Creating a global observatory for health R&D

Robert F. Terry,1# José F. Salm Jr.,2 Claudia Nannii,3 Christopher Dye4

A global map of health R&D activity would improve the coordination of research and help to match limited resources with public health priorities, such as combating antimicrobial resistance. The challenges of R&D mapping are large because there are few standards for research classification and governance and limited capacity to report on R&D data, especially in low-income countries. Nevertheless, based on developments in semantic classification, and with better reporting of funded research through the Internet, it is now becoming feasible to create a global observatory for health R&D.

The issue of knowing what research is currently being undertaken—where, by whom, and which organizations are supporting it—is a black hole in the public health landscape. In a funding environment still reeling from the economic crisis, being able to do more with what we have—thinking inside the box—is a priority, especially for those diseases disproportionately affecting developing countries and for which normal market incentives fail. Market failure describes those situations where the members of a population that suffer most from a disease—for example, leishmaniasis carried by sand flies—are too poor to provide a market that incentivizes industry to undertake research and development (1, 2).

How to finance research and development where normal market forces are absent has been the focus of a number of studies organized by the World Health Organization (WHO), culminating in 2012 with a report that assessed the strengths and weaknesses of more than 100 new financing mechanisms (3). One of the issues that became clear in compiling this report was the absence of good data. There is no global health R&D map that provides a comprehensive picture of research funding, ongoing research, and results that could be used to guide the allocation of the limited available funding. Consequently, the member states of WHO have called for the establishment of a global observatory on health R&D to address this lack of information (4). The vision is of an observatory that would compile data from bibliometrics and available databases on patents, registered clinical trials, the product pipeline, and the data related to current and ongoing R&D held in the databases of research funders (Fig. 1) (5).

Although some of the information necessary to create a global health R&D map can be found in existing databases, there is nothing that covers all diseases for all countries. There are a number of regular R&D surveys undertaken by, for example, the Organisation for Economic Cooperation and Development (OECD) for its members; Red de Indicadores de Ciencia y Tecnología (RICYT), covering Spain, Portugal and 29 countries in the Americas; and the UNESCO Institute. However, these data are not interoperable because of the varied methods of classifying the research and the different degrees of detail reported—for example, information on medical research as opposed to different diseases. This means that if you wanted to understand a global picture of, for example, research into new antibiotics, you would be left with undertaking a lengthy and expensive survey with all the difficulties such surveys entail.

One of the biggest challenges we face is developing capacity for collecting and reporting data on R&D funding when only 37% of countries are currently able to publicly report data of this type (5). Work in the WHO Regional Office for the Western Pacific has begun to address this challenge. The office has developed software that creates a management system to register a research project and prepare it for ethics approval. Technical assistance has been provided to five developing countries within the region: Cambodia, Laos, Fiji, Mongolia, and Papua New Guinea; subsequently some of them have launched national registries (6). Although this will not capture all health research, the approach of combining ethics approval with reporting on R&D funding provides a good example of getting multiple uses out of a single collection of data.

As with many WHO projects of this type, it is a new activity and will require new and additional funding outside of its existing budget. A conservative estimate is that $11.5 million will be needed in the first 5 years to cover project

---

**Fig. 1. Components of a global health observatory.** Much of the initial data for the observatory will be garnered from available sources such as the following (examples rather than complete lists are shown). Publications: Medline and PubMed, the Biblioteca Virtual de Saúde (BIHREM), Thomson Reuters, and others. Regional R&D hubs: Health Research Web, Regional Platform on Access and Innovation for Health Technologies (PRAIS), and others. Patents: World Intellectual Property Organization (WIPO) Patent Scope, Re:Search, WIPO Gold, and national and regional patent databases. Clinical trial registries: WHO International Clinical Trials Registry Platform. Product pipeline: Global Health Primer. Monitoring of R&D funding flows: UNESCO Institute for Statistics, OECD, Eurostat, RICYT, G-FINDER (31 neglected tropical diseases), Association of Southeast Asian Nations (ASEAN) science and technology indicators, Organization of Islamic States R&D, African Union disease-specific tracking (e.g., HIV). R&D funder databases: ÜberResearch, World RePORT, and others.

---

1/ Special Programme for Research and Training in Tropical Diseases, World Health Organization, Geneva, Switzerland.
2/ Stela Institute, Center for Management and Socioeconomic Science (ESAG), Santa Catarina State University, Florianópolis, Brazil.
# Corresponding author. E-mail: terryr@who.int
staff and software development and to build capacity in those countries (the majority) that do not report health R&D data.

A proposal: Automated mapping of the research landscape

The aim is to create a global health R&D observatory by adopting the latest approaches in data mining to maximize the collection, synthesis, and interpretation of data, and to minimize the cost of bringing the data together.

As more of the administration of R&D becomes digital, more of it comes online on the Web sites of individual researchers, their institutions, and the research funders. This provides raw data but adds little knowledge. For a global R&D observatory to be feasible, automating the coding and subsequent association of available data will be essential. When operational, this would enable data on research funding to be recorded locally in whatever format suits that context, but, through automated translation, it could be reported against a global standard.

This requires a culture shift to create greater transparency in the R&D system by publishing more information on current research that is open access, allowing automated reading by machines such as Web crawlers. To create a map of current activity, all public research funders and academic institutions (and eventually industry) should publish metadata, abstracts, or even whole texts of current research protocols on the Web in a machine-readable code, such as XML. This will require standards, because there are numerous variations in the current research information systems (CRIS) used by many universities, particularly in Europe and North America. However, if the funders of research could publish a common set of metadata points—for example, principal investigator, host institute, project title, project abstract, time scales, and amount and source of funding, then it would be technically feasible to create a current research map by means of search engines rather than surveys. This is not as insurmountable an obstacle as it might seem; access to Medline via PubMed only began in 1997, and the system has quickly moved from manual deposition of citation data to an automated system that is integrated into the workflows of medical publishers. We need to build a similar system into our research-funding processes.

The major challenge is classifying the research itself. There are high-level research classification systems that exist, and the OECD’s Frascati manual is a good example used to estimate gross domestic expenditure on R&D (GERD). However, the Frascati methodology aggregates R&D funding under medical and health sciences, with subcategories, but does not report the detail below this. It is used mainly by OECD countries.

So the opportunity cost of first agreeing on a standard and then getting every researcher and funding body to use it accurately is large. The alternative is to develop semantic-based classification using search engines that can automate the classification step. Health has an advantage over other sciences because it already has a number of comprehensive vocabularies with agreed-upon definitions or ontologies—for example, the Unified Medical Language System of the NIH and the International Classification of Diseases (ICD). Although there is no single global standard, these ontologies provide the seed from which automated classification can grow.

In South America at the end of the 1990s, the Brazilian Council of Science, Technology, and Innovation (CNPq) launched the Lattes Platform, with a goal of providing information services for researchers, students, institutions, research groups, and agency staff. Although originally conceived as a centralized database, during the first years of operation it developed a roadmap for improving access to data and providing an ontological classification to help share information (Consciencias da Informação) (9). Most of this standardization work was coordinated by the Stela Institute/Stela Group, a nonprofit Brazilian research organization that has been responsible for the development of other Brazilian government systems, including Aquarius, a newly launched “all federal funds” investment tracking system, and many other federal- and state-based infrastructure systems. Today, it is an exemplar of how individual researcher data (for example, standardized bibliographical data) can be integrated with other information (such as publications) to create a picture of research at institutional and national levels.

Using the experience of the Lattes Platform Stela Institute, the National Science Foundation, the National Institutes of Health, and the Office of Science and Technology Policy have discussed a similar architecture for the American Research Profile System. The intention is to use semantic technology and a data adapter to transform proprietary data (researcher IDs, CVs, funded protocols, patents, and publications) from different data sources to a common format to create Linked Open Data (LOD).

By starting with the individual researcher profile, an R&D map could be built from the bottom up—as opposed to surveys, which tend to be top down. Such a system enables social network analysis, identifying research collaborations, coauthorship distance networks, and the degree of cross-disciplinary research.

This vision of semantic, automated classification sounds a long way off, but current efforts demonstrate more than just proof of principle. For example, UberResearch has recently produced an analytical tool and an R&D database from a selection of public health research funders, including NIH and the Wellcome Trust. It analyzes the abstracts of current research projects available on their Web sites by means of semantics and coded them to the Medical Subject Headings (MeSH) of the National Library of Medicine.

The resulting database allows searches to be conducted from a user’s desktop and demonstrates the value of analyzing research trends, visualizing research networks, and identifying centers of excellence, even using the limited data set that UberResearch can currently access from public sites.

Next steps

At a technical workshop hosted by the Wellcome Trust in February 2013, a number of agencies that collect R&D data were brought together to discuss the feasibility of creating a global resource for health research. The recommendations were to begin by building on data that already exists, to understand what users want and what they will use in planning their own R&D activities. Understanding user needs will be vital for creating a resource that can be sustainably supported.

Using these criteria, two existing resources on R&D were investigated as a first step. Emory University manages a database called the Global Health Primer that seeks to map the pipeline for health technologies (diagnostics, vaccines, and drugs) in development for the neglected tropical diseases (NTDs). Separately, the G-FINDER database operated by Policy Cures tracks funding for R&D in the NTDs area by more than 3600 research funders. We therefore commissioned a stakeholder analysis to design a platform linking these data together. A consultation was conducted that included interviews with 27 of the major stakeholders of R&D in NTDs responsible for 73% of global...
funding for neglected diseases R&D in 2012. In addition, an online survey attracted 91 responses from 30 different countries. The findings support the need for a new platform, with the majority of respondents wanting a resource that they can rely on in terms of quality, that is regularly updated for use in planning, and that is relatively simple to use. Their key request was for a common tool or portal to provide access to both R&D funding and pipeline information, including details on diseases, products, funders, and developers. The stakeholders preferred to have the ability to access the underlying data for their own analyses rather than being given a suite of ready-provided analyses. Presenting the data as simple tables and graphs was preferred over interactive animations and maps, even if these were seen as “nice-to-haves.” With regard to content, in addition to the data providing a current picture of R&D funding by disease, product, and organization, there were requests for predictive analysis trying to estimate future funding needs. An example was the need for advance warning to fund expensive late-stage clinical trials if early development looks promising (13). On the basis of these findings, a business case for this platform—which would form one part of a global health observatory—is being designed, with a beta version expected to be launched in 2015.

The political driver for establishing such an observatory grew from discussions about non-commercial areas of health research in the NTDs, and this is the initial focus of the observatory. However, market failure affecting R&D is not only an issue in the prevention and treatment of diseases of poor countries. As antimicrobial resistance continues to rise, the pipeline for new antibiotics is not supplying enough new drugs (14, 15). So, among other factors, new financial incentives and more public money will be needed to cover the risk of R&D for antibiotic development, increasing the need to ensure efficiency in the public R&D system. Consequently, there have already been calls from the WHO member states for the observatory to expand its remit to include a mapping of antimicrobial resistance R&D as one way to facilitate global collaboration and improve efficiency (16).

Open access to research publications and, increasingly, to data is now well established as a principle of good research practice. To a large degree, this has been facilitated by the policies of the health research funders. In moving forward, what would be really beneficial is if those open-access principles could also be applied to the data sets that the research funders hold themselves.

With the call from health ministers through WHO and the recent findings of our consultation, there is a desire for better understanding of global health R&D and a need to improve the coordination of research, particularly around global efforts. This desire in itself is not new, but with the potential of the Internet and innovations in automated research mapping techniques, it is now becoming feasible to create a global observatory for health R&D.

REFERENCES AND NOTES

10. UberResearch database of public health research funders (subscription required); www.uberresearch.com
10.1126/science.1258737