

# Systematic archiving and access to health research data: rationale, current status and way forward

Manju Rani<sup>a</sup> & Brian S Buckley<sup>b</sup>

**Abstract** Systematically archiving data from health research and large-scale surveys and ensuring access to databases offer economic benefits and can improve the accountability, efficiency and quality of scientific research. Recently, interest in data archiving and sharing has grown and, in developed countries, research funders and institutions are increasingly adopting data-sharing policies. In developing countries, however, there is a lack of awareness of the benefits of data archiving and little discussion of policy. Many databases, even those of large-scale surveys, are not preserved systematically and access for secondary use is limited, which reduces the return on research investment. Several obstacles exist: organizational responsibility is unclear; infrastructure and personnel with appropriate data management and analysis skills are scarce; and researchers may be reluctant to share.

This article considers recent progress in data sharing and the strategies and models used to encourage and facilitate it, with a focus on the World Health Organization Western Pacific Region. A case study from the Philippines demonstrates the benefits of data sharing by comparing the number and type of publications associated with two large-scale surveys with different approaches to sharing.

Advocacy and leadership are needed at both national and regional levels to increase awareness. A step-by-step approach may be the most effective: initially large national databases could be made available to develop the methods and skills needed and to foster a data-sharing culture. Duplication of costs and effort could be avoided by collaboration between countries. In developing countries, interventions are required to build capacity in data management and analysis.

Abstracts in **عربي**, **中文**, **Français**, **Русский** and **Español** at the end of each article.

## Introduction

Despite repeated global calls for increased investment in health research,<sup>1–3</sup> securing investment can be challenging, especially in developing countries where research may compete with health service delivery for funding and personnel. Advocacy for increased investment can also be undermined by stakeholders' doubts about the efficiency and effectiveness of research, by failure to realize the potential of previous investment due to the poor utilization of research outputs and by a low level of public trust in research.<sup>4–6</sup> In this context, some way of increasing the accountability, efficiency and effectiveness of research is needed. In addition to universal clinical trial registration and open access to publications, two closely linked strategies have considerable potential: the systematic archiving of unaggregated data generated by research studies and wider access to databases. Both would facilitate the secondary use of data within and, preferably, between countries.

In recent decades, there have been several high-level initiatives advocating the routine archiving and sharing of health research data.<sup>7–11</sup> The rationale for this is both scientific and economic. Sharing data facilitates reinforces the collaborative and cumulative processes involved in creating scientific knowledge.<sup>7</sup> It can also promote new research and enable the testing of new or alternative hypotheses. For example, combination and meta-analysis of databases can allow researchers to examine trends through time and between regions.<sup>7–10,12,13</sup> In addition, archiving and sharing data can increase the transparency and accountability of research and bolster its reliability and authority by enabling other investigators to repeat or extend analyses. Since data collection is often a significant and expensive aspect of research, ensuring that databases can

be used repeatedly increases the financial return on research investment by reducing the possibility of data duplication.

Despite these benefits, systematic data archiving and sharing are not yet the norm, especially in low- and middle-income countries. Moreover, many health research databases are not efficiently cleaned, managed or used, even by primary researchers, and often data are stored informally by institutions or individual researchers, which makes secondary use impossible.<sup>14</sup> Systematic and secure data archiving can ensure that these valuable resources are available for answering future public health questions.

This paper discusses important developments in data-sharing policy and highlights factors in health research that may affect policy implementation, with particular reference to countries in the World Health Organization (WHO) Western Pacific Region. In addition, practical strategies for fostering data sharing are considered.

## Global context

In 1997, a collaboration of scientific bodies concluded that: "The value of data lies in their use. Full and open access to scientific data should be adopted as the international norm for the exchange of scientific data derived from publicly funded research."<sup>15</sup> Subsequently, in 2004, the 30 countries of the Organization for Economic Cooperation and Development, along with China, Israel, the Russian Federation and South Africa, adopted the Declaration on Access to Research Data from Public Funding and, in 2007, issued principles and guidelines.<sup>7</sup> In 2007, the European Union called upon member states to develop data-sharing policies.<sup>16</sup> The World Bank and the Health 8 group of international health agencies

<sup>a</sup> Western Pacific Regional Office, World Health Organization, Corner Taft and UN Avenue, Manila-1000, Philippines.

<sup>b</sup> Philippine General Hospital, University of the Philippines, Manila, Philippines.

Correspondence to Manju Rani (e-mail: ranim@wpro.who.int).

(Submitted: 4 April 2012 – Revised version received: 21 September 2012 – Accepted: 24 September 2012 – Published online: 10 October 2012)

made commitments towards data sharing in early 2010.

A global consultation in 2008 led to a joint statement of purpose on sharing research data by 17 major research funding organizations in 2011.<sup>10</sup> The statement asserted that: (i) data sharing should be equitable, ethical and efficient; (ii) the interests of researchers who create the data sets, researchers who want to reuse the data and the communities and funders who expect health benefits to ensue from the research should be recognized; (iii) the privacy of individuals must be protected; and (iv) data sharing should increase the quality and value of research.<sup>10</sup> Finally, both the WHO global strategy and plan of action for public health, innovation and intellectual property and the WHO research for health strategy call for greater access to data through improved sharing.<sup>17,18</sup>

Although important and influential, these high-level declarations were limited to general principles and provided little operational guidance on how or when data should be shared. They acknowledged the need for flexibility in individual countries' approaches to data sharing and recognized that the cost must be balanced against the potential benefits.

## Data-sharing policy

Dialogue at the global level has yet to trickle down to the national level and this is particularly true for developing countries. In the WHO Western Pacific Region, there is no ongoing discussion of policy and no specific policy on data sharing in most countries, apart from a few developed countries.<sup>19</sup>

However, in developed countries an increasing number of international research funding institutions are adopting data-sharing policies. The policies do not differ a great deal, although there is some heterogeneity. For example, the National Science Foundation in the United States of America, the Austrian Science Fund, the British Medical Research Council and Cancer Research UK require all applicants for funding to provide detailed strategies for data archiving and sharing, whereas the United States National Institutes of Health has this requirement only for research funding in excess of 500 000 United States dollars and the Wellcome Trust in the United Kingdom requires it for research that results in databases with significant value for the wider research community.

In the Western Pacific Region, the National Health and the Medical Research Council in Australia and the Health Research Council of New Zealand are signatories to the joint statement of purpose on data sharing.<sup>10</sup> The Australian body requires the open publication of and the sharing of data from any research that it funds unless a reason for limiting access can be demonstrated. In New Zealand, a data-sharing policy is being developed for introduction in 2012 through a consultation with stakeholders.<sup>19</sup>

No existing policy stipulates in detail how data should be archived or how access should be managed. This reflects a lack of consensus on best practice that may in part be due to the heterogeneity of health research, the need for flexibility and the lack of a single method that suits all forms of research and data.<sup>9,19</sup> The policy of the National Institutes of Health leaves researchers free to decide where data are archived and how databases are shared. Researchers can use a central repository or establish databases and manage access themselves. The policy of the Austrian Science Fund states that data should be housed in subject-specific or institutional repositories. The Wellcome Trust suggests which repositories could be used but makes no stipulations as long as the archive used is accessible and can be linked to others. In Australia, research institutions must provide adequate facilities and infrastructure for secure data archiving and have a policy on database management.

Approaches also vary on when data should be shared. Most policies explicitly acknowledge that primary researchers should have exclusive access for a certain period. The National Institutes of Health require data to be made available no later than the date of acceptance of the final research report; the Australian National Health and Medical Research Council require them to be available within 12 months of the publication of a peer-reviewed research paper; and the Austrian Science Fund, within 2 years of the end of the project. Other policies, such as those of the British Medical Research Council and the Wellcome Trust, simply refer to "timely" sharing of data. In both Malaysia and Thailand, primary researchers have exclusive use of data for 2 years before sharing is required.<sup>20</sup>

All existing policies maintain that secondary researchers should acknowledge both the primary researchers and

the data source. Some organizations, such as the National Institutes of Health and the Wellcome Trust, recommend that, where appropriate, data should be shared through the collaboration of, and by mutual agreement between, primary and secondary researchers to balance the need to maximize access with the need for safeguards.<sup>11,21</sup>

Existing policies acknowledge that it may not be possible to share some data for ethical, confidentiality or privacy reasons. International agreements on ethical principles in health research involving human subjects state that an individual's privacy and data confidentiality must be protected and that, with some exceptions, informed consent on the use of personal health information must be obtained from study participants.<sup>22,23</sup> It could be argued that these principles are upheld if data are made anonymous before archiving for possible sharing and if the risk to study participants is minimal. However, it has been pointed out that, to date, policies have been proposed without sufficient discussion of how ethical standards can be maintained during data archiving and sharing or how risks to participants can be prevented.<sup>24</sup> Nor do policies provide guidance on these matters.

## Policy effectiveness

There has been little evaluation of whether data-sharing policies are effective for ensuring that researchers comply with recommendations or for increasing the amount of research carried out.<sup>25</sup> Compliance with policy in genomics research, which is considered the front-runner in data sharing, seems to be fairly good, though it is far from universal. One study of papers published in six key journals that required data sharing as a precondition for publication found that at least 85% of authors reported depositing data in the global deoxyribonucleic acid (DNA) data repository.<sup>26</sup> Overall, however, a great deal remains unshared, especially data from studies of cancer and human subjects.<sup>27</sup> Despite claims that microarray data are now routinely stored in accessible archives, less than 50% of data sets are deposited, often because of technical difficulties.<sup>28,29</sup>

Though little "direct" analysis of the public health impact of formal data-sharing policies has been carried out, databases have been shown to have a huge impact on research when they are made accessible. Databases from Demo-

graphic and Health Surveys conducted in more than 70 developing countries since the 1980s are accessible globally, which demonstrates that cultural, ethical and technical barriers to data sharing can be overcome. In 2010 alone, there were nearly 4000 requests for data from these databases.<sup>30</sup> The number of peer-reviewed publications based on data from these surveys has increased substantially and has influenced health policy in many countries.<sup>31</sup> Similarly, by June 2011, some 650 peer-reviewed papers based on the United States National Cancer Institute's Surveillance, Epidemiology and End Results databases had been published; the influence of these databases on researchers' understanding of treatment and survival in cancer is undoubted.<sup>32</sup> The long-running Caerphilly Prospective Study of cardiovascular disease in the United Kingdom has resulted in the publication of some 150 peer-reviewed papers.<sup>33</sup> In addition, data from the nationally representative United Kingdom National Cancer Data Repository has helped in interpreting the results of less representative experimental research.<sup>34</sup> A case study that compared the outputs of two large-scale surveys in the Philippines demonstrated the value of data sharing but also indicated that data analysis capacity needs to be built up if the full benefits are to be realized (Box 1).

### Raising awareness

Currently, the attitude to data sharing in developing countries in the WHO Western Pacific Region is characterized by a widespread lack of awareness or appreciation of its benefits rather than active resistance.<sup>19</sup> In the region, the predominance of external funding for health research and a lack of clarity on the ownership of research outputs has contributed to the indifference observed, especially in low-income countries.

Proactive advocacy is required to ensure that the concept of data sharing becomes a mainstream consideration in national discussions of research management and governance. One way of increasing awareness may be to carry out a systematic assessment of the current situation to demonstrate its inefficiencies and to highlight the loss of valuable scientific data.

### Articulating and enforcing policies

Clear and enforceable data archiving and sharing policies are required. Since

#### Box 1. Data sharing and its effect on research output and efficiency: a case study from the Philippines

Several health programmes and research institutions in the Philippines undertake research based on large-scale surveys that produce data which may be valuable over the long term and help answer a wide range of public health questions. In the absence of an explicit policy or measure that requires individuals and institutions to archive and share data, these data are usually retained by the institutions or individuals that produced them. Consequently, the status and quality of the archives available for future use remain uncertain.

We assessed the effect of data archiving and sharing in the Philippines by comparing the utilization and influence of data collected in two large-scale, nationally representative surveys that are carried out every 5 years: externally funded Demographic and Health Surveys (DHS), which are implemented by the National Statistics Office in each country; and domestically funded National Nutrition Surveys, which are conducted by the Food and Nutrition Research Institute. The most recent surveys – the 8th DHS and 7th National Nutrition Survey – took place in 2008. Both surveys produced data on a range of health indicators.

Whereas DHS data are systematically archived and transparent access is provided via an internationally accessible repository maintained by the external funder, the Food and Nutrition Research Institute provides only aggregated results for National Nutrition Survey data and there is no official information on how to access microdata.

A PubMed search was conducted using the terms *Demographic* (and) *Health Survey* and *Philippines* and *National Nutrition Survey* (and) *Philippines* to identify scientific publications that used data from the surveys. In total, 58 records mentioning a DHS were retrieved; 21 of them directly reported DHS data (10 involved comparisons with other countries and 3 reported trends over time). In contrast, only 14 records were retrieved for National Nutrition Surveys, including one comparative study and one trend analysis.

According to statistics provided by the DHS data repository, between January 2007 and July 2012, 799 distinct users downloaded DHS data from the Philippines on 914 occasions. There were 183 downloads by users within the Philippines and 586 users were from universities. However, an analysis of the authorship of publications that used DHS data showed that external researchers predominated, which indicates that efforts may be needed to build data analysis capacity in developing countries. The wide access to DHS microdata since 1993 was not associated with any resistance from in-country researchers or from the implementing agency (the National Statistical Office). Nor were there any reports that sharing data was associated with misuse or misrepresentation of the data or cultural or ethical issues.

In coming years, the Philippines is likely to witness an increase in scientific output and a better return on investment in research with the development of systematic archiving and wider access to data. Several new initiatives have been launched in the last 3 years with the assistance of the Accelerated Data Program. The National Statistical Office set up a repository for archiving microdata in October 2009, although currently it includes only data from research or surveys conducted by the National Statistical Office. In addition, the National Statistical Coordination Board brought together several agencies to form the Philippine Statistical System, which aims to archive and document microdata using international standards, and recently the National Statistical Office was given the responsibility for maintaining a central repository for archiving microdata.

ensuring the efficiency and effectiveness of health research is a governance issue, it may be appropriate that the lead on data sharing be taken by national health research governance bodies, where they exist, or by elements within ministries of health, such as national health information units, in consultation with all stakeholders in health research. In addition, funders, research institutes and other stakeholders should have their own policies on data sharing, which may provide the opportunity to pilot different approaches while countries prepare national policies.

A policy should state clearly when, where, how and which data should be archived and made available. Heterogeneity in policies, a lack of clarity on

ethical considerations and uncertainty about archiving and sharing methods may both frustrate researchers who want to share data and provide loopholes for those who are unwilling to share.

Clear mechanisms for enforcing and monitoring compliance with data-sharing policies should be developed. In the United States, it has been reported that non-compliance with the National Institutes of Health data-sharing policy in cancer research may have been due to a lack of clarity about data-sharing requirements and the absence of enforcement.<sup>35</sup> The inability of funding agencies to enforce data-sharing policies has also affected compliance in genetics.<sup>29</sup> Partnerships with scientific publishers may be useful for enforcing compliance as

researchers report that the data-sharing policies of scientific journals influence their actions more than those of funders because publication is such an important currency in the world of academic research.<sup>35</sup>

### Overcoming researchers' reluctance

Many researchers have a proprietorial attitude towards data and are concerned that the benefits of data sharing might be outweighed by perceived disadvantages: the loss of academic advantage and independence; the possibility that their work may be misused, misinterpreted or misrepresented; the loss of intellectual property; and an increased workload for administration and data management.<sup>14,25,35–38</sup> A survey of the first authors of research articles published in the *Annals of Internal Medicine* in 2008 demonstrated a hesitancy about data sharing: only 4% said they would share data unconditionally, whereas 57% would do so only under author-defined conditions and most would not share data without personal contact with secondary users.<sup>36</sup> Hence, a period of exclusive data use for primary researchers – an approach advocated by data-sharing policies internationally – may be required to protect their interests and ensure they receive the appropriate benefit and recognition.<sup>39</sup>

Another issue affects the attitude of researchers and policy-makers in developing countries to sharing data internationally. Researchers in developing countries may have invested considerable effort in data collection and database generation, but often better-resourced researchers in developed countries analyse and publish data without sufficiently collaborating with or acknowledging the primary researchers. This inequity has been acknowledged by high-level global advocates of data sharing. Strategies are required to prevent this potential inequity and to encourage sharing of both skills and data between countries and regions.<sup>10,13,20</sup> However, researchers may be encouraged by the beneficial research collaborations that can result from sharing data.<sup>14,35,40</sup> In Thailand, national data sets are made available to international researchers only on the condition that they form skill-sharing and collaborative partnerships with local scientists.<sup>20</sup>

A realignment of the way in which research achievement is evaluated may

also be beneficial. Currently, recognition of individual researchers and institutions and researchers' advancement depend largely on peer-reviewed publications. This fosters competition and a degree of secrecy among researchers at the expense of collaboration. In this context, sharing data seems counter-productive and, consequently, the creation, curation and utility of databases are given relatively little attention. The joint statement of purpose by research funding organizations acknowledged that the generation of valuable databases deserved better recognition as a research activity.<sup>10</sup>

### Increasing skills and resources

Although generating and maintaining well organized and well-documented databases is part of good research practice,<sup>41</sup> researchers in developing countries may have neither the skills nor the resources required.<sup>14,40,42</sup> Data management training for researchers and the recruitment of dedicated support staff to document data and manage repositories may be needed.<sup>10,43</sup> In addition, data archiving and sharing may also be constrained by the lack of accepted protocols for data formats, security and transfer.<sup>19</sup> The introduction of modest minimum standards and the preparation of supporting materials for research databases, which may use different formats, would make data reformatting and interpretation easier for secondary users.

The International Household Survey Network and the Accelerated Data Program, both of which started in 2006, are important initiatives in this area. They are involved in developing standards for data documentation and in building national capacity in microdata preservation, analysis, anonymization and dissemination. In addition, the Accelerated Data Program is helping countries establish national data repositories using international data standards.

Although data archiving and sharing require financial and human resources, this is counterbalanced by the resulting rise in opportunities for collaboration and increased scientific output. The joint statement of purpose acknowledged that funders should underwrite the cost of data sharing.<sup>10</sup>

### Accessing databases

Both developed and developing countries in the WHO Western Pacific

Region report limited awareness of the existence of many databases currently available for secondary use and difficulty in locating them, which may decrease the return on research investment in these countries. Existing data archiving and sharing models recognize that some method for locating databases is needed.

Several models for archiving and sharing research databases exist (Box 2). The portal model may be the most effective for encouraging a culture of data sharing because it allows primary researchers to retain involvement with their databases while facilitating database searching, data sharing and collaboration between primary and secondary researchers. It may also minimize the resources required in developing countries. Other models, such as centralized archiving with disseminated expert support and the subject-focused repository model, necessitate greater investment in infrastructure and require coordination. This makes them more difficult to implement in settings lacking resources, capacity and cohesive health research governance.

Controlling access to data and quality control of data use are also important. Concerns have been expressed that unconditional access to databases may result in poor quality secondary studies, which could undermine the reputation of the data sources and primary studies.<sup>14,39,48</sup> Just as citations to papers are monitored, some way of monitoring database usage is also needed, both to evaluate the effectiveness of data-sharing policies and to ensure that databases are appropriately referenced and acknowledged.<sup>10,39,41,43</sup>

### Prioritizing data for archiving and sharing

Given the cost and infrastructure implications of data archiving and sharing, a good starting point in the short term could be the development and implementation of data-sharing policies for databases associated with large-scale surveys and registers, since these offer fewer challenges and provide the greatest benefit to health research. Many data sets from large-scale surveys, which are often externally funded and initiated, duplicate effort because separate surveys ask similar questions and the data are subsequently underused. In some countries, such as the Philippines and

**Box 2. Data archiving and sharing models****The centralized model**

In this model, data archiving and managing access to data are handled by a single repository to which researchers are obliged to submit their databases along with supporting documentation. Rights to the data are either ceded to the repository or terms are agreed to govern limits on access. The United Kingdom Data Archive, which has operated successfully since the late 1960s, uses this model.<sup>44</sup> Reported advantages include: cost-effective infrastructure use; the opportunity to train and retain highly skilled data archivists; removal of the burden of data archiving from institutions; and provision of a "one-stop shop" for researchers seeking data for secondary use. However, the centralized system is the least popular model with many researchers, who place a high value on being able to monitor, influence or participate in the secondary use of their data.<sup>14,35,39,45</sup> In addition, it has been argued that centralized data archives cannot provide the same expert understanding of research databases as the generating institutions or more specialized repositories.<sup>14</sup>

**Centralized archiving with disseminated expert support**

This model involves centralized data archiving and access combined with the support of experts in several participating institutions. These diverse centres of expertise, which may often include the original research groups, support the secondary use of databases but do not make decisions about access. The advantages of this model are broadly similar to those of the fully centralized model, plus the availability of expert support. However, researchers' concerns about the loss of control remain and are compounded by the requirement to provide ongoing support, which gives them no greater rights. In addition, the cost of coordinating expert support makes the model less cost-effective.<sup>14,46</sup>

**Subject-focused repositories**

Databases are archived in repositories that specialize in specific research areas. Infrastructure use is reasonably cost-effective and the support provided to users benefits from greater familiarity with the data being managed. However, it has been argued that this model can work against interdisciplinary collaboration because boundaries between research areas are not always well defined, making databases harder to locate.<sup>47</sup>

**Portal models**

A portal provides links to databases stored elsewhere, often in the institutions that created them. The portal relies on searchable metadata about the remote databases and provides users with a one-stop shop for identifying and locating databases. The portal service does not, however, manage access to databases or data transfer: these functions remain in the hands of the primary researchers and their institutions. This model is less costly than others and has the advantage that it can provide links to data archives anywhere in the world. In addition, researchers retain more control over data sharing and reuse. However, since there is no central data repository, data preservation standards cannot be assured and consistent access to data cannot be guaranteed.

**Conclusion**

Routine data archiving and sharing offers considerable benefits: the effectiveness and efficiency of health research could be increased and science and health-care policy could advance more rapidly. However, if the potential is to be realized equitably, especially in developing countries, advocacy and leadership are needed at both national and regional levels.

The most effective way of achieving the ultimate goal of universal data archiving and sharing may be to adopt a gradual, multistage approach. Increased access to national databases hosted by statutory bodies can pave the way for data sharing by smaller, but nonetheless valuable, individual research databases. Research funders should encourage researchers to maximize the value of their databases and adopt consistent data standards and management strategies when designing new studies.

The infrastructure, skills and standards needed for data archiving and sharing may be best developed through international partnerships and skill sharing, thereby avoiding the duplication of effort. The creation of good databases and good data management should be recognized as legitimate research activities by funders and academic culture alike, and developing countries should start building capacity in data management and analysis. ■

Viet Nam, aggregate data from national health surveys are published. However, data archiving is fragmented and there is no clear arrangement for accessing microdata.

The establishment of good archiving and data-sharing practices for these databases would enable the host bodies to achieve several objectives. First, valuable data would be made available for national health research. Second, the process of identifying, implementing

and evaluating contextually appropriate methods for the wider preservation and sharing of data could begin. Third, the growth of a "data-sharing research culture" would be encouraged. This could increase awareness and understanding of the rationale for, and benefits of, data sharing and pave the way for more wide-ranging policies and strategies that could be extended to academic institutions and investigator-initiated research databases.

**Acknowledgements**

We would like to thank Vicente Belizario, James D Best, Dave Carr, Jaime Montoya, Robin Olds, Kia Reinis, Robert Terry and all participants in the Expert Consultation on Health Research Management, Governance and Data Sharing in the Western Pacific convened by the WHO Regional Office for the Western Pacific in August 2011.

**Competing interests:** None declared.

**ملخص**

أرشفة بيانات البحوث الصحية والوصول إليها على نحو منهجي: الأساس المنطقي والوضع الراهن والخطوات التالية تتوافر فوائد اقتصادية نتيجة لأرشفة البيانات الواردة من البحوث الصحية والدراسات الاستقصائية واسعة النطاق وضمان الوصول إلى قواعد البيانات على نحو منهجي، كما يمكنها تحسين المساءلة والكفاءة والجودة الخاصة بالبحوث العلمية. ومؤخراً، زاد الاهتمام بأرشفة البيانات وتبادلها، وتقوم جهات تمويل الأبحاث والمؤسسات البحثية في البلدان المتقدمة باعتماد سياسات لتبادل

البيانات على نحو متزايد. وبرغم ذلك، ثمة قلة وعي في البلدان النامية بفوائد أرشفة البيانات ومناقشة السياسات على نحو غير مستفيض. ولا يتم حفظ العديد من قواعد البيانات حتى تلك الخاصة بالدراسات الاستقصائية واسعة النطاق على نحو منهجي، كما أن الوصول إلى الاستخدام الثانوي محدود، وهو ما يخفض المردود على الاستثمار البحثي. كما توجد العديد من العقبات: عدم

ثمة حاجة للدعوة والقيادة على المستويين الوطني والإقليمي بُغية زيادة الوعي. وقد يكون النهج التدريجي هو الأكثر فعالية: يمكن مبدئياً إتاحة قواعد البيانات الوطنية الضخمة لتطوير الأساليب والمهارات اللازمة وتعزيز ثقافة تبادل البيانات. ويمكن تجنب مضاعفة التكاليف وتكرار الجهود من خلال التعاون بين البلدان. ويلزم تنفيذ تدخلات في البلدان النامية بُغية بناء القدرات في مجال إدارة البيانات وتحليلها.

وضوح المسؤولية التنظيمية؛ وندرة البنية التحتية والأفراد الذين تتوافر لديهم مهارات مناسبة لإدارة وتحليل البيانات؛ واحتمال إحجام الباحثين عن تبادل البيانات.

تتناول هذه المقالة التقدم الراهن في تبادل البيانات والاستراتيجيات والنماذج المستخدمة لتشجيعه وتيسيره، مع التركيز على إقليم غرب المحيط الهادئ في منظمة الصحة العالمية. وتثبت دراسة حالة من الفلبيين فوائد تبادل البيانات من خلال مقارنة عدد ونوع المطبوعات المرتبطة بدراستين استقصائيتين واسعتي النطاق بنهج مختلفة لتبادل البيانات.

## 摘要

### 卫生研究数据的系统归档和访问:原理、现状及前进方向

系统地归档来自卫生研究和大规模调查的数据,确保数据库访问提供经济效益,并能改善科学研究的问责制、效率和质量。最近,发达国家对数据存档和共享的兴趣增加,研究资助者和机构越来越多地采用数据共享政策。然而,在发展中国家,缺乏对数据归档好处的认识,对政策的讨论很少。许多数据库(甚至是大规模调查的数据库)没有系统地保留,二次利用的访问很有限,这降低了研究的投资回报率。存在的若干障碍包括:组织的责任不明确、缺乏配有数据管理和分析能力的基础设施和人员、研究人员可能不愿意共享。

本文考虑了数据共享的最新进展和用于鼓励和促进数

据共享的策略和模式,将重点放在世界卫生组织西太平洋地区。来自菲律宾的案例研究通过将不同的共享方式的两个大规模调查的相关出版物数量和类型进行比较,证明数据共享的好处。

要提高认识,需要在国家和地区级别上进行宣传和领导。循序渐进的方法可能最有效:最初的全国性大型数据库可以用来发展所需的方法和技巧,并促进数据共享文化。各国之间的合作可避免重复的成本和工作。在发展中国家,需要采取干预措施来建立数据管理和分析的能力。

## Résumé

### Archivage systématique et accès aux données de recherche sanitaire: raisons d'être, état actuel et perspectives

Archiver systématiquement les données de la recherche sanitaire, ainsi que les enquêtes à grande échelle, et assurer l'accès aux bases de données offre des avantages économiques et peut améliorer la transparence, l'efficacité et la qualité de la recherche scientifique. Récemment, l'intérêt pour l'archivage et le partage des données s'est accru et, dans les pays développés, les organisations qui financent la recherche et les institutions adoptent de plus en plus de politiques de partage des données. En revanche, dans les pays en voie de développement, il y a un manque de prise de conscience des avantages liés à l'archivage des données et peu de discussions sur la politique de partage. De nombreuses bases de données, y compris celles des enquêtes à grande échelle, ne sont pas conservées de manière systématique, et leur accès pour une utilisation ultérieure est limité, ce qui réduit le retour sur investissement de la recherche. Plusieurs obstacles existent: la responsabilité organisationnelle n'est pas claire, l'infrastructure est limitée, tout comme le personnel disposant des compétences adéquates en matière d'analyse et de gestion des données, et les chercheurs peuvent être réticents à partager leurs

résultats.

Cet article examine les progrès récents réalisés dans le partage des données, ainsi que les stratégies et les modèles utilisés pour l'encourager et le faciliter, en mettant l'accent sur la région du Pacifique occidental de l'Organisation mondiale de la Santé. Une étude de cas aux Philippines démontre les avantages du partage de données en comparant le nombre et le type des publications associées à deux enquêtes à grande échelle, avec différentes approches du partage de données.

La défense du partage de données et une direction efficace sont nécessaires, aux niveaux national et régional afin d'accroître la sensibilisation. Une approche étape par étape peut être la plus efficace: de grandes bases de données nationales pourraient d'abord être mises à disposition pour développer les méthodes et les compétences nécessaires, et favoriser une culture du partage de données. La duplication des coûts et des efforts pourrait être évitée grâce à la collaboration entre les pays. Dans les pays en voie de développement, des interventions sont nécessaires pour renforcer les capacités de gestion et d'analyse des données.

## Резюме

### Систематическое архивирование и доступ к данным исследований в области здравоохранения: обоснование, современное состояние и дальнейшее развитие

Систематическое архивирование данных медицинских и других крупномасштабных исследований, а также обеспечение доступа к базам данных, представляется экономически выгодным и помогает улучшить подотчетность, эффективность и качество научных исследований. В последнее время интерес к архивированию и обмену данными вырос, и в развитых странах финансовые источники исследований и другие учреждения все чаще внедряют политику обмена данными. В развивающихся

же странах наблюдается недостаточная осведомленность о преимуществах архивирования данных и отсутствие активности в обсуждении данной политики. Многие базы данных, даже в составе крупномасштабных исследований, не сохраняются систематически, и доступ к ним для вторичного использования ограничен, что снижает рентабельность инвестиций в исследования. Причин тому несколько: неопределенная структурная организация, недостаток инфраструктуры и

персонала с соответствующими навыками управления и анализа данных, а также, возможно, нежелание исследователей делиться результатами своих работ.

В данной статье рассматриваются последние достижения в области обмена данными, а также стратегии и модели, используемые для поощрения и облегчения обмена данными, акцентируя внимание на Западном Тихоокеанском регионе Всемирной организации здравоохранения. Пример Филиппин демонстрирует преимущества совместного использования данных на основе сравнения количества и типа публикаций, связанных с двумя крупномасштабными исследованиями с различными подходами к обмену данными.

Для повышения осведомленности в данной области необходимы пропаганда и руководство как на национальном, так и на региональном уровнях. При этом пошаговый подход видится наиболее эффективным: первоначально может быть предоставлен доступ к крупным национальным базам данных в целях разработки методов и навыков необходимых для развития культуры обмена данными. Увеличения затрат и дублирования усилий можно избежать путем сотрудничества стран. В развивающихся странах необходимо внедрение практических мер по наращиванию потенциала в области управления и анализа данных.

## Resumen

### Almacenamiento y acceso a los datos de investigación sanitaria: fundamento, situación actual y camino a seguir

Almacenar de forma sistemática los datos de investigaciones sanitarias y encuestas a gran escala, así como asegurar el acceso a las bases de datos presentan ventajas económicas y pueden mejorar la responsabilidad financiera, la eficiencia y la calidad de las investigaciones en salud. El interés en el almacenamiento e intercambio de datos ha crecido recientemente y, en los países desarrollados, los proveedores de fondos para la investigación y las instituciones están adoptando cada vez más estrategias de intercambio de datos. En los países en desarrollo, sin embargo, se desconocen las ventajas del almacenamiento de datos y no hay apenas debates acerca de las estrategias. Muchas bases de datos, incluso las de encuestas a gran escala, no están sujetas a un mantenimiento sistemático y el acceso para el uso secundario es limitado, lo que reduce la rentabilidad de las inversiones en investigación. Existen numerosos obstáculos: la responsabilidad en las organizaciones no es clara, la infraestructura y el personal con las capacidades adecuadas de análisis y gestión de datos son escasos, así como una posible reticencia de los investigadores a compartir los datos.

Este artículo considera los progresos recientes en el intercambio de datos y las estrategias y modelos empleados para fomentarlos y facilitarlos, enfocados a la Oficina Regional para el Pacífico Occidental de la Organización Mundial de la Salud. Un estudio de caso de Filipinas demuestra las ventajas del intercambio de datos al comparar el número y tipo de las publicaciones asociadas con dos encuestas a gran escala con diferentes enfoques para compartir.

Se necesita apoyo y liderazgo tanto a nivel nacional como regional para aumentar la conciencia. Un enfoque paso a paso podría ser el más efectivo: en una etapa inicial, podrían ponerse a disposición grandes bases de datos nacionales a fin de desarrollar los métodos y las habilidades necesarias y para promover una cultura de intercambio de datos. A través de la colaboración entre países podría evitarse la duplicación de costes y esfuerzos. En los países en desarrollo, son necesarias intervenciones a fin de desarrollar la capacidad de análisis y gestión de datos.

## References

1. Commission on Health Research for Development. *Health research: essential link to equity in development*. New York: Oxford University Press; 1990. Available from: [www.who.int/rpc/summit/documents/summit\\_report\\_final2.pdf](http://www.who.int/rpc/summit/documents/summit_report_final2.pdf) [accessed 8 October 2012].
2. *Report from the ministerial summit on health research. Identify challenges, inform actions, correct inequities. Mexico City, 16–20 November 2004*. Geneva: World Health Organization; 2005.
3. Scaling up research and learning for health systems: now is the time. Report of a high level task force. *Presented and endorsed at the: Global Ministerial Forum on Research for Health, Bamako, 2008*. Geneva: World Health Organization; 2009.
4. Getz KA. Public trust and confidence today: a review of public opinion polls. *Monitor*. 2008 17–21 September. Available from: [www.ciscrp.org/downloads/articles/Getz\\_publicopinion.pdf](http://www.ciscrp.org/downloads/articles/Getz_publicopinion.pdf) [accessed 3 October 2012].
5. Goodyear M. Learning from the TGN1412 trial. *BMJ* 2006;332:677–8. doi:10.1136/bmj.38797.635012.47 PMID:16554332
6. Mullings AM. Research ethics committees: preserving research integrity and the public trust. *West Indian Med J* 2007;56:105–7. doi:10.1590/S0043-31442007000200001 PMID:17910138
7. *OECD principles and guidelines for access to research data from public funding*. Paris: Organisation for Economic Co-operation and Development; 2007.
8. Chan M, Kazatchkine M, Lob-Levyt J, Obaid T, Schweizer J, Sidibe M et al. Meeting the demand for results and accountability: a call for action on health data from eight global health agencies. *PLoS Med* 2010;7:e1000223. doi:10.1371/journal.pmed.1000223 PMID:20126260
9. *Final NIH statement on sharing research data*. Bethesda: National Institutes of Health; 2002. Available from: <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html> [accessed 3 October 2012].
10. Walport M, Brest P. Sharing research data to improve public health. *Lancet* 2011;377:537–9. doi:10.1016/S0140-6736(10)62234-9 PMID:21216456
11. *Policy on data management and sharing*. London: Wellcome Trust; 2010. Available from: <http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm> [accessed 3 October 2012].
12. *MRC policy on data sharing and preservation*. London: Medical Research Council; 2005. Available from: <http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Datasharinginitiative/index.htm> [accessed 3 October 2012].
13. Pisani E, Whitworth J, Zaba B, Abou-Zahr C. Time for fair trade in research data. *Lancet* 2010;375:703–5. doi:10.1016/S0140-6736(09)61486-0 PMID:19913902
14. Corti L, Wright M. *MRC population data archiving and access project: consultants' report*. London: Medical Research Council; 2002. Available from: <http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC003682> [accessed 8 October 2012].
15. Committee on Issues in the Transborder Flow of Scientific Data of the National Research Council. *Bits of power: issues in global access to scientific data*. Washington: National Academy Press; 1997.
16. *Communication from the Commission to the European Parliament, the Council and the European Economic and Social Committee on scientific information in the digital age: access, dissemination and preservation*. Brussels: Commission of the European Communities; 2007.
17. Resolution WHA61/21. Global strategy and plan of action on public health, innovation, and intellectual property. In: *Sixty-first World Health Assembly, Geneva, 19–24 May 2008. Official records: resolutions, decisions and annexes*. Geneva: World Health Organization; 2012. Available from: [http://apps.who.int/gb/or/e/e\\_wha61r1.html](http://apps.who.int/gb/or/e/e_wha61r1.html) [accessed 8 October 2012].

18. Resolution WHA63/22. WHO's role and responsibilities in health research: Draft WHO strategy on research for health. In: *Sixty-third World Health Assembly, Geneva, 17–21 May 2010. Official records: resolutions, decisions and annexes*. Geneva: World Health Organization; 2012. Available from: [http://apps.who.int/gb/e/e\\_wha63.html](http://apps.who.int/gb/e/e_wha63.html) [accessed 3 October 2012].
19. *Expert consultation on improving health research management, governance and data sharing in the western Pacific*. Manila: World Health Organization Regional Office for the Western Pacific; 2011.
20. Tangcharoensathien V, Boonperm J, Jongudomsuk P. Sharing health data: developing country perspectives. *Bull World Health Organ* 2010;88:468–9. doi:10.2471/BLT.10.079129 PMID:20539864
21. *NIH data sharing policy and implementation guidance*. Bethesda: National Institutes of Health; 2003. Available from: [http://grants.nih.gov/grants/policy/data\\_sharing/data\\_sharing\\_guidance.htm](http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm) [accessed 3 October 2012].
22. *International ethical guidelines for biomedical research involving human subjects*. Geneva: Council for International Organizations of Medical Sciences and World Health Organization; 2002.
23. *World Medical Association declaration of Helsinki: ethical principles for medical research involving human subjects*. Ferney-Voltaire: World Medical Association; 2008. Available from: <http://www.wma.net/en/30publications/10policies/b3/17c.pdf> [accessed 3 October 2012].
24. Dawson A, Verweij M. Could do better: research data sharing and public health. *Public Health Ethics* 2011;4:1–3. doi:10.1093/phe/phr011
25. Piwowar HA. *Foundational studies for measuring the impact, prevalence, and patterns of publicly sharing biomedical research data*. Pittsburgh: University of Pittsburgh; 2010.
26. Noor MAF, Zimmerman KJ, Teeter KC. Data sharing: how much doesn't get submitted to GenBank? *PLoS Biol* 2006;4:e228. doi:10.1371/journal.pbio.0040228 PMID:16822095
27. Piwowar HA. Who shares? Who doesn't? Factors associated with openly archiving raw research data. *PLoS One* 2011;6:e18657. doi:10.1371/journal.pone.0018657 PMID:21765886
28. Thou shalt share your data. *Nat Methods* 2008;5:209. doi:10.1038/nmeth0308-209
29. Ochsner SA, Steffen DL, Stoeckert CJ Jr, McKenna NJ. Much room for improvement in deposition rates of expression microarray datasets. *Nat Methods* 2008;5:991. doi:10.1038/nmeth1208-991 PMID:19034265
30. Reinis K. Data sharing in the Demographic and Health Surveys. In: *Expert consultation on improving health research management, governance and data sharing in the western Pacific*. Manila: World Health Organization Regional Office for the Western Pacific; 2011.
31. Short Fabric M, Choi Y, Bird S. A systematic review of Demographic and Health Surveys: data availability and utilization for research. *Bull World Health Organ* 2012;90:604–12. doi:10.2471/BLT.11.095513 PMID:22893744
32. *SEER-Medicare publications by journal & year*. Bethesda: National Cancer Institute; 2011. Available from: [http://healthservices.cancer.gov/seermedicare/overview/pubs\\_jour\\_year.php](http://healthservices.cancer.gov/seermedicare/overview/pubs_jour_year.php) [accessed 22 October 2011] doi:10.1056/NEJMs0810119 PMID:19144931
33. *Caerphilly prospective study*. Bristol: University of Bristol / MRC Epidemiology Unit (South Wales); 2011. Available from: <http://www.epi.bris.ac.uk/caerphilly/caerphilly.htm> [accessed 21 October 2011].
34. *Data centres: their use, value and impact: a research information network report*. London: Research Information Network; 2011.
35. Tucker J. *Motivating subjects: data sharing in cancer research* [thesis]. Blacksburg: Virginia Polytechnic Institute and State University, 2009.
36. Laine C, Berkwijs M, Mulrow C, Shaeffer MG, Griswold M, Goodman S. Reproducible research: biomedical researchers' willingness to share information to enable others to reproduce their results. In: *Proceedings of the Sixth International Congress on Peer Review and Biomedical Publication*; 2009 10–12 September; Vancouver, Canada. Available from: <http://www.ama-assn.org/public/peer/abstracts-0910.pdf> [accessed 3 October 2012].
37. Reidpath DD, Allotey PA. Data sharing in medical research: an empirical investigation. *Bioethics* 2001;15:125–34. doi:10.1111/1467-8519.00220 PMID:11697377
38. Savage CJ, Vickers AJ. Empirical study of data sharing by authors publishing in PLoS journals. *PLoS One* 2009;4:e7078. doi:10.1371/journal.pone.0007078 PMID:19763261
39. Lowrance WW. *Access to collections of data and materials for health research: a report to the Medical Research Council and the Wellcome Trust*. London: Medical Research Council & Wellcome Trust; 2006.
40. Rice R. *DISC-UK DataShare project evaluation report*. Edinburgh: EDINA and University Data Library, University of Edinburgh; 2009. Available from: <http://www.disc-uk.org/docs/Datashare-eval-final.pdf> [accessed 3 October 2012].
41. den Eynden VV, Corti L, Woollard M, Bishop L, Horton L. *Managing and sharing data*. Colchester: UK Data Archive; 2011.
42. *Interview with Susanna-Assunta Sansone*. San Francisco: Scientific Data Sharing Project; 2011. Available from: <http://scientificdatasharing.com/general/interviewwith-susanna-assunta-sansone> [accessed 31 October 2011].
43. Pisani E, AbouZahr C. Sharing health data: good intentions are not enough. *Bull World Health Organ* 2010;88:462–6. doi:10.2471/BLT.09.074393 PMID:20539861
44. *Across the decades: 40 years of data archiving*. Colchester: UK Data Archive, University of Essex; 2007.
45. Piwowar HA, Chapman WW. Identifying data sharing in biomedical literature. *AMIA Annu Symp Proc* 2008;2008:596–600. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2655927/> [accessed 8 October 2012].
46. *Economic and Social Data Service*. Swindon: UK Data Archive; 2011. Available from: <http://www.data-archive.ac.uk/about/services/esds> [accessed 3 October 2012].
47. Natural Environment Research Council [Internet]. Data centres. Swindon: Natural Environment Research Council; 2011. Available from: <http://www.nerc.ac.uk/research/sites/data/> [accessed 8 October 2012].
48. Green A, Macdonald S, Rice R. *Policy-making for research data in repositories: a guide*. Edinburgh: EDINA and University Data Library, University of Edinburgh; 2009. Available from: <http://www.disc-uk.org/docs/guide.pdf> [accessed 3 October 2012].