

COMMENTARY

Open Access

# Mapping global health research investments, time for new thinking - A Babel Fish for research data

Robert F Terry<sup>1\*</sup>, Liz Allen<sup>2</sup>, Charles A Gardner<sup>3</sup>, Javier Guzman<sup>4</sup>, Mary Moran<sup>5</sup> and Roderik F Viergever<sup>6</sup>

## Abstract

Today we have an incomplete picture of how much the world is spending on health and disease-related research and development (R&D). As such it is difficult to align, or even begin to coordinate, health R&D investments with international public health priorities.

Current efforts to track and map global health research investments are complex, resource-intensive, and caveat-laden. An ideal situation would be for all research funding to be classified using a set of common standards and definitions. However, the adoption of such a standard by everyone is not a realistic, pragmatic or even necessary goal.

It is time for new thinking informed by the innovations in automated online translation - e.g. Yahoo's Babel Fish. We propose a feasibility study to develop a system that can *translate* and map the diverse research classification systems into a common standard, allowing the targeting of scarce research investments to where they are needed most.

## A need for accurate information on research resource flows

Our understanding of how much is spent on health and disease-related research and development (R&D), who is spending it, on what, and where, is very limited. This failure to accurately oversee financial flows for health research hinders our ability to make health research investments effectively address international public health priorities. This inability is especially critical where resources and the capacity to undertake research are low [1].

The Global Forum for Health Research has published estimates for global spending on health research for the past ten years based primarily on a biennial science and technology (S&T) spending survey conducted by the OECD [2]. However, teasing out what proportion of S&T investments actually go to health research is not easy, requires significant assumptions, and the resulting global numbers cannot be disaggregated by disease or purpose of the research (e.g., biomedical versus health systems

research). The interpretation of such data is therefore complex and difficult for strategists and policy makers.

So the existence of comprehensive and accurate information on research expenditure across the globe would bring a range of benefits to individual researchers, funders, national governments, and to those involved in managing and directing research resources. Better estimates of R&D resource flows would enable the comparison and benchmarking of what is being spent in for example malaria drug development versus drug delivery research, or biomedical versus health systems research, or how much is spent on cancer research in different countries. It would help research funders to make strategic investments, enable coordination and reduce duplication to increase the impact of the billions of dollars that are invested in health research every year.

## A kaleidoscope of R & D classification systems

Our current view of the global health research landscape is further obscured by the diverse classification systems and nomenclatures that the funders of research have adopted to define their portfolios and deliver against their own remits. A multitude of such classification

\* Correspondence: [terryr@who.int](mailto:terryr@who.int)

<sup>1</sup>Public Health, Innovation and Intellectual Property, World Health Organization, Avenue Appia, Geneva CH-1211, Switzerland  
Full list of author information is available at the end of the article

systems for health, disease and the related research is currently in use across the globe. These systems typically combine a description of the health or disease topic, often using the International Statistical Classification of Diseases and Related Health Problems (ICD10) or Medical Subject Headings developed by the US National Library of Medicine (MeSH), with a description of the objective, purpose or type of research. While the resulting outputs differ markedly the similarity in the approach to classification - with use of a disease code combined with a description of the research purpose - suggests there is a common understanding or principle of what a classification system should incorporate.

An ideal improvement would be that all research investments are classified using an agreed set of standards and definitions. However, encouraging research funders to harmonize and align their individual classification systems to a common standard may be unwieldy, impractical and perhaps an unrealistic expectation. In addition, it is no small challenge for many countries to provide even the most basic data on their R&D resource flows over time.

For example, in the UK biomedical field, UK research funders have classified their funding portfolios against the Health Research Classification System (HRCS) developed by the UK Clinical Research Collaboration (UKCRC) [3]. The HRCS was first used in 2005 to classify the portfolio of a number of major biomedical funding bodies and allowed some of the first analysis of research foci across UK organisations; a second wave of classification was repeated for 2010 resource flows across the same organisations to compare and to describe any trends that may have emerged over time. While there is interest in using the HRCS by other European research funders [4], undertaking the actual classification requires considerable manual coding and in addition the integration of the classification system with multiple funders' grants systems is not a simple process. This requirement for a manual coding step, common to the use of any classification system, is both costly and time consuming.

### **Translation not standardization**

Today the revolution in text, data mining and semantic web analysis presents us with new opportunities to achieve automated and large scale translation efforts. Instead of organizations being required to classify their investments using a common standard, it should be feasible to develop a system that can translate diverse research funding descriptions to a "*lingua franca*" that delivers systematic and comprehensive maps of resource flows for the first time.

The precedent is there with several recent innovations. For example Natural Language Processing (NLP)

algorithms used by Collexis to search and match related documents using free text rather than complicated Boolean search terms and translational software systems used by the French Multi-Terminology Indexer (F-MTI) [5]. Online language translation, such as Google Translate or Yahoo's Babel Fish, is improving and will get better with cloud technology. These innovations suggest that such a research translation system might be achievable. The use of translation tools is currently being explored to track the impact and outcomes of research [6].

G-FINDER has provided a recent practical demonstration of how a translation approach can work to produce insight into international research commitments. Supported by the Bill & Melinda Gates Foundation, the G-FINDER survey aims to provide "comprehensive data to help funders and product developers better understand where funding gaps lie and how their investments fit into the global picture". It has done so for one area of health research, that of product R&D for neglected diseases. G-FINDER now covers all major public, private and philanthropic funders in high-income countries, and major funders in some middle-income "innovative developing countries" [7].

The work involved in collating and reconciling the data provided by all the funders contributing to the G-FINDER survey is, however, substantial. The G-FINDER team has developed its own way of mapping existing classifications of neglected disease R&D against an agreed, central code frame, but this is currently a largely manual process.

### **First steps: exploring the feasibility of a translation system**

Automation of mapping individual classification systems against a commonly agreed standard using new software tools would be a major breakthrough innovation. In our view, a feasibility study is needed to explore whether such a mapping and translation approach can deliver more accurate and insightful resource flow mapping for health and disease-related R&D.

As part of this study, many practical issues must be resolved including how to build on what is already strong and familiar across existing classification systems that are now in use, and how to generate agreement on the classification standard. In addition, the mechanism of translation itself and the degree to which it could be automated will be a key element in the usefulness of the envisioned translation system. Different options for such a mechanism should be explored, as well as how it would be maintained, curated and governed and how any reporting and analysis would be delivered.

We propose that a number of principles guide the development of a translation system to improve its chances of success. In our view, a translation system should be:

accurate, cost effective and sustainable; flexible and able to evolve over time; equitable i.e. it meets the needs of all (or most) global users; not burdensome to users that supply information and/or interface with the system; and able to generate output that is open and accessible to all.

The initial resources required to scope and develop a translation system may be significant. However, longer term efficiencies that will be gained through our ability to track global, regional and national investments in health and disease R&D, coupled with improved coordination and more strategic investments, should far outweigh that initial investment.

Our ultimate goal is to develop a translation tool that is of value to as many stakeholders as possible, and that is sufficiently detailed to enable its use in resource flow mapping and strategy and priority setting. The premise of our paper is that there is a role for automation to improve the efficiency of R&D classification and, as a consequence increase the likelihood of gaining access to better data. Such efforts will be essential if the desire for greater harmonization in global health R&D are ever to be realised [8].

#### Disclaimer

The views and opinions expressed in this article are those of the authors and do not necessarily reflect those of the organizations they represent.

#### Author details

<sup>1</sup>Public Health, Innovation and Intellectual Property, World Health Organization, Avenue Appia, Geneva CH-1211, Switzerland. <sup>2</sup>Strategic Planning & Policy Unit, The Wellcome Trust, 215 Euston Road, London NW1 2BE, United Kingdom. <sup>3</sup>Health, UBS Optimus Foundation, Augustinerhof 1, Zürich CH-8098, Switzerland. <sup>4</sup>Policy Cures, 36 Gordon Square, London WC1H 0PD, United Kingdom. <sup>5</sup>Policy Cures, 46 Market Street, Sydney, NSW 2000, Australia. <sup>6</sup>Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, 15-17 Tavistock Place, London WC1H 9SH, United Kingdom.

Received: 28 April 2012 Accepted: 6 August 2012

Published: 1 September 2012

#### References

1. World Health Assembly resolution WHA61.21; 2008. [http://www.who.int/phi/implementation/phi\\_globstat\\_action/en/index.html](http://www.who.int/phi/implementation/phi_globstat_action/en/index.html).
2. Global Forum for Health Research: *Global Forum for Health Research*. Monitoring financial flows for health research series; 2001–2009.
3. UK Clinical Research Collaboration (2005) Health Research Classification System. 2005. [http://hrcsonline.net/sites/default/files/HRCS\\_Document.pdf](http://hrcsonline.net/sites/default/files/HRCS_Document.pdf).
4. Health Research Classification Systems – Current Approaches and Future Recommendations. European Science Foundation: Policy Brief 43; 2011.
5. Pereira S, *et al.*: Multi-terminology indexing for the assignment of MeSH descriptors to medical abstracts in French. *AMIA Annu Symp Proc* 2009, 521–525.
6. Lane J: Measuring the Results of Science Investments. *Science* 2011, **331**(6018):678–680. doi:10.1126/science.1201865.
7. Moran M, *et al.*: *The G-Finder report 2010. Neglected Disease Research & Development*. Policy Cures: Is the global financial crisis changing R&D? 2010. [http://www.policycures.org/downloads/g-finder\\_2010.pdf](http://www.policycures.org/downloads/g-finder_2010.pdf).
8. Report of the Consultative Expert Working Group on Research and Development: *Financing and Coordination Research and Development to Meet Health Needs in Developing Countries*. World Health Organization, Geneva: Strengthening Global Financing and Coordination; 2012.

doi:10.1186/1478-4505-10-28

**Cite this article as:** Terry *et al.*: Mapping global health research investments, time for new thinking - A Babel Fish for research data. *Health Research Policy and Systems* 2012 10:28.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

