



Stanford Center for
Biomedical Informatics Research

Connection

Spring 2021

What's Inside

- Controlled Terminologies in Africa
- Open Research Tools
- Vascular Surgery & Informatics
- Tackling Biases
- Promoting Data Sharing
- Tuberculosis

We Connect Data to Health

Encouraging Collaboration through Our New Website

If you haven't visited BMIR's [website](#) since we launched a new version on April 30, I urge you to check it out.

Our new site offers an improved user experience, one that provides clear information about BMIR's mission and research.

We've redesigned the site to improve navigation. Our research groups are now organized into categories that correspond with our

researchers' areas of focus. That makes it much easier to quickly see each individual's field of interest and special expertise. The site also features more complete profiles of our faculty.

We believe this new site highlights the achievements of BMIR, makes our tools and services easy to find, and demonstrates our connection to the clinical enterprise.

Finally, we hope the new site encourages future collaboration, which is at the heart of how BMIR advances informatics research here at Stanford and across the globe.

Mark Musen, MD, PhD
Director, Stanford Center for Biomedical Informatics Research

TRAINING

Every Monday you'll find [CEDAR](#) Project Manager John Graybeal on a Zoom videoconference call with up to two dozen health care data specialists on the African continent. On any given call, participants might include providers at stand-alone health clinics, developers creating systems for various health care delivery locations, researchers, and senior leaders of health care organizations and universities in countries like Kenya, Uganda, Ethiopia, and Zimbabwe.

The calls are part of a Virus Outbreak Data Africa Network ([VODAN](#)), organized by the [GO FAIR](#) initiative to enable distributed access to the critical data needed from Africa and the rest of the world to fight and contain the COVID-19 pandemic. Dr. Mirjam van Reisen of Leiden University Medical Center leads the VODAN initiative, coordinating its technical, political, and team-building activities.

Graybeal's job is to consult with the local health care data experts as they implement controlled terminologies and metadata templates reflecting the virus reporting requirements of their participating health facilities and organizations.

Expanding Adoption of Controlled Terminologies and Metadata in Africa



While medical care is advancing in Africa, the use of pen and paper is often the primary method for keeping medical records, which makes data virtually unmanageable for reporting purposes. And most reporting systems in Africa are implemented top-down, eliminating the opportunities for local personnel to adapt and re-use their own reported data.

VODAN was envisioned as a way to build a more equitable and sustainable digital health infrastructure so that technologists and health centers in African countries can

take greater ownership of their data and their technology.

Graybeal's [overview of basic vocabulary development](#) is a practical, minimalist tutorial to help the African participants start working with controlled terminologies fast.

Graybeal explained why he is excited for BMIR to be involved in the VODAN initiative: "Our powerful technology can be useful anywhere, and it can be accessed by anyone to meet their needs for managing their data descriptions. Being an active participant in GO FAIR led to our involvement in the VODAN initiative, which enables our technology to be shared more widely across the continent of Africa."

"That increased visibility is certain to lead to adoption by a broader group of organizations, and that will increase the number of people who are able to use, maintain, and improve our technology. It will also invariably lead to the creation of a community centered around our technology, which is what happened with our [Protégé](#) and [Bio-Portal](#) software. That creates an ecosystem that benefits everyone," he said.

BMIR is working with the [Research on Research Institute](#) (RoRI) and an international consortium of research funders, academics, and technologists to develop new open research tools under the banner ‘FAIRware’. The project is supported by a group of five research funders: the Wellcome Trust (UK), the Austrian Science Fund, the Canadian Institutes of Health Research, the National Institute for Health Research (UK), and the Swiss National Science Foundation.

A request for proposals originally issued by RoRI asked for a tool to help identify whether data sets online are compliant with the FAIR principles, verifying that the data are Findable, Accessible, Interoperable, and Reusable.

“Although RoRI was asking for a tool that would assess the FAIRness of data sets after they were already created, we pointed out that what really matters is making sure that scientists have a tool to guide them to create FAIR data from the start,” said BMIR Director Mark Musen, MD, PhD.

“[CEDAR](#) is that tool,” he added. He indicated that this was the first time in his

career that he received a research contract by proposing to do something different from what was actually mandated in the request for proposals. RoRI, however, actively encouraged proposals that might take a different point of view.



“The CEDAR tool allows a scientist to create metadata that are comprehensive, rich, and based on community standards. By using CEDAR, grantees of the FAIRware Funders Group will be guaranteed that their corresponding data sets will be FAIR,” Dr. Musen said.

During the next year, BMIR will focus on three activities:

1. Designing the metadata needed to ensure FAIRness for several areas of science by holding Metadata for Machines (M4M) workshops.

2. Using CEDAR to give comprehensive support to investigators authoring metadata to annotate experimental datasets, software, and other digital research objects.
3. Designing and building a FAIRware Workbench that will assess metadata quality and that will offer concrete steps for enhancing those metadata.

Dr. Margreet Bloemers is project leader for FAIR Data & Data Management at the Netherlands Organization for Health Research and Development ([ZonMw](#)), the Dutch equivalent of the National Institutes of Health. In commenting on BMIR’s proposal to the FAIRware Funders Group, Dr. Bloemers said, “A tool like CEDAR, combined with M4M workshops for defined research communities, is the way to go... We want to provide added value for research. With the metadata schemes for research communities, we contribute to the FAIRness of research outputs. Researchers benefit from this, and—after all—funders have a way to review the output from the research project!”

CLINICAL CARE

Combining Informatics Research and Vascular Surgery

As both a vascular surgeon and a biomedical informatics research scientist, Elsie Ross, MD, MSc, brings a unique perspective to both surgery and informatics.

She splits her week evenly between clinical work and research. Mondays are spent



mainly in the operating room, and on Tuesdays, Dr. Ross is in the clinic. Wednesdays are her “hybrid” days – with clinical activities in the morning, and research activities in the afternoon.

Thursdays and Fridays are full research days that include writing, a meeting with those who share her interest in machine learning applied to electronic health record data, and a lab meeting “focused on precision medicine for vascular disease, which is an outgrowth of work I performed while serving as a biomedical informatics postdoctoral fellow in the lab of Nigam Shah,” Dr. Ross said.

It’s highly unusual to combine two such distinct disciplines, and one of Dr. Ross’s recent grant proposals illustrates how she melds the two.

“A lot of work in applying machine learning to electronic health records has focused on trying to understand or predict medical

issues, but there has been less work done on the *surgical* side,” she said.

“Vascular surgeons are particularly interested in outcomes after inserting a stent or doing a bypass. We want to know if that repair is going to be durable, and I don’t know that you can get all that information from just the health records themselves – it probably requires understanding some imaging characteristics. Can we apply deep learning analysis to the images that you get during that procedure to then predict durability?” she asked.

“So my proposal is aimed at trying to understand how to apply machine learning/deep learning to images to get some kind of automated analysis that can then be used to predict whether revascularization (inserting a stent or doing a bypass to improve blood flow) intervention is durable,” she added.

Her unique expertise in both fields will undoubtedly pay rich rewards.

Modern evidence synthesis is a major focus for Maya Mathur, PhD, who joined the QSU faculty as assistant professor in 2019.

"Researchers are increasingly, and rightly, concerned about the credibility of the published literature. For example, published studies could be filtered by publication bias, or they could have internal biases like confounding. So it's critical to try to develop methods for evidence synthesis that don't take the published literature at face value but rather that try to adjust in a principled way for biases," she said.

Several of her recent [papers](#) address that by developing statistical methods that do a better job of adjusting for these biases than previous methods. One example of publication bias is where the literature selectively publishes results that are positive and that look good. In other cases, bias can occur within the study itself, such as confounding

that might take place in an observational study that isn't randomized, for instance.

Something that attracted Mathur to the QSU is the freedom given to faculty "to explore methodological research directions that benefit the departments that we are affiliated with," she said.

In her case, she is primarily affiliated with pediatrics and the maternal and child health research unit. Within pediatrics, many health conditions are quite rare, which means that most studies are pretty small with low statistical power.

"That's a situation where it's especially important to be able to synthesize evidence across studies, and the QSU has supported me by giving me the freedom and responsibility to pursue methods research that is applicable in this case," she said.

Apart from her methodological research, Mathur is also pursuing empirical research

around behavior interventions, with a current focus on interventions that would reduce consumption of meat.

"That's been really fun because I've been able to develop partnerships with some great folks here at Stanford as well as in nonprofits whose expertise nicely complements mine," she said.

She described the QSU as a very rich and vibrant environment where biostatisticians of differing levels of seniority are able to partner with one another. Sometimes that could be a mentoring situation between a senior and junior statistician.

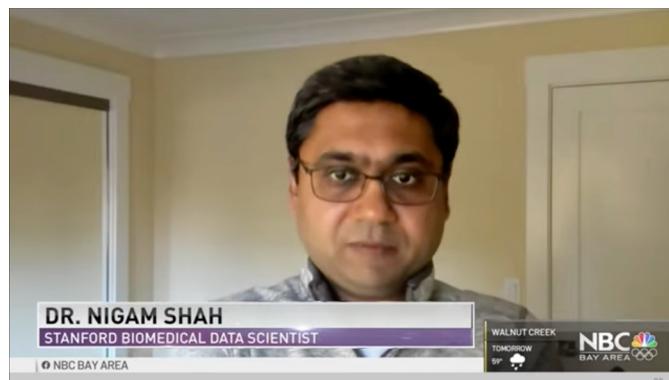
That was true in Mathur's own circumstance, when QSU Director Manisha Desai, PhD, served as "an amazing mentor to me during my first job at the QSU. She has always put a lot of emphasis on making sure that more junior members have a chance to learn new statistical methods," she said.

COLLABORATION

Promoting the Value of Industry-wide Data Sharing

During the early days of the coronavirus pandemic, Stanford Medicine took only 48 hours to develop temporary policies to reduce the load on the ICU at Stanford Hospital. Soon after that, Kevin Schulman, MD, who led the effort to develop the policies, approached Nigam Shah, MBBS, PhD, who leads the [Green Button](#) project, which is aimed at providing physicians with the information they need to make evidence-based decisions. Dr. Schulman sought help with a review of four months of data to assess whether the new suggested policies were in alignment with what the data was saying.

To do that, and to make sure the data were useful, Dr. Shah and team members Allison Callahan, a BMIR scientist, and Birju Patel, a clinical informatics fellow, worked with the [Technology and Digital Solutions Group](#) and the STanford Research data Repository ([STARR-OMOP](#)) database. STARR-OMOP is a clinical data warehouse (supported by Spectrum – the Stanford Center for Clinical and Translational Research) that updates patient data on a weekly basis for supporting such research.



Callahan and Patel provided evidence to answer multiple questions such as: *Compared to patients not discharged on oxygen, were patients discharged with home oxygen more likely to be readmitted within 30 days?* (The evidence showed that readmission rates were not higher for patients discharged with home oxygen, thus supporting the policy of allowing discharging with home oxygen.)

While the current database was only drawing from, and serving, patients in Stanford Healthcare, earlier in the pandemic, Dr. Shah proposed [a vision of sharing data among all local hospitals](#) to shape best prac-

tices based on learnings from a much larger database.

Dr. Shah, who was interviewed earlier this year for a related [story](#) on NBC Bay Area, observed that "doctors use information of prior patients to care for the next one, but the information they use is the information they remember from their own experience." With the help of data from a much larger base of patients, Dr. Shah explained that the system could then "see millions of patients and I can answer questions that no human doctor would be in a position to answer because it's just not possible to hold the collective experience of a million patients in your head."

"Pretty much every other industry that I can think of uses information on what they have done in the past to make their services better, and I think it's time that health care be able to do that," he added.

TUBERCULOSIS

Khatri Lab Translates a Gene Signature in a Point-of-Care Test

Purvesh Khatri, PhD, and his team at BMIR have transformed their genomic methods for diagnosing tuberculosis (TB) into a point-of-care cartridge that has been shown to work in the real-world patient population.

In less than five years, the Khatri lab moved from discovery to point-of-care, a process that would typically take a decade.

Khatri's team identified three genes that, in people infected with TB, tend to be activated. The team made the discovery by methodically mining data from publicly available datasets from gene-expression experiments performed by other investigators.

Dr. Khatri and his group published a [paper](#) in 2016 showing that a three-gene signature can diagnose TB using peripheral blood. Two years later, he and his team prospectively [validated](#) that the three-gene signature can diagnose and predict progression to TB. Several international groups in China and England independently [validated](#) that the three-gene signature indeed predicts and diagnoses TB.

The following year, in 2019, Cepheid, a commercial company, in partnership with the Khatri lab, translated this academic work into a point-of-care test for clinical use. The Cepheid point-of-care cartridge can measure three genes in a very precise manner using a completely automated device. The results are borne out by a [study](#)

led by the Foundation for Innovative New Diagnostics that was published in 2020 in the *Journal of Clinical Microbiology*.

"That paper illustrates the multiple stages leading to development of this particular product," Dr. Khatri said.



"First you must pick a problem and identify variables that you need to measure in a clinic so you can be confident that the diagnosis is correct. That was the bulk of the work performed in our lab between 2016 and 2018.

"Next you must transfer that to a point-of-care cartridge, which we did through a partnership with Cepheid. As described in

the paper, we showed that the cartridge does exactly what we want it to do," said Dr. Khatri.

The tuberculosis three-gene signature is one of five projects in the Khatri lab that are moving toward clinical translation. All are being done rapidly – in a four to five year time span – across multiple diseases.

"Our work is focused on accelerating clinical translation – not waiting 10 years to get something into clinical use," Dr. Khatri said.

With regard to the tuberculosis work, Cepheid will perform more testing to show that the cartridge works in different patient populations. The product will then be submitted for regulatory approval.

The World Health Organization has set certain performance criteria toward a goal of ending tuberculosis by 2035.

"Our three-gene signature was the first diagnostic to meet that target profile and the only method that translated into a point-of-care cartridge," said Dr. Khatri.

One of the most virulent diseases, it is estimated that tuberculosis has killed more than one billion people over time.

"Because of its contagious nature, it is important to diagnose active tuberculosis accurately and quickly, which is still not possible. It is hoped that the point-of care cartridge will change that," Dr. Khatri said.



BMIR

Stanford Center for
Biomedical Informatics Research

WE CONNECT DATA TO HEALTH

Connection
Spring 2021

Stanford Center for Biomedical Informatics Research
1265 Welch Road, Stanford, California 94305-5479
<https://bmir.stanford.edu>

The Stanford Center for Biomedical Informatics Research (BMIR) uses advanced research techniques to discover, apply, translate, and organize data that make a difference for health and health care. With its expertise in clinical and translational informatics research and biostatistics, the division works to uncover new ways to ad-

vance personalized medicine and to enhance human health and wellness.

Collaboration is in our DNA. We are excited about the prospect of working with other experts who share our goal to connect data to health and medicine. We encourage you to contact Mark Musen, Director of BMIR (musen@stanford.edu), to learn more.