ-tlse3.fr)

Collaboration avec Ubiquity Press

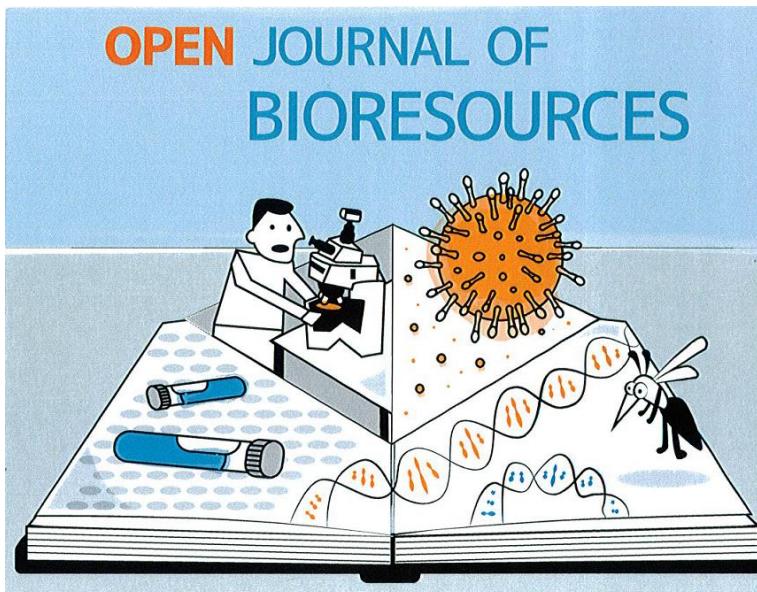
pour le lancement d'un journal en accès-libre dédié à la publication de "description méthodologique de BR"

'Open Journal of Bioresources'

<http://openbioresources.metajnl.com/>

But:

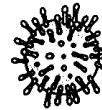
Augmenter la visibilité des BR en leur offrant une possibilité de "marker paper" méthodologique en accès-libre, sur la base d'une matrice d'article destinée aux auteurs et aux reviewers afin de recueillir les métadonnées nécessaire à la description précise de la BR



OPEN JOURNAL OF BIORESOURCES

UPmetajournals

OJB features peer reviewed short papers helping researchers to locate and cite bioresources with high **reuse** potential.



Making bioresources more openly discoverable has enormous benefits not only for the research community and the **wider public**, but for the producers of the bioresources as well. Both the resources and the OJB papers are **citable** and this is tracked to provide authors with metrics on reuse and **impact**.

We work with major biobanks and other repositories to ensure that the associated bioresources are professionally preserved, and **accessible**.



Accepting submissions now:
<http://openbioresources.metajnl.com>

[u] Ubiquity Press
<http://www.ubiquitypress.com>

(3) Bioresource
description

(1) Bioresource
Overview

(2) Methods

(4) Reuse
potential

(5) Copyright Notice

STANDARD BIORESOURCE CITATION

BRIF workshop in Rome, June 2013

Poster Presentation in Chicago at Peer review congress September 2013

Publication in 2015 in BMC Medicine

BMC Medicine, 17 Feb 2014 : <http://www.biomedcentral.com/1741-7015/13/33>

Developing a guideline to standardize the citation of bioresources in journal articles (CoBRA)

Elena Bravo^{1§}, Alessia Calzolari¹, Paola De Castro¹, Laurence Mabile², Federica

Napolitani¹, Anna Maria Rossi¹ and Anne Cambon-Thomsen²

¹Istituto Superiore di Sanità, Rome, Italy

²UMR U 1027, Inserm, Université Toulouse III - Paul Sabatier, Toulouse, France

Comment l'implémenter?

- Adoption et diffusion par des institutions
- Conseiller de l'utiliser dans les publications
- Dissémination via des sociétés scientifiques et professionnelles, des infrastructures
- Implémentation via les instructions aux auteurs d'éditeurs variés
- Sensibiliser les étudiants aux ressources utilisées dans les travaux

Conclusion

- Les données cliniques : place particulière au sein des données – régulation
- L'anonymat : aujourd’hui pas absolu quand accumulation de données – autres moyens de régulation
- Accéder /manager les données pour sa recherche
- Point de mire : ce que l’on produit va aussi se partager
- Le travail de « rendre partageable » doit être reconnu comme contribution à la connaissance

Merci de votre attention