Science AAAS

Advancing Global Health Research Through Digital Technology and Sharing Data

Trudie Lang

Science **331**, 714 (2011);
DOI: 10.1126/science.1199349

ownership, latency, and release, even within the academic community, are diffuse and polarized, which calls for standards set by publishers, societies, and funding agencies. Data quality and formats vary, and different disciplines have much to gain by developing their own standards, architectures, interface designs, and metadata tied to particular species of data. Many professional societies, however, still seem unaware that voxel data fundamental to new discoveries in their own disciplines are not being released, much less validated or reused, and in many cases, are not being saved or curated at all.

Online supplemental data limits are too small for voxels, nor is it the role of publishers to manage primary data collections. Extending the GenBank model to voxels is a solution within reach, if not without obstacles. Sustained funding is paramount. Another need is professional advancement, still bound to metrics of conventional publication, while the more fundamental tasks of data generation and management go unrewarded. Young careers are still best served by publishing words and pixels, and abandoning used voxels to get on to the next project. As funding agencies pour increasing millions into scanners and scanning, only negligible funding and thought have gone to data archiving or leveraging their initial investments. There is urgency to act. As second-generation voxel scientists have now begun to retire, their data are on track to die with them, as it did with the first voxel pioneers, even as we now train a third generation in 3D imaging and computation. Funding agencies can rejoice in the unexpected longevity and growing value in voxels they have already produced. But they must first secure the basic tenet of science by ensuring that researchers have the means to archive, disclose, validate, and repurpose their primary data.

References and Notes

1. B. A. Price, I. S. Small, R. M. Baecker, in *Proceedings of the 25th Hawaii International Conference on System Sciences*, 7 to 10 January 1992, vol. 2, p. 597.
2. J. L. Contreras, *Science* **329**, 393 (2010).
3. C. Hess, E. Ostrom, *Understanding Knowledge as a Commons: From Theory to Practice* (MIT Press, Cambridge, MA, 2007).
4. E. C. Beckmann, *Br. J. Radiol.* **79**, 5 (2006).
5. G. C. Conroy, M. W. Vannier, *Science* **226**, 456 (1984).
6. T. B. Rowe, W. Carlson, W. Bottorff, *Thrinaxodon: Digital Atlas of the Skull* (Univ. of Texas Press, Austin, TX, CD-ROM ed. 1, 1993).
7. www.DigiMorph.org/
8. www.ctlab.geo.utexas.edu/
9. http://digitalfishlibrary.org/ and http://csci.ucsd.edu/
10. www.sio.ucsd.edu/ and http://collections.ucsd.edu/mv/

11. This study was based on funding to L.R.F. from NSF (DBI-0446389) and T.R. from NSF (EAR-0948842, CNS-0709135, IIS-0208675, IIS-9874781, EAR-940625B, BSR-89-58092), the Intel Foundation, and the Jackson School of Geosciences.

10.1126/science.1202828

PERSPECTIVE

Advancing Global Health Research Through Digital Technology and Sharing Data

Trudie Lang

The imperative for improving health in the world's poorest regions lies in research, yet there is no question that low participation, a lack of trained staff, and limited opportunities for data sharing in developing countries impede advances in medical practice and public health knowledge. Extensive studies are essential to develop new treatments and to identify better ways to manage healthcare issues. Recent rapid advances in availability and uptake of digital technologies, especially of mobile networks, have the potential to overcome several barriers to collaborative research in remote places with limited access to resources. Many research groups are already taking advantage of these technologies for data sharing and capture, and these initiatives indicate that increasing acceptance and use of digital technology could promote rapid improvements in global medical science.

Clinical research in the world's poorest regions lags behind the rest of the globe, but, because these communities carry the highest burden from disease, data from studies conducted in these areas could make the biggest impact on global health. Lack of trained staff, low investment in health research, and remote communities with poor infrastructure combine to make clinical research in these settings challenging (*1*). Innovative use of digital technology has the potential to drive important changes in global health research and in many cases is doing so already. Within the convention of collecting data through clinical studies, there are good examples of how novel data-capture mechanisms are being used across a spectrum of global health research areas. These technologies in combination with open access, data sharing, and knowledge exchange (*2*) could transform clinical research in the world's poorest regions.

Global Health Clinical Trials, Centre for Tropical Medicine, University of Oxford, Oxford OX3 7LJ, UK. E-mail: trudie.lang@ndm.ox.ac.uk

Communities most affected by diseases of poverty are held back in economic development by the perpetual cycle of ill health and low income. The world's poorest communities experience many health issues, often concomitantly. Therefore, rather than working in separation on a specific disease, groups need to work together on complex and overlapping challenges (*3*). This paper surveys how digital technology is being harnessed to capture, record, store, combine, and share a diversity of data sets. Nevertheless, there are both practical and notional problems to overcome if this technology is to deliver its full potential in advancing improvements in global health.

Data Capture and Sharing in Global Health Research

The process of tackling a disease or health issue begins by characterizing the problem, initially by the collection of epidemiology and laboratory data. Once a disease or health issue is understood (even if only in part), targets can be sought and interventions designed. An intervention could be a new drug or vaccine, but it might be a new way of diagnosing a disease, introducing a change in the training of healthcare workers, or a new practice in managing a disease or healthcare situation. Any new interventions like these must be tested through clinical trials, and ongoing observational studies are necessary to support implementation and policy by monitoring the effectiveness of the intervention, its cost effectiveness, and associated changes to quality of life. These measures are required in any healthcare setting, but in the field of global health capturing data to address these questions can be problematic if populations are hard to reach, infrastructure is poor, and there is a lack of trained staff to conduct such studies. Technologies, ranging from geographic information systems to gene sequence databases, are beginning to be used to address key public health questions in resource-limited settings and are already making impressive advances.

Data to Quantify and Describe Health Issues

In developing countries, key data from health and demographic surveillance systems have to be accessed by both researchers and policymakers; hence, there are efforts to standardize and link such databases. In remote areas, handheld satellite Global Positioning Systems are proving valuable for surveying communities and collecting data to characterize and quantify diseases and to identify exactly where a new intervention or change in management is needed (Fig. 1). The International Network for the Demographic Evaluation of Populations and Their Health in Developing Countries (INDEPTH) is a global network engaged in conducting longitudinal health and demographic evaluation of populations in low- and middle-income countries (*4*). The capacity of this and similar networks has been greatly enhanced by advances in satellite technology, Internet access speed, and increased access to handheld data-entry devices. Development of open-source software such as openXdata (*5*) has transformed surveillance studies by bringing scale, quality, coverage, and, importantly, knowledge and sharing of best practice. These surveillance networks are succeeding as groups become willing to share data and collaborate and are beginning to make global disease surveillance a reality. Similarly, the East African Community has made an important step forward in establishing the East African Integrated Disease Surveillance Network (EAIDSNet), in which countries share data on communicable diseases to improve public health in their region (*6*). In Asia, the South East Asia Infectious Diseases Clinical Research Network is successfully sharing data across dozens of sites and several countries and makes many resources for researchers available on the Web (*7*, *8*).

There are many other types of networks operating in combination to give a more comprehensive understanding of specific diseases, and when data are combined a powerful picture emerges. The Malaria Atlas Project (*9*) works with geographers, statisticians, epidemiologists, biologists, and public health specialists in endemic countries to assemble a spatial database combining medical intelligence and satellite-derived climate data to define the limits of malaria transmission. This initiative has succeeded in compiling an archive of community-based estimates of parasite prevalence drawn from 85 countries.

In the search for new interventions to combat diseases of poverty, sharing and linking databases are clearly advantageous. Assembling findings on biochemical targets, genetics studies, and the pathogens themselves is vital to improve the speed of new drug and vaccine development. One such project is MalariaGen (*10*), in which researchers from 21 countries are collaborating to build a malaria genome database. This project has addressed some of the fundamental challenges that are inherent to data sharing, such as the ethical challenges of recruiting participants and setting out clear agreements on data linking and release. Establishing a set of policies to address these issues has worked well for this network (*11*).

Fig. 1. A health worker uses EpiSurveyor (www.episurveyor.org) free mobile data-collection software in Cameroon. [Photo reproduced with permission from J. Selanikio, www.DataDyne.org]

Measuring the Potential of New Interventions

Clinical trials are highly challenging in resource-limited settings, but obstacles are being countered by remote data-collection technology, as well as by distance learning and knowledge-sharing strategies. Open-source clinical trial data management systems (Fig. 2) permit international standard data management for noncommercial organizations by removing the cost impediments of commercial clinical trial software (*12*). The rapid expanse of mobile phone networks, and more recently 3G (third-generation) technology, has transformed the potential for electronic direct data entry in even the most remote corners of the globe. The continual evolution of mobile networks expands the possibility for clinical trials to be conducted to high data-quality standards in many regions and increases the professional networking opportunities for local researchers.

A scarcity of trained staff has limited some countries' capacity to design and manage independent trial programs without the involvement of external sponsors (*13*). Knowledge sharing will not only increase the numbers of skilled staff but also improve methods. The wiki concept (Web sites where content can be openly shared, changed, and developed) is likely to be an important contribution, because through this route researchers can share tools and protocols and improve the design and conduct of clinical trials.

Large clinical trials to establish novel approaches to disease management are needed, for example, to evaluate new uses for antibiotics or to assess treatment and management options for a variety of healthcare issues. Such trials generate large data sets that are too cumbersome for conventional publication, but increasingly these are being made available online. A good example of one recent large trial is the AQUAMAT study of hospitalized malaria patients, which enrolled 5425 patients across 11 centers in nine African countries. This trial provides supplementary data online, including enrollment figures, endpoint review, and quality assessments (*14*), all of which will benefit others planning similar trials or policy-makers needing more detail than that provided in the primary publication.

Data to Drive Policy Change and Support Implementation

Licensing of new drugs and vaccines, or gaining evidence for a new public health measure, is not the end in terms of collecting and sharing data. Changes in national treatment and management policies require ongoing data on safety and efficacy, quality of life outcomes, and health economic impacts; hence, it is important that after implementation of any new intervention data continue to be gathered. For instance, pathogens can become resistant to drugs; hence, networks have been established that monitor changes in drug efficacy. For example, the World Wide Anti-Malarial Resistance Network of disease-endemic country researchers collates data to inform and respond rapidly to the malaria parasite's ability to adapt to drug treatments (*15*).

Pharmacovigilance provides long-term safety monitoring vital to the success of drug implementation programs, but these activities generate

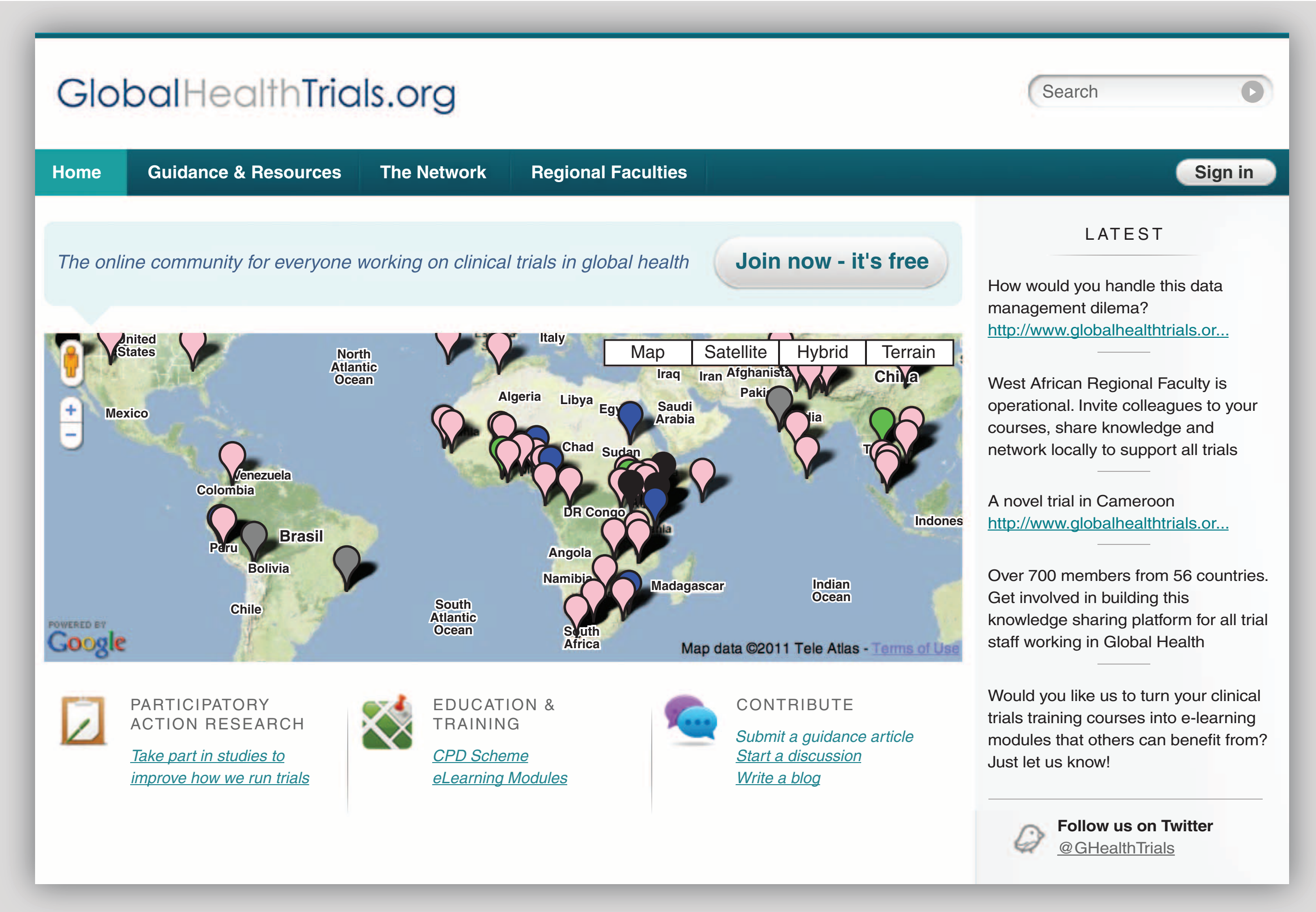

Fig. 2. Global Health Trials is a free, open-access collaborative program that aims to promote and to make easier the conduct of noncommercial clinical trials across all diseases in resource-poor settings by providing guidance and support and enabling the sharing of best practice. [Image reproduced with permission from www.globalhealthtrials.org]

large volumes of data that are difficult to handle. The World Health Organisation Collaborating Centre for Advocacy and Training in Pharmacovigilance aims to improve drug monitoring in developing countries (*16*) through the provision of large shared databases, training programs, and resource development. Another project that embraces the advantages of open-source software and the ethos of data sharing is the Millennium Global Village Project in sub-Saharan Africa. In this project, many ongoing research activities are linked by their Web-based program (*17*). Organizations are also using digital technology to support point-of-care diagnostics and treatment in remote regions poorly supplied with medical expertise and where treatment is hard to access, through the use of decision tree tools and access to resources and guidance. (*18*). There's no doubt that taking knowledge to the community via smart phones and laptops holds enticing potential for health care under any circumstances.

Issues and Challenges of Data Sharing and Capture

The examples present the illusion that digital technology is being readily adopted in global health research; unfortunately, this is not yet the case. The reality is one of unreliable electricity supply, not enough computers, and numerous and seemingly trivial (but cumulatively limiting) frustrations. Even simple equipment failures can become insuperable problems if parts are expensive to obtain and engineers are rare. In describing the issues they faced in setting up an electronic system to support clinical decisions in HIV care, a group in Kenya concluded that the ability to recognize and adapt to the specific needs of resource-limited settings was fundamental to successful implementation (*19*).

Thus, the difficulties of applying digital technology and data capture in developing-world settings are both practical and philosophical. The practical challenges encompass the range of physical and technical mechanisms of capturing, handling, and storing the data. This includes access to the technology and equipment, as well as skills training. There are still gaps in funding and knowledge to be met. The matters of individual and organizational attitudes to digital technology and adoption of a wiki culture are not so easy to address and depend on the acceptance and understanding of new technology and concepts to realize the potential they hold.

An essential component for successful adoption of digital media is the willingness of scientific communities to share data (*20*). Researchers acknowledge that data sharing increases the impact, utility, and profile of their work. Conversely, research is highly competitive (*21*, *22*), and publications depend on individual ability to produce novel data, which can be a disincentive for collaboration. There are also major ethical considerations in sharing data between researchers and between countries and in making data available for open access. The issues around consent and ownership are yet more complex within networks. Common frameworks and defined principles first need to be established if a data-sharing network is to succeed, particularly when it comes to the ethical and privacy issues surrounding patient data (*23*, *24*).

Shifting Attitudes

Widely dispersed researchers in resource-limited countries may have few opportunities to travel to courses or attend meetings, but they can meet online and share experiences, guide each other, and access resources. Learning and knowledge sharing online could play a vital role in adjusting the imbalance in research capacity. However, this medium for learning needs to become accepted, and senior research staff need to encourage and enable their colleagues to take up the numerous free and open-access learning opportunities that are increasingly available online (*13*).

Undoubtedly integration and knowledge sharing can be vastly improved to make the most use of gathered data, but many organizations in global health exist to address a single disease or work in a specific sector. There is a real need for mechanisms allowing research organizations, governments, and universities to collaborate outside their usual remits and locations to maximize the impact of data and available resources.

Governance and ethical issues are also a major concern, because if mistakes are made trust will be quickly lost and enthusiasm for opening access could be stifled. A particular anxiety resulting from disparities between wealthy and resource-limited nations is the removal of data and loss of ownership. Ownership and governance arrangements need to be made transparently for fair access and maintenance of security, and whenever possible the technology should be transferred rather than the data. These issues therefore need to be tackled openly and comprehensively early in the formation of data-sharing collaborations. Groups would be advised to seek advice and obtain example policy documents (such as agreements and terms of reference) from other successful data-sharing groups.

A striking range of data sets spanning a wide range of healthcare issues, including infectious and noncommunicable diseases, are accumulating with use of new technology and online collaboration. All this stands to make real changes in the lives of people affected by diseases of poverty. While scientists are rapidly adapting and taking up these approaches, funding agencies and regulators also need to adapt to ensure that all interested communities are able to take maximum advantage of the digital environment to drive improvements in global health.

References and Notes

1. P. Mwaba, M. Bates, C. Green, N. Kapata, A. Zumla, *Lancet* **375**, 1874 (2010).
2. E. Wenger, W. Snyder, *Harv. Bus. Rev.* **2000**, 139 (Jan.-Feb. 2000).
3. A. de-Graft Aikins *et al.*, *Global. Health* **6**, 5 (2010).
4. P. Kowal *et al.*, *Glob. Health Action* **3** (suppl. 2), 10.3402/gha.v3i0.5302 (2010).
5. OpenXData, www.openxdata.org.
6. EAIDSNet, www.eac.int.
7. H. F. Wertheim *et al.*, *PLoS Med.* **7**, e1000231 (2010).
8. The South East Asia Infectious Disease Clinical Research Network, www.seaicrn.org.
9. S. I. Hay, R. W. Snow, *PLoS Med.* **3**, e473 (2006).
10. The Malaria Genomic Epidemiology Network, *Nature* **456**, 732 (2008).
11. M. Parker *et al.*, *PLoS Med.* **6**, e1000143 (2009).
12. G. W. Fegan, T. A. Lang, *PLoS Med.* **5**, e6 (2008).
13. T. A. Lang *et al.*, *PLoS Negl. Trop. Dis.* **4**, e619 (2010).
14. A. M. Dondorp *et al.*, *Lancet* **376**, 1647 (2010).
15. P. J. Guerin, S. J. Bates, C. H. Sibley, *Curr. Opin. Infect. Dis.* **22**, 593 (2009).
16. M. Pirmohamed, K. N. Atuah, A. N. Dodoo, P. Winstanley, *Br. Med. J.* **335**, 462 (2007).
17. A. S. Kanter *et al.*, *Int. J. Med. Inf.* **78**, 802 (2009).
18. D-Tree International, www.d-tree.org/.
19. S. F. Noormohammad *et al.*, *Int. J. Med. Inf.* **79**, 204 (2010).
20. B. A. Fischer, M. J. Zigmond, *Sci. Eng. Ethics* **16**, 783 (2010).
21. E. Pisani, C. AbouZahr, *Bull. W. H. O.* **88**, 462 (2010).
22. J. Whitworth, *Bull. W. H. O.* **88**, 467 (2010).
23. B. Malin, D. Karp, R. H. Scheuermann, *J. Investig. Med.* **58**, 11 (2010).
24. R. Horton, *Lancet* **355**, 2231 (2000).
25. The author received no specific funding for this work and has no conflicts of interest to declare.

10.1126/science.1199349

PERSPECTIVE

More Is Less: Signal Processing and the Data Deluge

Richard G. Baraniuk

The data deluge is changing the operating environment of many sensing systems from data-poor to data-rich—so data-rich that we are in jeopardy of being overwhelmed. Managing and exploiting the data deluge require a reinvention of sensor system design and signal processing theory. The potential pay-offs are huge, as the resulting sensor systems will enable radically new information technologies and powerful new tools for scientific discovery.

Until recently, the scientist's problem was a "sensor bottleneck." Sensor systems produced scarce data, complicating subsequent information extraction and interpretation. In response to the resulting challenge of "doing more with less," signal-processing researchers have spent the last several decades creating powerful new theory and technology for digital data acquisition (digital cameras, medical scanners), digital signal processing (machine vision; speech, audio, image, and video compression), and digital communication (high-speed modems, Wi-Fi) that have both enabled and accelerated the information age.

Department of Electrical and Computer Engineering, Rice University, Houston, TX 77251–1892, USA. E-mail: richb@rice.edu

These hardware advances have fueled an even faster exponential explosion of sensor data produced by a rapidly growing number of sensors of rapidly growing resolution. Digital camera sensors have dropped in cost to nearly $1/megapixel; this has enabled billions of people to acquire and share high-resolution images and videos. Millions of security and surveillance cameras, including unmanned drone aircraft prowling the skies, have joined high-resolution telescopes, digital radio receivers, and many other types of sensors in the environment. As a result, a sensor data deluge is beginning to swamp many of today's critical sensing systems.

In just a few years, the sensor data deluge has shifted the bottleneck of many data acquisition systems from the sensor back to the processing, communication, or storage subsystems (Fig. 1). To see why, consider the exponentially growing gap between global sensing and data storage capabilities. A recent report (*1*) found that the amount of data generated worldwide (which is now dominated by sensor data) is growing by 58% per year; in 2010 the world generated 1250 billion gigabytes of data—more bits than all of the stars in the universe. In contrast, the total amount of world data storage (in hard drives, memory chips, and tape) is growing 31% slower, at only 40% per year. A milestone was reached in 2007, when the world produced more data than could fit in all of the world's storage; in 2011 we already produce over twice as much data as can be stored. This expanding gap between sensor data production and available data storage means that sensor systems will increasingly face a deluge of data that will be unavailable later for further analysis. Similar exponentially expanding gaps exist between sensor data production and both computational power and communication rates.

The danger is that more sensor data can lead to less efficient sensor systems. Consider two brief illustrations. The first is the Defense Advanced Research Projects Agency (DARPA) Autonomous Real-Time Ground Ubiquitous