Computing Community Consortium (CCC) Response to National Institutes of Health Request for Information (RFI): Strategic Opportunities and Challenges for the National Library of Medicine

October 2020

Dan Lopresti (Lehigh University/CCC Vice Chair), Lawrence Hunter (University of Colorado Denver), Beth Mynatt (Georgia Institute of Technology), and Shwetak Patel (University of Washington)

Science

We applaud NLM's focus on "curation at scale" and the creation of computable libraries of data, models and literature. We imagine a future when published papers shift to vetted models as scientists, practitioners and the general public look to NLM models for the latest knowledge in cancer care, for example. However, for this approach to be successful, NLM must also provide paths for new knowledge processes with metadata so that the trajectory of these models can be indexed in evolving literature. More fundamentally, these models should also delineate gaps explicitly. What are the "known unknowns" that bring to the forefront potential data biases through the absence of data based on gender, race, geography and more?

It will also be important to develop new methods that leverage and integrate the vast quantities of knowledge embodied in the literature with new data that is rapidly flowing into the system (e.g., data associated with the COVID-19 pandemic), placing each on an appropriate equal footing. This also raises interesting and important questions in the areas of privacy and consent and how they are adapted under critically urgent circumstances such as the current pandemic.

One significant concern in the Strategic Plan is the categorization of algorithms as "technology" and not as "science." While a catch-all term, "algorithms" includes an incredibly broad range of fundamental research and techniques that has powered the global data revolution, from large-scale search, to heterogenous data mining, to network analytics, to computer vision, to natural language understanding, to advanced AI and machine learning including issues of bias and fairness. The underlying assumption seems to be that other entities will develop these algorithms and that NLM will use them, perhaps adapting them to specific

applications. But NLM should take the lead in supporting research to develop these new algorithmic methods. As the curator of tremendously valuable data maintained for social good, and possessing the trust of the public, NLM is in a unique position that is undeniably distinct from the industrial applications pursued for profit by wealthy tech companies. NLM occupies a position that should be exploited, not ceded to others.

Technology

As the NLM plan notes, clinical care is migrating from hospital to home. The scientific path for this evolution must wrangle with the challenges of incorporating data from homes into clinical care systems. This messy data requires new AI approaches to discern signal from noise and to mitigate interruptions in data streams. Associated research issues and investment opportunities are identified in the CCC "Aging in Place" workshop report (https://cra.org/ccc/wp-content/uploads/sites/2/2015/05/CCC-AiP-White-Paper-.pdf) and the March 2016 PCAST report titled "Independence, Technology, and Connection in Older Age" (https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_independence_tech_aging_report_final_0.pdf)

In addition, we see opportunities for increased investment in sensing, screening, and triage tools to produce the data which we are learning to profitably mine.

Public Health, Consumer Health, and Outreach

NLM needs to drive socio-technical approaches to reduce health disparities, including determining the breadth of data to collect so scientists can understand larger patterns in health disparities and outcomes, and so that care providers better utilize broader socio-economic information in their practice, as outlined in the CCC "Sociotechnical Interventions for Health Disparity Reduction" workshop report (https://cra.org/ccc/wp-content/uploads/sites/2/2018/01/17602-CCC-Health-Disparities-ReportFinal.pdf). Time-stamped geographic data for patient homes will open up new areas of discovery regarding social determinants of health and should include economic data and psychosocial data (e.g., measures of stress). Clearly privacy protections are paramount and need to be developed alongside these more holistic discovery and care paradigms.

As we are now witnessing, approaches to enable more efficient and effective contact tracing are vitally important and would benefit from increased investment in the associated data science and AI research. This could permit contact tracing to take into account significantly more social, geographical, and health-related context than what is possible now, making it a more powerful tool for dealing with pandemics.

Scholarly Communication

We encourage the NLM to produce information that is "app friendly" so current knowledge can be consumed by web and mobile applications that seek to provide personalized and adaptive support for

patients (e.g., MyPath as described "Improving Cancer-Related Outcomes with CONNECTED HEALTH" President's Cancer Panel Report, 2016). One possible challenge to ponder: how could a conversational assistant (e.g. Alexa) answer questions for a newly diagnosed cancer patient starting treatment based on NLM curated information?

Likewise, communication approaches need to recognize and mitigate distrust between previously marginalized, stigmatized and communities abused by the medical system. To that end, the NLM should foster meaningful community partnerships so that these communities can shape, amplify, and at times combat healthcare communication. In doing so, NLM will model for others how community partnership can improve health outcomes and decrease disparities for underserved populations. See, for example, the recommendations arising from the CCC workshop on "Research Opportunities in Sociotechnical Interventions for Health Disparity Reduction" (https://cra.org/ccc/wp-content/uploads/sites/2/2018/01/17602-CCC-Health-Disparities-ReportFinal.pdf).

Perspectives, Practices, and Policies

We encourage NLM to promote science and medical practice to accelerate research in "intervention-generated inequality" (IGI) to recognize and prevent further harm caused by well-intentioned digital health approaches (see "Good intentions are not enough: how informatics interventions can worsen inequality," https://doi.org/10.1093/jamia/ocy052). Questions of IGI should stand alongside concerns in reproducibility and bias as more and more of healthcare practices shift to online and mobile platforms.

The pandemic has materialized serious health equity issues among large classes of the workforce, including gig workers, essential businesses that need to operate in person, etc. More attention needs to be focused on how we can better support these vulnerable members of our society.