

# Grid-Enabled Measures

## Using Science 2.0 to Standardize Measures and Share Data

Richard P. Moser, PhD, Bradford W. Hesse, PhD, Abdul R. Shaikh, PhD, MHSc, Paul Courtney, MS, Glen Morgan, PhD, Erik Augustson, PhD, MPH, Sarah Kobrin, PhD, Kerry Y. Levin, PhD, Cynthia Helba, PhD, David Garner, Marsha Dunn, MPH, Kisha Coa, MPH

---

**Abstract:** Scientists are taking advantage of the Internet and collaborative web technology to accelerate discovery in a massively connected, participative environment—a phenomenon referred to by some as Science 2.0. As a new way of doing science, this phenomenon has the potential to push science forward in a more efficient manner than was previously possible. The Grid-Enabled Measures (GEM) database has been conceptualized as an instantiation of Science 2.0 principles by the National Cancer Institute (NCI) with two overarching goals: (1) promote the use of standardized measures, which are tied to theoretically based constructs; and (2) facilitate the ability to share harmonized data resulting from the use of standardized measures. The first is accomplished by creating an online venue where a virtual community of researchers can collaborate together and come to consensus on measures by rating, commenting on, and viewing meta-data about the measures and associated constructs. The second is accomplished by connecting the constructs and measures to an ontological framework with data standards and common data elements such as the NCI Enterprise Vocabulary System (EVS) and the cancer Data Standards Repository (caDSR). This paper will describe the web 2.0 principles on which the GEM database is based, describe its functionality, and discuss some of the important issues involved with creating the GEM database, such as the role of mutually agreed-on ontologies (i.e., knowledge categories and the relationships among these categories—for data sharing).

(Am J Prev Med 2011;40(5S2):S134–S143) Published by Elsevier Inc. on behalf of American Journal of Preventive Medicine

---

### Introduction

The digital and information age has transformed how we access information and interact with others; a transformation that has implications for the conduct of health science. For example, the NIH was one of the first scientific agencies to use networked information technologies (a prototype cyberinfrastructure) to coordinate input from literally hundreds of laboratories from around the world to document the more than 3 billion base pairs comprising the human genome.<sup>1</sup> This technology-mediated effort virtually defined the concept of “team science”<sup>2</sup> in an era of distributed computing. Likewise, current efforts to extract research value from the terabytes of data existing within

electronic medical records hold the promise of reducing healthcare costs through comparative effectiveness studies,<sup>3</sup> reducing health disparities by informing policy decisions at the systems level,<sup>4</sup> and accelerating discovery by closing the gap in translational research.<sup>5,6</sup>

A limiting factor in taking full advantage of these networked information technologies, however, has been the organizational challenge of deriving agreement from scientific communities on the common terms, measurements, and data elements that will make up the content and structure of the interconnected data systems.<sup>7–9</sup> This paper describes one solution to that problem: the Grid-Enabled Measures (GEM) database. Whereas other systems have used consensus panels<sup>7</sup> and psychometric analytic techniques<sup>10</sup> to select common measures for very specific purposes, the GEM database is distinct in that it uses “web 2.0” functionality to solicit, comment, vet, and select measures from the behavioral and population science communities in open and transparent ways. It is an example of what some National Science Foundation (NSF) grantees have referred to as “Science 2.0.”<sup>11,12</sup>

---

From the National Cancer Institute, NIH (Moser, Hesse, Shaikh, Morgan, Augustson, Kobrin), Bethesda; Clinical Monitoring Research Program, SAIC-Frederick, Inc., NCI-Frederick (Courtney), Frederick; and Westat Inc. (Levin, Helba, Garner, Dunn, Coa), Rockville, Maryland

Address correspondence to: Richard P. Moser, PhD, National Cancer Institute, 6130 Executive Boulevard, MSC 7326, Bethesda MD 20892-7365. E-mail: moserr@mail.nih.gov.

0749-3797/\$17.00

doi: 10.1016/j.amepre.2011.01.004

## Web 2.0, Health 2.0, and Science 2.0

In 2004, computer publishing entrepreneur Tim O'Reilly hosted a conference of information technology specialists to identify characteristics of successful websites that seemed to be thriving in spite of the financial downturn associated with the “dot.com” implosion. He referred to this new generation of the web as “web 2.0” to provide a sense of “lessons learned” along with a forecast of rising trends. Among the most notable changes he foresaw in the emerging online ecosystem was a movement away from passive information dissemination (a hallmark of the mass media age) to active user participation. Wikis and blogs were beginning to dominate the Internet, as were sites for user-generated content such as YouTube, Flickr, MySpace, Delicious, and later Facebook, Four-square, and Twitter. Commercial vendors began to incorporate usage and other types of data into user-friendly evaluations of their products, and the Internet community began to value the notion of “collective intelligence” as valid input when making decisions.

These changes in the social ecology of online information systems did not escape the notice of informatics researchers and health scientists. Health-information seeking and active participation in health-related discussions rose substantially with the diffusion of the web during the first decade of the new millennium.<sup>13–15</sup> Frustrated by the slow movement of the traditional medical community, some patients began joining online participative sites, such as PatientsLikeMe, in which they could freely share data on their own personal recoveries with others with similar diagnoses in an online environment.<sup>16</sup> Advocacy groups began encouraging their members to participate actively in their own health care and to engage proactively in researchers' efforts to recruit for clinical trials.<sup>17</sup> Because of these noticeable changes, the Silicon Valley–based Institute for the Future predicted a trend toward “Open Health,” in which biomedical science, healthcare practice, personal responsibility, and community advocacy merged.<sup>18</sup> Gunther Eysenbach, editor of the online *Journal of Medical Internet Research*, concluded that medicine and health were moving toward a paradigm of “Medicine 2.0” or “Health 2.0” as patients and practitioners began to engage collectively in the new environs of web 2.0.<sup>19,20</sup>

Likewise, the nation's science institutions began reflecting on how the change in Internet culture might influence the conduct of science itself. In the spring of 2010, NSF convened a series of technical meetings to explore the phenomenon of “Technology Mediated Social Participation.”<sup>21</sup> “The World Wide Web is the ultimate open, global, collaborative communication medium,” explained a representative from Macmillan

Publishers in summarizing implications of the meetings, and “Science is the ultimate open, global, and collaborative human endeavor.”<sup>22</sup> Nevertheless, the culture of science often moves more slowly than the technology it helped create. The change is occurring; it is just not happening as quickly as we may like.<sup>22</sup>

Still, the changes that are occurring are impressive. In 2010 the National Aeronautics and Space Administration demonstrated how it could effectively take advantage of crowd-sourcing (i.e., engaging a full community through the web to solve a particular problem) to solve some of the more elusive computational challenges posed by long-range space flight.<sup>23</sup> Examples of “participatory research” sponsored by NSF have illustrated how individuals equipped with cell phones could contribute in parallel to document environmental health hazards.<sup>24</sup> Based on the potential of collaborative technologies to improve efficiencies in science, the President's Council of Advisors on Science Technology directed the federal government to invest “in new multi-agency NIT [Network Information Technologies] initiatives in areas of particular importance to our national priorities,” beginning with health and health care.<sup>25</sup>

## Open Science in Support of the National Institutes of Health

The need to marshal collective efforts in addressing the challenges of modern technology is readily apparent at the NIH. After succeeding in documenting the full human genome in 2003, the NIH community faced the task of identifying the precise connections between the DNA-sequenced base pairs and predictions for disease process. This link is necessary for advances in combating diseases such as cancer, cardiovascular disease, and neurologic disorders.<sup>26</sup> This “needle-in-a-haystack”<sup>27</sup> problem can be solved only by developing the cyberinfrastructure needed to connect the research community in focused, productive ways.<sup>28,29</sup>

To address the problem of data coordination at the National Cancer Institute, the largest institute within the NIH, biomedical informaticians embarked on an effort to create the next generation of computing infrastructure. Termed the “cancer Biomedical Informatics Grid,” or caBIG, the purpose of this new cyberinfrastructure was to connect the cancer research community in secure, interoperable, standardized, and semantically cohesive ways.<sup>30,31</sup> This new platform for cancer-related research would allow scientists to share protocols, operate in a harmonized way with standard measures, upload and share data, and obtain secure access to a virtual lattice of interconnected data sets. The spirit of the project matches the charge given to the secretary of the DHHS by the

Committee on Vital and Health Statistics in 2001: to “connect the dots” in biomedical and public health research.<sup>32,33</sup> Through its connection to cyberinfrastructure such as caBIG, the GEM database is a tool to help connect these dots and increase scientific productivity.

### **What Is the Grid-Enabled Measures Database?**

The purpose of the GEM database is to serve as a portal for health scientists who wish to take advantage of the benefits of Science 2.0 to accelerate scientific discovery. The GEM database has two overarching goals: (1) promote use of standardized measures, which are tied to theoretically based constructs; and (2) facilitate sharing of harmonized data resulting from the use of standardized measures. Although the process by which the GEM database achieves these goals is unique, the overall objectives are commensurate with other efforts such as consensus measures for phenotypes and eXposures (PhenX), the Patient-Reported Outcomes Measurement Information System (PROMIS), and the NIH Toolbox (see the section on The Changing Landscape).

The GEM database originated from strategic efforts to bring population science into the collective domain of network and information technologies, more commonly described as grid and cloud computing; in other words, using an Internet-based research workspace. Admittedly, the standard classifications guiding much of molecular medicine had been determined in the years preceding the massive infrastructure undertaking. Therefore, bringing the components of molecular medicine into semantic agreement was primarily a matter of assuring compliance with existing vocabularies and ontologies.

Bringing the components of population science into harmonization, however, was more challenging. By its nature, population science is made up multiple scientific disciplines aimed at different levels of analysis in solving health-related problems.<sup>34</sup> Epidemiologists traditionally look at variables distributed across different contextual environments in order to identify risk factors for disease processes and to determine optimal treatment approaches in clinical or public health practice. Behavioral researchers often look “intrapyschically” to determine how cognitive processes, emotional responses, and situationally based knowledge may interact in the prevention or control of disease. Until recently, these disciplines felt little urgency to agree on and share standardized measures. The task is daunting and few incentives have been in place.<sup>35</sup>

The landscape, however, is beginning to change especially for clinical researchers and those with a prevention focus. A good example is the “meaningful-use” require-

ment for determining eligibility for financial incentives as mandated by Congress in the Health Information Technology for Economic and Clinical Health Act of 2009.<sup>36,37</sup> Increasingly, standardized measures within the EMR will be used to assess patients’ prevention-related outcomes (smoking status, diet, nutrition, screening).<sup>38</sup> Selecting the right measures—measures that are both valid and practicable—will be a necessary first step in creating a new healthcare system that can be used to nudge healthier behavior by assessing system outcomes.<sup>39</sup>

### **Principles Guiding Development of the Grid-Enabled Measures Database**

The GEM database was designed to facilitate these standardization and sharing processes across scientific disciplines. Necessarily, the GEM database was built to be flexible enough to accommodate different types of measures (e.g., from self-report to biological). The system can accommodate independent and dependent variables across the health continuum, including prevention, diagnosis, treatment, and end-of-life issues, regardless of disease/wellness focus. Using principles of Science 2.0, GEM solicits community participation in contributing, vetting, and selecting measures for harmonizing data in a grid-enabled world. More importantly, it is a doorway—or portal—to the use of harmonized measures and creation of a more semantically integrated science.

As a result, the GEM database creates an environment for “prospective meta-analyses” in which research is designed for integration. Comparisons across studies can become the norm, rather than the result of the current, often awkward, retrofitting process. The purpose and structure of the GEM approach can be seen from the design of the online tool’s primary interface. The GEM database interface uses a tab-based architecture with three main tabs: constructs, measures, and data sets. Briefly, users can search for or add information about theoretically based constructs (including meta-data such as the definition and synonyms), find linked measures in a one-to-many relationship (along with associated meta-data such as author, reliability, and validity) and search for and share harmonized data that use any common measures (Figure 1).

The GEM database is based on four primary design principles: (1) architecture for participation—barriers to participation (or collaboration, in the scientific realm) are removed, that is, it is easy to “plug and play” (think of how easy it is to upload videos to YouTube); (2) data-driven decision making—decisions, such as whether to buy a stereo or choose a certain doctor, hospital, or self-report measure are based on objective data; (3) wisdom of the



then immediately find data sets that include these measures. At a minimum, this function allows researchers to review the type of data produced by a measure of interest. Depending on a researcher’s objectives and technical capabilities, a query could then be executed to bring together data from several data services all using common constructs and associated measures. This interoperability could open the door to new data-sharing opportunities within the scientific community. Very likely, novel statistical techniques will need to be applied. Ultimately this approach to common data collection and resulting sharing has the potential to accelerate the accumulation of knowledge (Figure 2).

### Templates Available for Uploading Meta-Data

From its inception, one of the core goals of GEM has been data sharing. As discussed previously, the GEM database shares meta-data via a public data service on the Grid and promotes the sharing of data through its “Datasets” tab. However, true data sharing is reciprocated. That is, the system shares its information but also receives information from outside sources. Since there are outside database systems that contain meta-data about measures and constructs, the creators of the GEM database wanted to make sure it could benefit from these existing repositories without requiring duplicate data entry. The ideal solution would be to have the GEM database import these meta-data using another data service on the Grid. However, at

this time, most outside sources of meta-data for measures and constructs are not connected to the Grid. These sources are primarily proprietary data belonging to individual investigators. Therefore, a file-based approach to importing data into the GEM database needed to be developed.

During the initial deployment phase, the GEM database supported the importation of information about constructs and measures via Microsoft Excel spreadsheets. Standardized import template spreadsheets were developed for both constructs and measures. Institutions wishing to contribute meta-data from an existing database used these templates to create import spreadsheets. These spreadsheets were then processed and the contents imported into the GEM database. Although the GEM database team successfully imported meta-data from several outside institutions using these Excel spreadsheets, automating the import process was challenging. The majority of the barriers to automation were related to handling data values that were not consistent with the database’s internal checks and balances. A future update will address these issues using Extensible Markup Language (XML), a schema for encoding documents in machine-readable form.

### Wiki-Like Processes to Edit Meta-Data

In the simplest terms, a wiki is an open, collaborative website where anyone can contribute (see for more infor-

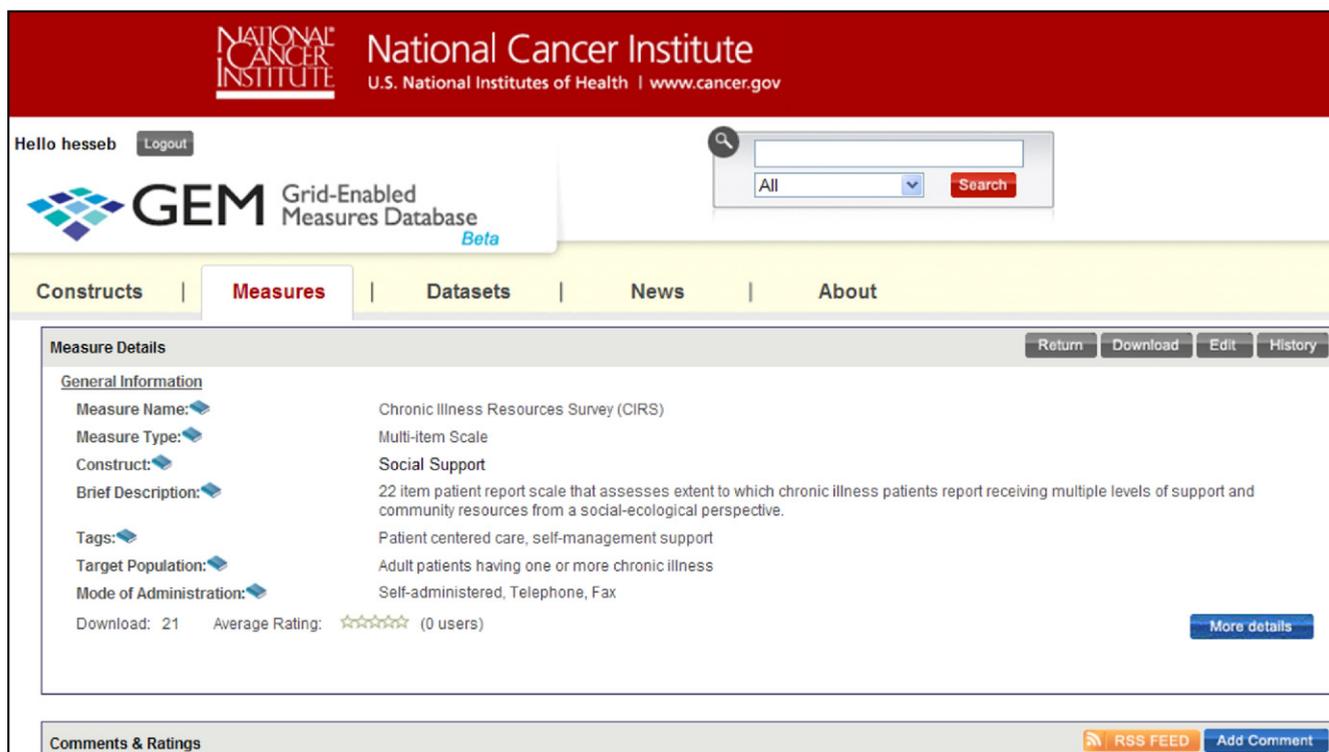


Figure 2. Illustration of meta-data for an individual measure, in this case a multi-item scale for social support

mation: [en.wikipedia.org/wiki/Wiki](http://en.wikipedia.org/wiki/Wiki)). It is in this spirit of enabling community-driven science that the GEM database was developed. It provides a means by which any registered user within the research community can contribute by adding meta-data about constructs, measures, and research data. The contributor does not need to be the author of a construct or measure in order to document it in the GEM database. Any registered user can make additions or changes to any meta-data in the database. But it cannot be done anonymously. To ensure a sense of accountability, the author of the additions/edits will be known.

Just as with other wikilike systems, the GEM database is dependent on the community of users to monitor the accuracy and validity of the information being contributed. Information found to be erroneous or suspect by the community can be discussed and subsequently corrected by the community itself. As a safeguard, the GEM database keeps a full audit trail so that the “who, what, and when” associated with each update is always available to every user. Despite these safeguards, it may seem to those not familiar with wikis that having a site where anyone who has an authorized login may contribute and edit existing information would lead to chaos and inaccurate entries. However, it has been demonstrated by *Nature* that the accuracy of Wikipedia when covering the subject of science is on par with that of the Encyclopedia Britannica.<sup>41</sup>

Although the GEM database promotes wikilike open authorship and editing of meta-data, it is more structured than a traditional wiki. For example, all contributed information must be entered into a fixed set of attributes instead of an open narrative format. These attributes are specific to constructs, measures or data sets. The attributes associated with measures include name, description, construct measured, target population, validity, and reliability among others. Also, some of these attributes will be required before a construct or measure can be submitted for public review. This process ensures a minimum level of completeness before the content is visible to the larger research community.

### Rating/Commenting on Measures

An important aspect of community-driven websites is communication among various members of the community. The ability to provide feedback and comment on important concepts gives participants a voice and encourages them to fully engage in the project. The GEM database allows for community feedback in two ways. First, users are able to post comments about any aspect of the constructs, measures, and data sets. These comments are visible at the bottom of each detail page. Commenting can be used to provide feedback to con-

tent authors and potential users of measures and data sets. In addition to general comments, the GEM database encourages researchers to rate measures using a 5-star rating system. The scale has no pre-specified meaning and instead will access the user’s evaluative gestalt of the measure. Users can always change their previous ratings, but each user can have only one active rating for a given measure so there is less concern that someone may “pad” the rating of their favorite measure. An average rating for each measure is calculated and made available through both the GEM database interface and other grid-enabled services such as the caGrid data service. This rating system allows a community of researchers to integrate data in the process of identifying measures that should become the standard.

### Data-Driven Decisions

Identification of one measure as better than others and a decision to promote for standard use should be data-driven. In the GEM database, users are given data to help with this process. These data include subjective outcomes such as averages and distributions of ratings, objective measures such as the number of times a measure has been downloaded from the database, and more traditional psychometric information such as a measure’s reliability and validity.

### Wisdom of the Masses

Though crowd sourcing has its detractors, and even proponents acknowledge that it doesn’t work effectively under all circumstances, the characteristics of the GEM database and its user community would suggest that the conditions for its use are optimal under the following conditions:<sup>40</sup> (1) diversity of opinion: The users of the database will likely come from a variety of academic disciplines and have different scientific interests; (2) independence of members: Users will not be collectively tied to any particular group or a particular way of seeing the world; (3) decentralization: Final decisions are not made by one or a few experts (top-down) but by the entire community (bottom-up); and (4) a good method for aggregating opinions: All input from all users can be found, summarized, and utilized by the community of users to make decisions. Although these factors do not guarantee the best measures will be identified, the circumstances are very supportive of that outcome, far more so than reliance on dissemination of findings by peer-reviewed publication alone. And, as the number of users increases, the more likely their collective wisdom will efficiently advance the field.

## Open Access

Open access is an essential component of the philosophy underlying the GEM database. Users can easily find meta-data about measures and data sets. The measures themselves can be downloaded easily (where publicly available), and data can be identified and accessed in several different formats that should facilitate data sharing. However, the move toward open collaboration works only insofar as those working together are willing to share information or data and are willing to sublimate their own concerns to a collective benefit. To achieve this openness, the scientific community must break down the barriers—real or perceived—that often hinder open sharing of information.<sup>42</sup> Concerns about sharing data such as confidentiality and HIPAA requirements need to be addressed. Some of these can be accomplished relatively easily by removing or obscuring any identifying information.

Other concerns, such as sharing data that a scientist may consider their property or fears of others trumping their results, may take longer to overcome. These barriers are particularly strong in academic centers where people compete for limited resources and promotions. Some of this can be accomplished only by embracing a new collaborative philosophy (Figure 3).<sup>43</sup>

## The Importance/Challenges of Creating Ontologies

To meet the goals of the GEM database, common understanding of terms is essential. An ontology is a way of representing a knowledge domain to enable “a shared and

common understanding that can be communicated between people and heterogeneous and distributed application systems.”<sup>44</sup> Development of a useful ontology requires that the terms are defined with sufficient context so if two people—or computer systems—use the same terms, there is complete and correct communication between them. A good example is the sharing of information through electronic medical records discussed earlier where two record systems need to be measuring the same construct in the same way with complete understanding of terms. Differences in the definition of terms used can make the goal of merging across data systems untenable.

There currently exist many efforts to reconcile definitions of terms such as Systematized Nomenclature of Medicine—Clinical Terms (SNOMED CT®) and Logical Observation Identifiers Names and Codes (LOINC®). These systems can potentially interact with the GEM database to further facilitate sharing of information. For example, if there were a relationship between a measure in the GEM database and a LOINC code that describes a term, the reference to the LOINC code could be a piece of meta-data for that measure in the GEM database. In essence, this would standardize terms across the two systems and would ensure the compatibility in terms. And by extension, any other data systems that use the LOINC codes could share information with the GEM database.

In the domain of behavioral research, an additional type of misunderstanding can occur when social scientists and clinicians use the same words to mean different

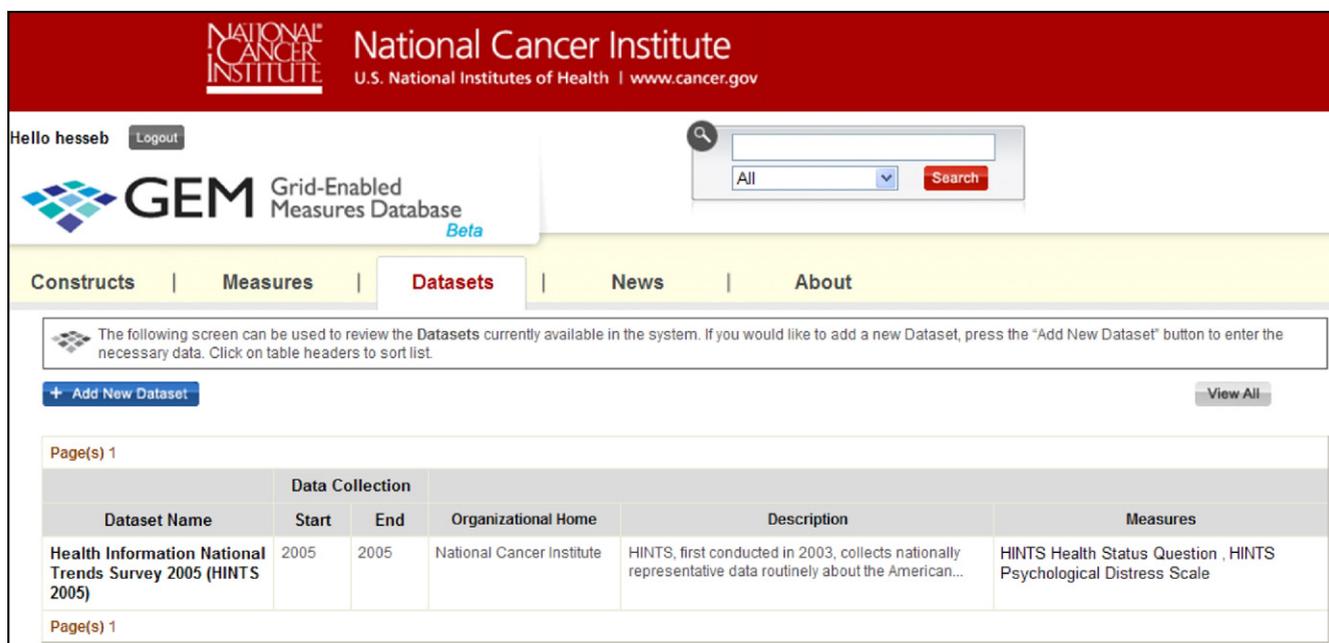


Figure 3. Information about publically available data sets ready for download through the Grid-Enabled Measures database

things. For example social and behavioral scientists often define depression as a self-reported mood state assessed with a self-report measure such as the Beck Depression Inventory. Conversely, for the clinician, “depression” may refer to a clinical state that is evaluated by a medical professional according to the DSM-IV criteria.

The solution to the problem of how to integrate such heterogeneous definitions of terms has been the development of semantic web technologies that can accomplish this integration by utilizing the semantics of the meta-data about the terms (context, provenance, and relationships to other meta-data) that are made explicit through the use of ontologies. One example of this is the Open Biological and Biomedical Ontologies website ([www.obofoundry.org/](http://www.obofoundry.org/)), which is a consortium of scientists with the goal of “. . . creating a suite of orthogonal interoperable reference ontologies in the biomedical domain.” In order for collaborators in behavioral research to move forward most efficiently, a similar ontology of the respective terms is necessary. Professional organizations such as the American Psychological Association have attempted to achieve this end through development of a thesaurus of terms, but this approach has limited usefulness if it is not adopted by a wider range of users.

### **The Grid-Enabled Measures Database As a Tool to Promote Integrative Data Analysis**

Using standardized measures and sharing harmonized data are useful only if they ultimately result in more-efficient science and improved outcomes. As a tool, the GEM database can facilitate this improved way of doing science through its ability to combine shared constructs and associated measures across independent data sets. Scientific discovery can advance more quickly by combining data to create a cumulative knowledge base—literally standing on the shoulders of giants—as opposed to the traditional way of doing science where scientists work independently and do not integrate information across studies. These data sets can then be analyzed using statistical techniques that have been termed integrative data analysis (IDA).<sup>45</sup>

Integrative data analysis is similar to meta-analysis in that both have the goal of combining information across studies. However, in meta-analysis, only summary statistics of empirical research (e.g., effect sizes) are analyzed, and in IDA the actual data are combined together, with the merged data set being analyzed as a whole. Generally, IDA is more likely to be hypothesis-driven. The GEM database facilitates both IDA and the use of other exploratory methods, such as data mining, which use advanced computational methods to look for unseen patterns

within a large data set. All of these techniques have been termed “discovery informatics” and make for meaningful use of secondary data.<sup>46,47</sup>

### **Providing Incentives for Use of the Grid-Enabled Measures Database and Addressing Issues Regarding Data Sharing**

Scientists, especially those in academia, typically do not have institutional incentives to agree on standard measures and to share data with other scientists. Academia tends to reward researchers for individual scholarly productivity, not those who work collaboratively and share resources. Historically, there has been little support to engage with others outside of one’s own discipline.<sup>48</sup> But this individualism leads to only a continued fractured scientific landscape. Using agreed-on measures and sharing data will require a change in thinking about how science is best conducted and a turn toward a collective mindset.

One way to promote this change is to appeal to the scientists’ interest in having a voice in setting the agenda for their respective fields. Using the GEM database can help scientists influence their colleagues to use a particular measure. Data sharing requires adopting a “you’ve got to give to get” mentality. While most scientists would be willing to receive data from others to enhance scientific discovery, requiring them to share data first may push them to think differently and more creatively. Many legitimate concerns exist; examples include concerns about participant confidentiality and other misuses of data. Under some circumstances, sharing data would be irresponsible or even illegal.

Institutional review boards may raise concerns about the ability to get informed consent or to identify respondents, and they will need reassurance that study participants will not be harmed in any way. For the majority of situations, though, concerns can be addressed. Often, what is needed is to change scientific cultural norms and expectations about collaboration<sup>49</sup> in addition to providing procedures and processes to address legal and/or ethical issues. For instance, dbGaP, a website to access to data from genomewide association studies ([www.ncbi.nlm.nih.gov/gap/](http://www.ncbi.nlm.nih.gov/gap/)), requires users to receive permission from a Data Access Committee before allowing access to restricted data. Those who submit data must provide a statement that they have met all applicable laws and regulations. Similarly, the GEM database will require users to follow existing data-sharing procedures using guidelines provided by the caBIG Data Sharing & Intellectual Capital Workspace ([https://cabig.nci.nih.gov/working\\_groups/DSIC\\_SLWG/](https://cabig.nci.nih.gov/working_groups/DSIC_SLWG/)).

## The Changing Landscape

The GEM database is certainly not the first tool of its kind to promote standardizing measures and sharing data. In fact, numerous other examples of both exist within and outside of government. Projects such as PROMIS,<sup>10</sup> PhenX, (<https://www.phenx.org/>; see Schad et al. this issue), and the NIH Toolbox (for the assessment of neurologic and behavioral functioning; [www.nihtoolbox.org/default.aspx](http://www.nihtoolbox.org/default.aspx)) have similar goals but use different processes. The GEM database tends to focus on the use of a broader community input to drive consensus (more bottom-up) rather than relying on a selected number of experts (more top-down) to make decisions. Despite these practical differences, the rapid proliferation of data sharing and standardizing projects indicates a clear paradigm shift in science. People are discovering the tremendous advantages of consensus—by sharing as opposed to hoarding data. The field is embracing an NIH (Not Invented Here) attitude.

The irony of the concurrent and independent development of several systems to bring consensus is not lost here. As a step toward ameliorating this situation, efforts are underway to import measures from some of these systems (e.g., PROMIS) into the GEM database. Continued effort is needed so that the lack of integration—and thus lack of interoperability—is lessened and not perpetuated.

Grid-enabled measures have the potential to help facilitate a new way of doing science that supports collaboration and sharing of information and data. This new approach, however, has many challenges and requires scientists to move from their comfort zone into a new scientific arena where the necessary incentives—both personal and systemwide—are not well understood or tested. Time will tell whether scientists are eager and able to meet these challenges in the name of advancing scientific discovery.

---

Publication of this article was supported by the National Institutes of Health.

No financial disclosures were reported by the authors of this paper.

---

## References

- Langenberg DA. Information technology and the conduct of research: the user's view. Washington: National Academies Press, 1989.
- Stokols D, Misra S, Moser RP, Hall KL, Taylor BK. The ecology of team science: understanding contextual influences on transdisciplinary collaboration. *Am J Prev Med* 2008;35(2S):S96–S115.
- Buetow KH, Niederhuber J. Infrastructure for a learning health care system: CaBIG. *Health Aff (Millwood)* 2009;28(3):923–924; author reply 924–25.
- Gibbons MC. eHealth solutions for healthcare disparities. New York: London: Springer, 2007.
- Etheredge LM. A rapid-learning health system. *Health Aff (Millwood)* 2007;26(2):w107–18.
- Khoury MJ, Gwinn M, Yoon PW, Dowling N, Moore CA, Bradley L. The continuum of translation research in genomic medicine: how can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention? *Genet Med* 2007;9(10):665–74.
- Stover PJ, Harlan WR, Hammond JA, Hendershot T, Hamilton CM. PhenX: a toolkit for interdisciplinary genetics research. *Curr Opin Lipidol* 2010;21(2):136–40.
- Psaty BM, O'Donnell CJ, Gudnason V, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet* 2009;2(1):73–80.
- President's Council of Advisors on Science and Technology. Realizing the full potential of health information technology to improve health-care for Americans: the path forward. Washington: Executive Office of the President, 2010.
- Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol* 2010;63(11):1179–94.
- Waldrop MM. Science 2.0. *Sci Am* 2008;298(5):68–73.
- Shneiderman B. Computer science. *Science* 2008;319(5868):1349–50.
- Hesse BW, Arora NK, Beckjord EB, Finney Rutten LJ. Information support for cancer survivors. *Cancer* 2008;112(11S):S2529–S2540.
- Hesse BW, Moser RP, Rutten LJ. Surveys of physicians and electronic health information. *N Engl J Med* 2010;362(9):859–60.
- Fox S. *Cancer 2.0: a summary of recent research*. Washington: Pew Research Center, 2010.
- Hesse BW, Hansen D, Finholt T, Munson S, Kellogg W, Thomas JC. Social participation in health 2.0. *IEEE Computer* 2010;43(11):45–52.
- Ferguson T, Group e-PSW. E-patients: how they can help us heal healthcare. Princeton NJ: Robert Wood Johnson Foundation, 2007.
- Institute for the Future. *Open Health Toolkit: innovation in the global economy*. Palo Alto CA: Institute for the Future, 2008.
- Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. *J Med Internet Res* 2008;10(3):e22.
- Hughes B, Joshi I, Wareham J. Health 2.0 and Medicine 2.0: tensions and controversies in the field. *J Med Internet Res* 2008;10(3):e23.
- Pirolli P, Preece J, Shneiderman B. Cyberinfrastructure for social action on national priorities. *Computer* 2010;43(11):20–1.
- Hannay T. What can the web do for science? *Computer* 2010;43(11):84–7.
- Globe Staff. NASA teams up with InnoCentive. *Boston Globe* 2010 Jan 15. [http://www.boston.com/business/ticker/2010/01/nasa\\_teams\\_up\\_w.html](http://www.boston.com/business/ticker/2010/01/nasa_teams_up_w.html).
- Burke J, Estrin D, Hansen M, et al. Participatory sensing. Paper presented at: WSW'06 at SenSys2006; Boulder CO, Oct 31–Nov 3, 2006.
- President's Council of Advisors on Science and Technology. *Designing a digital future: federally funded research and development in networking and information technology*. Washington: Executive Office of the President, 2010.
- Collins FS. NIH in the 21st century: the director's perspective. Subcommittee on Health; Committee on Energy and Commerce. Washington: NIH, 2010.
- Braun R, Rowe W, Schaefer C, Zhang J, Buetow K. Needles in the haystack: identifying individuals present in pooled genomic data. *PLoS Genet* 2009;5(10):e1000668.
- Buetow KH. Cyberinfrastructure: empowering a “third way” in biomedical research. *Science* 2005;308(5723):821–4.
- Buetow KH. An infrastructure for interconnecting research institutions. *Drug Discov Today* 2009;14(11–12):605–10.

30. Hesse BW. Of mice and mentors: developing cyberinfrastructure to support transdisciplinary scientific collaboration. *Am J Prev Med* 2008;35(2S):S235–S239.
31. von Eschenbach AC, Buetow K. Cancer informatics vision: caBIG. *Cancer Inform* 2007;2:22–4.
32. National Committee on Vital and Health Statistics. Information for health: a strategy for building the National Health Information Infrastructure. Washington: DHHS, 2001.
33. Forman MR, Greene SM, Avis NE, et al. Bioinformatics: tools to accelerate population science and disease control research. *Am J Prev Med* 2010;38(6):646–51.
34. Abrams DB. Applying transdisciplinary research strategies to understanding and eliminating health disparities. *Health Educ Behav* 2006;33(4):515–31.
35. Breckler S. Psychology and the grid. American Psychological Association Science Directorate Newsletter. Washington: American Psychological Association, 2008.
36. Blumenthal D. Promoting use of health IT: why be a meaningful user? *Conn Med* 2010;74(5):299–300.
37. Blumenthal D, Tavenner M. The “meaningful use” regulation for electronic health records. *N Engl J Med* 2010;363(6):501–4.
38. Hesse BW. Time to reboot: resetting health care to support tobacco dependency treatment services. *Am J Prev Med* 2010;39(1S):S85–S87.
39. Hesse BW, Woods SS, Ahern DK. Nudging best practice: the HITECH Act and behavioral medicine. *Translational Behavioral Medicine*. 2011; 1(1).
40. Surowieski J. *The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economics, societies, and nations*. New York: Doubleday, 2004.
41. Giles J. Internet encyclopaedias go head to head. *Nature* 2005; 438(7070):900–1.
42. Gadaleta E, Lemoine NR, Chelala C. Online resources of cancer data: barriers, benefits and lessons. *Brief Bioinform* 2011;12(1):52–63.
43. Committee on Science Engineering and Public Policy (U.S.). Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age. Ensuring the integrity, accessibility, and stewardship of research data in the digital age. Washington: National Academies Press, 2009.
44. Lopez JM, Gil R, Garica R, Cearreta I, Garay N. Towards an ontology for describing emotions. *Lecture Notes in Computer Science* 2008;5288:96–104.
45. Curran PJ, Hussong AM. Integrative data analysis: the simultaneous analysis of multiple data sets. *Psychological methods* 2009; 14(2):81–100.
46. Agresti WW. Discovery informatics. *Commun ACM* 2003;46(8):25–8.
47. Safran C, Bloomrosen M, Hammond WE, et al. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. *J Am Med Inform Assoc* 2007;14(1):1–9.
48. Nash JM. Transdisciplinary training: key components and prerequisites for success. *Am J Prev Med* 2008;35(2S):S133–S140.
49. Kleppner D, Sharp PA, Whitesides GM. Ensuring the integrity, accessibility, and stewardship of research data in the digital age. Washington: National Academies Press, 2009.

**Did you know?**

You can search over 500 top health sciences journals online, including MEDLINE, via the *AJPM* website.

Visit [www.ajpm-online.net](http://www.ajpm-online.net) today!