Targeted format: 2500 words, 30 ref

Title: The BRIF (Bioresource Research Impact Factor) as a tool for improving bioresource sharing in biomedical research

Authors: Laurence Mabile, Raymond Dalgleish, Gudmundur A. Thorisson, Mylène Deschênes, Robert Hewitt, Jane Carpenter, Elena Bravo, Pierre Antoine Gourraud, Paul Hofman, Jennifer Harris, Mª Angeles Munoz-Fernandez and Anne Cambon-Thomsen as members of the BRIF working group.

Table of Contents

Introduction

Definition of bioresources and their importance in biomedical research

An increasing portion of biomedical research relies on the use of bioresources [1], many of which are being already embedded in medical practice. Bioresources and a number of related objects need to be reminded (box1). These resources are increasingly important in the modern era of "omics" medicine [2]: high throughput 'omics' platforms require material from and generate data on large numbers of both patients and/or healthy individuals. It is difficult or impossible, to obtain to do this on an individual basis, and this need is being met by contemporary bioresource facilities. Biobanks used for research aim at providing reliable and sustainable solutions with respect to several key issues, in order to provide high-quality material based on optimized procedures for the various steps, from acquisition, processing and storage to transfer procedures [3]. One of those issues concerns measuring the actual use of bioresources.

As more long term clinical and outcome data are generated, the value of a bioresource tends to increase exponentially with time. Indeed, data accrual is a major activity of many bioresource projects. Also, as research progresses, projects supplied with material will return research results (which can include both raw and curated data) to the resource and thus enhance the data set available; indeed the data alone may be sufficient to mine and answer research questions. Excellent examples are: the Framingham cohort [], variant-

frequency data in the dbSNP database [4] which are derived from the analysis of panels of DNAs from specific human populations; Genbank [5], OMIM [6], CEPH families [7]…..

A commitment to share this information with the research community is paramount [8,9]. The 2011 joint statement of 17 major national funders sent a powerful signal that research resources must be shared to maximize the potential of publicly funded resources [4,10,11]. However, an open-access policy for research data needs to be backed up by appropriate measures to enable the proper recognition for research resources being used in the development of novel scientific knowledge. Failure to do so will have a major negative impact on the global sharing vision which funders hope to encourage. Whilst promoting measures to improve access to biobanks and databases, we must develop policies mandating end-users to recognize and acknowledge the provenance of these resources. An appropriate set of tools is needed to implement such policies. Some tools currently exist, but an insufficient level of coordination and systematic implementation makes it difficult to see their positive impact on the overall organization of scientific activities.

What are the obstacles for recognition of the work involved in setting up and maintaining bioresources?

Establishing a valuable bioresource requires considerable time and effort. To provide appropriate rewards and recognition, it is necessary to be able to measure the performance of bioresources and there are various ways to do so, using a range of indicators (see Box 2) including management indicators showing that the bioresource is efficiently run and well utilized, indicators of the quality of the bioresource at various levels (biological samples, annotations and associated data, and search tools), of quantity available and of value of the samples or datasets, and indicators of research productivity.
Since the purpose of the bioresources considered here is to enhance research productivity, this last indicator may provide the most reliable assessment of an appropriate use. However, each current indicator has serious drawbacks, as outlined in box 3. The process of tracking publications and quantifying their impact is not straightforward. To track the publications made possible by a bioresource, it is essential that researchers consistently acknowledge use of the bioresource by placing a unique and traceable citation in all their relevant publications in a defined section of articles. To some extent this is already possible, *i.e.* if researchers acknowledge the bioresource in their articles or refers to bioresources's publications and if the effort is made to search through publications to find appropriate acknowledgements. To optimise this process and standardise it, it needs to be automated and based on systematic use of traceable and unique resource identifiers to support a routine and widely accepted method of assessment.

What are the obstacles for recognition of the work involved in setting up and maintaining databases?

A database differs in many ways from other types of bioresource such as a biospecimen bank. The tangible commodity in a database is information, as opposed to the samples, cell lines and DNA found in biological samples biobanks. The existence of a physical commodity affords biobanks the ability to control the distribution of their resources. This facilitates the later collection of data relating to the impact of the distributed items. By comparison, databases deal solely in information, with the traditional expectation on the part of the end-user that access to the data will be free. A notable exception to the free model of data distribution is BIOBASE [12] which charges for the distribution of its own and

that of third parties (such as HGMD Professional) with whom it has commercial agreements.
Within the framework of the free-access expectation in which they operate, databases have only very limited scope to measure their impact. It is feasible to log visits or "hits" to web pages and to broadly identify from where the requests are being made, but it is much more difficult to obtain detailed information. Because no contract exists with users, it is impossible to impose requirements on users with respect to how they are expected to acknowledge their use of the data in publications. However, imposition of a "contract" as rudimentary as the need to register to access data would only amount to being a crude control measure and would probably be a disincentive for many users.
Even when authors make an honest attempt to properly acknowledge a database, the results can be patchy. Publication of the URL for a database in biomedical papers provides some evidence of reuse, but if it is not present in the abstract the URL will not be indexed in PubMed or other bibliographic databases. Hence there is a need to establish a clear policy on the part of publishers concerning the citation and referencing of database. Otherwise, measuring the impact of open-access online database resources will remain an imprecise process.

The BRIF concept and objectives

To address the above issues, the Bioresource Research Impact Factor (BRIF) was introduced in 2003 [13], and later further developed [14,15]. The central aim of the BRIF is to construct a quantitative parameter to evaluate bioresources, modeled to some degree on the Journal Impact Factor (JIF) [16], and to provide guidance and methodology for optimising recognition of bioresources, their use and their sharing at international level. This would make it possible to document: i) the quantitative use of a bioresource, ii) the type and the importance of research results involving it, and iii) the scientific and management efforts of those who established and made available a bioresource, and provide recognition for their institution. Such a framework could be used much more rationally than "reputation" for evaluating bioresource activities over time. When taken into account in assessing researchers/contributors' professional results, this would increase the quality and sharing of bioresources.
To implement this concept an international working group has been set up. The working group comprises 123 members from 22 countries. Specific tasks have been assigned to several sub-groups as presented below.

Digital identifier schemes

To address the various issues referred to above concerning identification, bioresources need to be assigned actionable digital identifiers or IDs [17]. In order to fulfil the requirements of the scholarly record, bioresource ID should be persistent, globally unique and citable. The ID subgroup focuses on exploring and assessing existing and emerging technical solutions suitable for bioresource identification, as well as addressing key related questions such as what to identify (biobank projects, sample collections, databases, datasets) and which international and independent body or bodies should be responsible for assigning bioresource ID.

It is worth emphasizing here that the aim is not to create a new identifier scheme specifically for bioresources. Rather, the aim is to identify frameworks which are already established (or well on their way to becoming so) and to subsequently recommend their use as appropriate with respect to: i) resource providers (*e.g.* what type of IDs to use for biobank projects, clinical trial registering system or other systems [18]); ii) end users (e.g. guidelines from journal editors to authors about how to properly cite biobank projects and databases using unique ID).
Preliminary conclusions of the sub-group are that the field is already moving in the right direction in several areas. Notably, the DataCite initiative [19] has established a worldwide data registration agency which reuses and extends the Digital Object Identifier (DOI) scheme already widely used in the scholarly publishing domain [20,21]. Digital records describing bioresources (*e.g.* sample collections), or borne-digital resources (datasets associated with a resource or generated by its use) could well be assigned data DOIs. This could be an important step towards treating these important research outputs as first-class contributions to science [22]. DOIs could then serve as first-step identifiers to be used in the BRIF assessment.

A highly-related issue is the current lack of global infrastructure for identifying and attributing researchers who contribute to bioresources. The centralized contributor ID system now being built by the international ORCID initiative [23] will be launched mid-2012. It seems reasonable to assume that this infrastructure will be at the heart of any attribution schemes and therefore very relevant for BRIF in the near future.

Parameters, measures and indicators

To identify the different parameters to take into account when calculating the BRIF, it has been decided to focus on two types of entity providing a service to the scientific community: i) biobanks of human biomaterials and ii) databases of information relating to human subjects research. The aim is to provide a measure of the extent to which bioresources contribute to research. Downstream effects on healthcare and the economy will not be assessed. The parameters need to be objective and easily verifiable, and the calculation of a BRIF needs to be as simple as possible. A wide range of parameters are being considered for inclusion, including some indicators of bioresource quality, value, efficiency and research productivity as noted above. The most obvious is a simple metric based on citation counts and the traditional notion of journal-level impact. More sophisticated factors which incorporate measurements of bioresource quality, value and efficiency can also be devised (see Box 4).

Journal guidelines for resource citing and referencing

A key element for assessing the use and the research impact of bioresources is their systematic citation in journal articles. However, we are far from having standards and guidelines for the citation of such resources. A specific task is a proposal for sensitizing journal editors to BRIF issues and modifying their editorial guidelines accordingly**.** The necessity of recognition by journal editors of the need to properly acknowledge the bioresources used, using proper terminology and/or identifiers and agreeing on standards of citation (format/marker paper, location(s), institutions, people, *etc.*) have been extensively discussed. As “Uniform Requirements for Manuscripts Submitted to Biomedical Journals” [24] already exist, a proposal has been sent to the International Committee of

Medical Journal Editors to be considered in a future amendment of these requirements. This proposal highlights some additional requirements (Box 5) that may be needed to address the editorial problems concerning bioresources. Furthermore, additional actions have been planned to sensitize other committees and institutions concerned with editorial and ethical issues.

Policies for resource access and sharing

Attempting to measure the impact of a bioresource supposes, as its premise, that the research resource is actually being used. Use of a biobank (or research database) is contingent upon many factors, but the access and sharing policies certainly play a major role in facilitating or hindering use. Various components such as the level of constraints imposed on users or the level of simplicity/complexity of the procedures to gain access are pivotal in creating an environment that will stimulate or discourage use of a given bioresource.

Sharing, access and publication policies, and the agreements that support the 'transaction' of sharing material or data, appear as the appropriate vehicles to consider in implementing enforceable means of measuring the impact of a bioresource [25]. Through such guidelines or contracts, a bioresource can impose requirements on eventual users requirements to empower itself to measure its impact. Let us consider two categories of tools that are likely to contribute to the capacity of a bioresource to measure its impact: the dissemination and the 'control' measures.

Publications, academic presentations and other less traditional means of disseminating research results have a prime importance in measuring the impact of a scientific contribution. Bioresources must thus ensure that users will recognize the resources that were used in whatever medium of dissemination the researchers choose to communicate the results to the scientific community and the public. This recognition must be written in such a way as to allow for a systematic search to track uses as described above. Bioresources might also consider requiring the users to report on their use (*e.g.* sending their publication or a summary report) back to its source. However, a balance must be struck between imposing a series of requirements on users and on bioresource managers and still maintaining conditions that foster resource use.

Another element is the level of control that the bioresource can exercise over the secondary use of its material/data. If a bioresource hopes to keep track of the use of its content, it must ensure that all eventual users will comply with its dissemination requirements. This is particularly a challenge for research databases where the data can be copied and circulated easily and *ad infinitum*. In a context where international collaboration is increasing and pooling of research resources is necessary to conduct research on complex diseases and health, it is difficult for the bioresource to keep track of all eventual users. The identity of the source of a material may be lost in the chain of multiple exchanges and amalgamation with others. A bioresource can thus require that users do not share with third parties the material/data. Under such circumstances, it is expected that eventual users will have to deal directly with the initial bioresource provider to gain access, and will thus have the same requirements imposed upon them to recognize the original resources. However, here again, a balance must be struck between imposing constraints on users and making use of the bioresources appealing. Even if the correct balance is struck, there are specific issues relating to databases where no physical entity is necessarily provided. To some extent, commercial data providers can impose constraints on the onward distribution of the data. A breach of corresponding terms and conditions might then allow the data provider to restrict future access. Among those who provide free access to their data some large organisations do can support a number of

actions [25]. Smal data providers, for example, curators of LSDBs for a small number of genes have fewer opportunities to exert access control and simply rely on database copyright protection [26]. Given the delicate balance required between stimulating uses and supporting the capacity to measure the impact, the BRIF sub-group proposes to further study and develop an appropriate set of tools that could eventually be integrated in the overall access and sharing policies of the bioresources.

Conclusion

There is definite pressure to develop a chart of principles, tools and guidelines for bioresources management, uses and referencing on which the medical and scientific community could rely for their research practice. However this will not become tangible unless proper tools and frameworks for recognising such activities are first in place. This forum article provides the foundations for such policies and hopes to stimulate discussion among relevant stakeholders.

Box 1: Definitions.

Biospecimen.: A quantity of tissue, blood, urine, or other human-derived material. A biospecimen can comprise subcellular structures, cells, tissue (e.g. bone, muscle, connective tissue, and skin), organs (e.g., liver, bladder, heart, and kidney), blood, gametes (sperm and ova), embryos, fetal tissue, and waste (urine, feces, sweat, hair and nail clippings, shed epithelial cells, and placenta). Portions or aliquots of a biospecimen are referred to as samples (*NCI*** *Best Practices* working definition).

Database: An organized set of data or collection of files that can be used for a specified purpose (definition from *A dictionary of Epidemiology 4th Ed. by J.M. Last*)

Biorepository. An organization, place, room, or container (a physical entity) where biospecimens are stored (*NCI*** *Best Practices* working definition).

Biological resource centres: consist of service providers and repositories of the living cells, genomes of organisms, and information relating to heredity and the functions of biological systems. BRCs contain collections of culturable organisms (e.g. genomes, plasmids, viruses, cDNAs), viable but not yet culturable organisms cells and tissues, as well as databases containing molecular, physiological and structural information relevant to these collections and related bioinformatics…

Biospecimen resource. A collection of human specimens and associated data for research purposes, the physical entity in which the collection is stored, and all associated processes and policies. Biospecimen resources vary considerably, ranging from formal institutions to informal collections in a researcher's freezer (*NCI Best Practices* working definition).

Bioresources : include both biological samples with associated data (medical/epidemiological, social), databases independent of physical samples and other biomolecular and bioinformatics research tools (*BRIF group* working definition).

*Organisation for Economic Co-operation and Development
** From National Cancer Institute Best Practices for Biospecimen Resources

Box 2: Examples of indicators describing an efficient bioresource management and use.

►*Indicators of management:*

- Number of projects supported per year
- Number of biospecimens entering in the biobank / number of biospecimens used for developing a research project by year
- Sustainable maintenance

►*Indicators of quality*:

- Participation in external assessment programmes such as certification or accreditation
- Performing a morphological control of frozen specimens used for "omics" programme (biobanks)
- Assessment of the extent and richness of the datasets collected (including follow up of patients and treatments when relevant)

►*Indicators of research productivity*:

- Counting and measuring
    - The cumulated impact of publications that result from research supported by the bioresource
    - The number of patents that result from the use of the bioresource
    - The number of material (data) transfer agreements and contracts signed per year

Box 3: Current key elements impeding proper tracking of bioresources use in scientific literature.

Difficulties related to identification and acknowledgement of bioresources:

- multiplicity of sections where bioresources can be acknowledged (Material & Methods, Acknowledgements, References…)
- bioresource acknowledgement or citation placed outside the title or abstract in the main paper (or in online supplementary materials) and can therefore only be detected via full-text mining and is not indexed in Pubmed or Web of Science
- typing errors or approximation of the bioresource name/identification
- multiplicity of names for a given bioresource
- acknowledgement of persons instead of the bioresource itself
- absence of acknowledgement for the bioresource used (negligence)
- no standardized way to incentivise researchers to acknowledge properly the bioresource used
- ….

Difficulties encountered with marker papers*:
-Suitable to refer to one type of bioresource but not for any derived, or secondary bioresources

***Peterson & Campbell** *Nature Genetics* 42 (11), 919 (2010).

Box 4: Possible BRIF parameters to take into consideration.

First-line parameters	*Second-line parameters*
a/ Indicators of research productivity: - Number of articles citing the bioresource itself or the staff - The quality of the journal (impact factor...)	-Grants obtained by the users of the bioresource or to support the bioresource - Patents/licenses based on research supported by the bioresource - Economic impact
b/ Indicators of quality -Richness of data associated with samples -Compliance with data reporting nomenclatures and sharing standards	-Official recognition from Regional/National Health Institution
c/ Indicators of high value - Depth of data associated with samples - Rare disease samples or data - Existence of a quality control policy for samples and data	
d/ Indicators of usage - Number of requests filed per year (to be balanced with the type of resource) - Number of web page accesses per year for data resources	- Number of samples received and distributed per year - Number of material/data transfer agreements - Number of contracts or conventions
e/ Indicators of ethical standards - General policies of transparency, access rules, use of appropriate MTAs/DTAs - Consent forms - Data protection measures	
f/ Indicators of biobank efficiency - Turnaround time for requests - Time to include new data	
	Other factors - Age of bioresource - Size of bioresource - Return of research policy - impact of data cost on inclination to correctly cite the source of data - Past achievements of the bioresource…

Box 5*: Main suggestions for the *Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication (www.icmje.org):

in II. ETHICAL CONSIDERATIONS IN THE CONDUCT AND REPORTING OF RESEARCH

II.A Authorship and contributorship

II.A.2 Contributors listed in acknowledgements:

Proposition**: 'Biobankers should always be acknowledged for their contribution in providing "bioresources" useful for the conduct of the study. The name of the biobank (and identifier, if available) should also be reported here in full.'**

in IV. MANUSCRIPT PREPARATION AND SUBMISSION

IV.A Preparing a manuscript for submission to a biomedical journal

IV.A.2 Title page

Propositions:

'8. List of bioresources and/or biobanks used as sources of samples and/or data (and their identifier, if available). Bioresources include both biological samples with associated data (medical/epidemiological, social) and biomolecular research tools. The biosamples and biomolecular resources include any "physical" specimen derived from biological organisms, as well as antibody, affinity binder collections, clone collections, siRNA and microarrays libraries. Research tools include any data directly or undirectly derived from biosamples such as databases, locus specific-databases, registries of disease patients and any specific tool for molecular characterization of biobanked samples.'

'9. Infrastructures. National, European and/or international infrastructure that has evaluated the project.'

Corresponding author

Dr Anne Cambon-Thomsen, Inserm, UMR1027, Université de Toulouse - Université Paul Sabatier - Toulouse III, Epidémiologie et analyses en santé publique
Faculté de médecine. Equipe 4: "Génomique, biothérapies et santé publique : approche interdisciplinaire"
37 allées Jules Guesde F-31073 Toulouse Cedex 7 France
Tel : +33 (0)5 61 14 59 59
Fax : +33 (0)5 61 14 56 23
E-mail : cambon@cict.fr

Acknowledgements

We would like to thank Federica Napolitani, Paola De Castro and Anna Maria Rossi for fruitful discussions.

Financial Disclosure

The BRIF initiative was funded within the European Community's Seventh Framework project "Biobanking and Biomolecular Resources Research Infrastructure" (BBMRI), Grant agreement 212111. The BRIF research also receives funds from FP7 collaborative projects GEN2PHEN (Genotype-to-Phenotype Databases: A Holistic Solution), grant agreement 200754, and BioSHaRE-EU (Biobank Standardisation and Harmonisation for Research Excellence in the European Union), grant agreement 261433.

"The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript."

Author Contributions

Jane Carpenter and Paul Hofman have been involved in providing with the introductory section of the article, Robert Hewitt with the two paragraphs entitled '*What are the obstacles for recognition of the work involved in setting up and maintaining bioresources?*' and the '*Parameters, measures and indicators*', Raymond Dalgleish with the section '*What are the obstacles for recognition of the work involved in setting up and maintaining databases?*', Gudmundur Thorisson with the '*Digital identifier schemes*' paragraph and with other parts of the text, Elena Bravo with the '*Journal guidelines for resource citing and referencing*', Mylène Deschênes with the '*Policies for resource access and sharing*' section. Laurence Mabile and Anne Cambon-Thomsen wrote the '*The BRIF*

concept and objectives' and '*Conclusion*' paragraphs and were in charge with ensuring the global coherence of the manuscript.

Competing Interests

None

Abbreviations

BRIF: Bioresource Research Impact Factor
CEPH family: Centre d'Etude du Polymorphisme Humain family
dbSNP: Single Nucleotide Polymorphism database
dbVar: database of large scale genomic variants
DOI: Digital Object Identifier
HGMD: Human Gene Mutation Database
ID: digital identifier
ICMJE: International Committee of Medical Journal Editors
OMIM: Online Mendelian Inheritance in Man
ORCID: Open Researcher and Contributor ID
URL: Uniform Resource Locator

References

1. Yu Y, Zhu Z (2010) Signficance of biological resources collection and tumour tissue bank creation. World J Gastrointest Oncol 2: 5-8.
2. Kaye J (2011) From single biobanks to international networks: developing e-governance. Hum Genet 130: 377-382.
3. Baust JG (2008) ISBER: Best practices for repositories and trends at the Institute for Problems of Cryobiology and Medicine. Cell Preservation Technology 6: 1-1.
4. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, et al. (2001) dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 29: 308-311.
5. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2011) GenBank. Nucleic Acids Res 39: D32-37.
6. OMIM (1985) Online Mendelian Inheritance in Man. Available: http://omim.org. Accessed 18 September 2011.
7. CEPH (1991) The CEPH family panel. Available: http://hg.wustl.edu/hdk_lab_manual/14/14_1.html. Accessed 4 October 2011.
8. Walport M, Brest P (2011) Sharing research data to improve public health. Lancet 377: 537-539.
9. Altshuler JS, Balogh E, Barker AD, Eck SL, Friend SH, et al. (2010) Opening up to precompetitive collaboration. Sci Transl Med 2: 52cm26.
10. Knoppers BM, Harris JR, Tasse AM, Budin-Ljosne I, Kaye J, et al. (2011) Towards a data sharing Code of Conduct for international genomic research. Genome Med 3: 46.
11. Kaye J, Heeney C, Hawkins N, de Vries J, Boddington P (2009) Data sharing in genomics--re-shaping scientific practice. Nat Rev Genet 10: 331-335.
12. BioBase (2011) BIOBASE biological databases. Available: http://www.biobase-international.com. Accessed 18 September 2011.
13. Cambon-Thomsen A (2003) Assessing the impact of biobanks. Nat Genet 34: 25-26.
14. Cambon-Thomsen A (2004) The social and ethical issues of post-genomic human biobanks. Nat Rev Genet 5: 866-873.
15. Cambon-Thomsen A, Thorisson GA, Mabile L (2011) The role of a bioresource research impact factor as an incentive to share human bioresources. Nature Genetics 43: 503-504.
16. Garfield E (2006) The history and meaning of the journal impact factor. JAMA 295: 90-93.
17. Paskin N (2000) E-citations: actionable identifiers and scholarly referencing. Learned Publishing 13: 159-166.
18. Kauffmann F, Cambon-Thomsen A (2008) Tracing biological collections - Between books and clinical trials. Jama-Journal of the American Medical Association 299: 2316-2318.
19. DataCite (2009) Available: http://www.datacite.org. Accessed 18 September 2011.
20. CrossRef (2000) Available: http://www.crossref.org/. Accessed 17 October 2011.
21. DOI (1998) Digital Object Identifier. Available: http://www.doi.org/. Accessed 17 October 2011.
22. Thorisson GA (2009) Accreditation and attribution in data sharing. Nat Biotechnol 27: 984-985.
23. ORCID (2010) Open Researcher & Contributor ID. Available: http://orcid.org. Accessed 4 October 2011.
24. ICMJE (2009) International Committee of Medical Journal Editors. Available: http://www.icmje.org. Accessed 18 September 2011.
25. Joly Y, Zeps N, Knoppers BM (2011) Genomic databases access agreements: legal validity and possible sanctions. Hum Genet 130: 441-449.
26. EU (1996) legislation. Legal Protection: Databases http://europa.eu/legislation_summaries/internal_market/businesses/intellectual_property/l26028_en.htm. Accessed 17 October 2011.