October 16, 2020

Patricia Flatley Brennan, RN, PhD
Director
National Library of Medicine
National Institutes of Health
9000 Rockville Pike
Bethesda, Maryland 20892

Request for Information (RFI): Strategic Opportunities and Challenges for the National Library of Medicine (NLM), National Institutes of Health (NIH)

Dr. Brennan:

The American Medical Informatics Association (AMIA) appreciates the opportunity to provide input on strategic opportunities and challenges facing the NLM.

Health Informatics is the science of how to use data, information, and knowledge to improve human health, the delivery of health care services, and the execution of scientific research. AMIA is the professional home for more than 5,500 informatics professionals, representing frontline clinicians, biomedical researchers, public health experts, and educators who bring meaning to data, manage information, and generate new knowledge across the healthcare system and research enterprise. AMIA members advance health and wellness by implementing and evaluating informatics interventions, innovations, and public policy across settings and patient populations, adding to our collective understanding of health in the 21st century through peer-reviewed journals and scientific meetings.

AMIA strongly supports NLM's mission to acquire, organize, curate, and disseminate health-related information across the United States and the world. We view these as inter-related functions within the core of what the NLM is and we recommend that NLM view responses to this RFI through a prism that determines whether stakeholder responses further, or detract from, this remit. Of course, the health informatics community has much to contribute to these core functions, and several relevant initiatives from within our diverse community are highlighted across the RFI's sections and reproduced below in Table 1. Specifically, we highlight informatics-driven efforts to:

- **Develop standards to facilitate use of clinical data for research and improving interoperability**

  - Project Vulcan and similar efforts: Project Vulcan is an HL7 FHIR Accelerator project that brings together stakeholders across the clinical care, translational research, and clinical research communities to facilitate the use of and democratize clinical data for research purposes.[i] Efforts like Project Vulcan are incredibly important and whether (or not) it is successful, accomplishment of the group's goals is imperative. The value and volume of clinical data will only increase, so developing a process to make such data facile for research purposes should be a primary focus for NLM support.
  - Coordinate and support standards R&D around terminologies and code sets: NLM should invest in terminology management support beyond its current code sets. Resources are separate and guidance for terminology use is limited for those developing technology that utilize them. We recommend NLM facilitate a model throughout NIH to bring high quality standards and terminologies to development of Content Implementation Guides. Additional stakeholders, such as NIDDK, AHRQ, and CDC, which has launched MEDMORPH – an effort to develop computable clinical guidelines – should be included, as well.
- **Foster computable and executable knowledge artifacts and a knowledge ecosystem**
  - Several efforts, including Mobilizing Computable Biomedical Knowledge (MCBK), [ii] FHIR Clinical Guidelines (CPG-on-FHIR)[iii] and Evidence-Based Medicine Knowledge Assets project (EBMonFHIR),[iv] and the AHRQ evidence-based Care Transformation Support (ACTS) Initiative are working to develop computable knowledge artifacts via computable standards as part of a knowledge ecosystem. These, and other similar efforts, seek to deliver evidence back to the point of care electronically and in executable ways to facilitate decision-making. We recommend the NLM consider these efforts as use cases in addressing the 'core of NLM's service model' as it relates to meeting the biomedical information needs of all who participate in this work.
- **Advance how scholarly contributions are recognized**
  - NLM should also leverage its position to help the NIH address systemic inhibitors to open science. We lack incentives, or have perverse incentives, when it comes to data sharing and academic promotion related to development of useful datasets and software. Important work to finalize 42 CFR Part 11 for depositing data to ClinicalTrials.gov should be supported by similar requirements for research funded by NLM. Similarly, NLM should credit scholarship that develops useful datasets and software as a catalyst to overcome systemic disadvantages for such aspects of research in academia and other non-profit research entities.

Beyond these important initiatives, AMIA recommends the NLM ensure its 2017-2027 strategic goals are met by (1) establishing evaluation criteria for its programs vis-à-vis these core functions described in its mission and (2) strive to be the intellectual and programmatic epicenter for these functions across all NIH institutes and centers. By serving as lead acquirer, organizer, curator, and disseminator of health-related information across the NIH, the NLM will be

AMIA | American Medical Informatics Association
4720 Montgomery Lane, Suite 500 | Bethesda, Maryland 20814

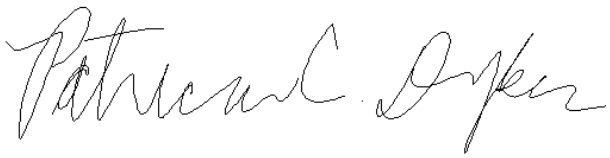
positioned to both advance the necessary suite of standards, processes, and technologies necessary to service all NIH ICOs and improve data-driven discovery across the entire extramural research enterprise.

In addition to these core library functions, the NLM should continue funding advanced biomedical informatics and data science research through pre- and post-doctoral training. AMIA recommends NLM prioritize multi-disciplinary training programs. AMIA members note that NLM has an opportunity to ensure its products and services are more broadly applicable to biomedicine because it is not focused on a singular disease. This quality gives NLM the capacity to leverage what other Centers and Institutes do, by abstracting resources that can be generalized and shared. This same characteristic of NLM naturally encourages multi-disciplinary, team-based training, which should be embraced through increased funding opportunities, as well as curriculum development beyond PhDs. We note that almost none of what we recommend in the table below can happen without advanced informatics expertise, and these experts are often the ones responsibe for brining in grants to their respective institutions. Given the rapidity of growth and interest in the fields of data science, open science and bioinformatics, NLM should allow for more frequent opportunities to apply for funding and it should more frequently seek ways to partner with other NIH Centers & Institutes on joint training opportunities. Additionally, NLM should convene non-profit partners, such as IMLS, MLA and AMIA to create bridges between informatics, librarianship, and health professions to fund training programs at the associate's degree (e.g. community college health professions), bachelor's degree, and Master's degree levels. These training programs should leverage role-based informatics competencies.

Finally, the NLM should work with AMIA, the MLA, the Friends of NLM, and others to establish an advocacy agenda focused on promoting uniform data sharing and management policies, advancing individuals' access to their health information, and advancing public health literacy.

Below we outline our recommendations in more detail, and we address NLM's specific questions related to this RFI. Should you have any questions or require additional information, please contact Public Policy Specialist Scott Weinberg, MPA at scott@amia.org. We, again, thank NLM for the opportunity to comment and look forward to continued dialogue.

Sincerely,



Patricia C. Dykes, PhD, RN, FAAN, FACMI
Chair, AMIA Board of Directors
Program Director Research

AMIA | American Medical Informatics Association
4720 Montgomery Lane, Suite 500 | Bethesda, Maryland 20814

October 16, 2020

Center for Patient Safety, Research, and Practice
Brigham and Women's Hospital

(Enclosed: Detailed AMIA Comments regarding major opportunities and challenges)

AMIA | American Medical Informatics Association
4720 Montgomery Lane, Suite 500 | Bethesda, Maryland 20814

October 16, 2020

Major opportunities or challenges that have emerged over the last five years and that have implications for the future of NLM in the areas of:

<table>
<tr><th colspan="2">a. Science (including clinical health sciences, biomedical science, information science, informatics, data analytics, data science, etc.)</th></tr>
<tr><td colspan="2">AMIA Response</td></tr>
<tr><td>Challenge:
The electronic health record (EHR) remains a deficient tool for documenting and communicating patient stories and it sits at the center of an increasingly complex environment of systems ill-equipped to interoperate or deliver myriad functions necessary for modern care delivery, research, public health, or human services.</td><td>Opportunities:
NLM should support design-level research to improve EHRs for patient care, while facilitating research, public health surveillance and reporting, and human services coordination.
NLM should support systems-level research to improve our collective understanding of how EHRs and other information systems should interact.</td></tr>
<tr><td>Challenge:
Evidence generation is occurring at increasingly rapid pace – sometimes on a daily basis, as has been the case with COVID-19 – and we do not have efficient ways to integrate new knowledge/evidence at the point-of-care. This is especially problematic at scale or with rapidly changing environments (like COVID-19).</td><td>Opportunities:
NLM should support research on executable/computable knowledge to keep CDS, for example, updated as evidence is generated or changes.
Standards for expression of scientific evidence is underway at HL7 as part of a group called Evidence-Based Medicine on FHIR (EBMonFHIR).[5] In less than 2 years, the EBMonFHIR project established draft standards for expression of evidence,[6] evidence variables,[7] statistics,[8] and ordered distributions for statistical arrays.[9]
Meanwhile, others are working on a possible common metadata framework with the Mobilizing Computable Biomedical Knowledge (MCBK) effort. A draft manuscript from MCBK participants notes, “A Common Metadata Framework will advance</td></tr>
</table>

	efforts to mobilize CBKs. A Common Metadata Framework addresses both the opportunities and challenges inherent in coordinating relationships across CBKs throughout clinical care, public health, education, and biomedical science. Developing a Common Metadata Framework requires attention to common needs within the wide variety of complexity, goals, tangible factors, and expected audiences across CBKs." Opportunities • Consistent with recommendations provided to the NIH in 2019,[10] the NLM should devote substantial resources to develop a strategy and a framework for FHIR research and development (R&D) to maximize the utility and secondary use of EHR-based, clinical data for research and to encourage the development of clinical informatics innovations aimed at improving care. o By working through the established standards maturity model progressions and balloting processes of HL7, this strategy and framework for FHIR R&D should result in FHIR Resources, FHIR Profiles, and FHIR Implementation Guides (IGs) to facilitate a broad range of research using clinical data. AMIA recommends the NIH focus on three categories of activity, including the use of FHIR to: 1. Facilitate access to clinical data for research, such as by supporting development of bulk access to clinical data; 2. Enable clinical research using FHIR, such as refinement of the Consent Resource and development of Privacy Labels; and 3. Translate clinical evidence back into practice, such as efforts to use FHIR for clinical guidelines. • In addition to supporting FHIR R&D, the NLM should support standard development efforts for metadata as well as support research that
Challenge: • Data liquidity and interoperability of health data presents ongoing challenges for clinical care, research, and public health.	

	demonstrates how to integrate data from different sources within the ecosystem. • NLM can support the additional exploration of enhanced and standardized documentation of implementation and usage of standards, including HL7, FHIR, terminologies, and coding systems. Such work could shed light on gaps in interoperability. At the same time, it would encourage exploration of using existing standards to demonstrate different levels of interoperability and pinpoint the obstacles to achieving interoperability.
Challenge • Data quality of clinical and real-world data (RWD) is low, and analytics maturity is underdeveloped because of this low data quality.	Opportunities • NLM should support research that systematically improves data quality. The new 2020 NEST Coordinating Center Data Quality Framework is now being piloted for generation of real-world evidence. Stakeholders include the medical device community, hospitals, PCORnet researchers, and the FDA.[11] Additionally, the Health Catalyst Healthcare Analytics Maturity Model, which is being updated,[12] delineates many secondary uses of clinical and RWD desired for a Learning Healthcare System.[13] Maturity in both frameworks is predicated on higher levels of data quality for secondary use than most institutions have achieved.
b. Technology (including biotechnology, platforms, hardware, software, algorithms, processes, systems, etc.)	
Challenge: • Algorithmic fairness and identification/mitigation of bias is emerging as a critical aspect of AI/ML-driven health and research applications.	Opportunities: • NLM can support efforts to develop consensus measures and standards for algorithmic transparency and communication devices (i.e. product labels) for machine learning-driven applications in health. o A growing body of literature indicates that concerted efforts are needed to better ensure fairness and mitigate AI/ML-derived bias in health and research applications.[14,15,16,17] o A forthcoming AMIA Position Paper presents a policy framework that spans the design and development, implementation, evaluation, and on-going maintenance of Adaptive CDS.[18] This work envisions

	an extensive policy landscape that includes transparency metrics for Adaptive CDS training datasets, communications standards to provide accurate information about the intended uses of Adaptive CDS, and dedicated actors and protocols to evaluate, test, and monitor Adaptive CDS *in situ*.
Challenge: • An expansive and expanding ecosystem of health technology – much of it coming out of non-EHR space in research – portends a need to integrate informatics into that work outside traditional EHR ecosystem.	Opportunities: • The medical context is very specific. Adoption of general technology (i.e., web standards and web-based services), however, can make the technology more specific or more suitable in medical contexts. We should thus leverage web standards and web-based services, rather than focus on technologies that are strictly medical. • Publish standards as authoritative, persistent URLs – vocabularies, terminologies, etc. – that can be referenced via a Wikipedia-like application for healthcare concepts in web-standards format. ○ The Value Set Authority Center is an exemplar resource for this opportunity and we recommend it is enhanced and promoted widely.[19] • In addition to publishing the terminologies (value sets and code systems), there is a need to develop terminologies for multiple aspects of reporting scientific knowledge. The COVID-19 Knowledge Accelerator project developed a Code System Development Protocol (tinyurl.com/CodeSystemDevelopmentProtocol, registered at https://osf.io/8d62q) to create code systems for statistic type, statistic model, study design, and risk of bias concepts. The protocol can be used for other terminologies as well and NLM would be a natural home for such development. • NLM is a logical owner for master data management (MDM), including relevant metadata management. There are several layers of topics associated with MDM including implementation frameworks, terminologies, statement models, assertional knowledge networks, and process representations. There is also a need for translation of MDM

	practices from other industries. MDM also overlaps with curation of algorithms and code and what we are calling analytics DevOps.
c. Public health, consumer health, and outreach (including epidemic disease surveillance, culturally competent engagement, optimizing the experience of resource users, etc.)	
Challenge: • Data reporting from clinical care to public health authorities is inconsistent, insufficiently coordinated, and siloed.	Opportunities: • While NLM may only play a supporting role in addressing clinical-to-public health data reporting, it can support research to: o Express evidence and guidelines in computable formats during dynamic public health emergencies. The COVID-19 Knowledge Accelerator (https://www.gps.health/covid19_knowledge_accelerator.html) and CPG-on-FHIR effort at the CDC are exemplars of such work. o Help coordinate ontologies and standards to improve consistency in how data is captured and reported to public health (MEDMORPH at CDC)[20] o The Reportable Conditions Knowledge Management System (RCKMS) is an authoritative, real-time portal to improve disease surveillance. RCKMS stores comprehensive information on public health reporting requirements and acts as a decision support service (DSS) to determine if a potential case is reportable and to which jurisdiction(s).[21] o Digital Bridge – public/private partnership coordinating discussion around many public health and information exchange initiatives, including viewpoints from a multi-stakeholder group of health care organizations, public health authorities, and industry partners.[22]
• Difficulty in getting feedback from public health back to the providers	• Opportunity for NLM to work with other groups to make public health information and analysis available back to providers and community partners to enhance situational awareness as well as to provide updates on community health status to facilitate local decision-making.

	• While much data reported to public health agencies is generally made public in various ways, there is often a significant lag time before such data is made available. Efforts outlined above to improve and enhance electronic reporting to public health agencies should also enable more rapid creation of consolidated electronic data that could more easily be returned to providers. • Research efforts around optimal multi-directional flow of data for public health should be encouraged to inform the development of such systems that address the needs of multiple stakeholders.
d. Library functions (including collection development, access, preservation, indexing, library metadata, service agreements with other libraries, etc.)	
Challenge: • The US citizenry has generally poor health literacy, poor scientific literacy, and poor data literacy.	Opportunities: • The pandemic has shown a need for libraries to engage in health literacy campaigns designed to reach broader audiences, including both basic and more advanced levels. • Data literacy standards for those in healthcare, as well as the public and other stakeholders that NLM works with – re: algorithms, ML performance, etc. Educational role for data literacy and beyond data literacy to disseminate standards. • Need to teach the public about science and data and how to interpret results. At a time in our nation when dis-information is often rapidly disseminated, it is extremely important for individual citizens to have a reliable, timely, and apolitical, source of comprehensible health-related information. • Citizen Science
Challenge: • Finding and disseminating digital medical content and resources is suboptimal	Opportunities: • The NLM could provide digital content delivery to health care organizations and collect and share digital resources for clinicians. We envision a collection of tools to help with different areas (e.g., mental

	health, prevention, etc.) and serve as a repository of digital resources worldwide. • The NLM could also develop ways to help individuals assess the trustworthiness of health information.
e. Modes of scholarly communication (including researchers' use of social media, preprints, living papers, changes in the roles and practices of publishers, data-driven approaches to studying historical medical texts, images, and datasets, etc.)	
Challenge: • Researchers who produce valuable software code/algorithms, datasets, and knowledge resources do not routinely receive academic credit for this work, unless it is joined by a publication. Researchers deserve credit for their work, especially given that career and professional development is dependent on scholarly references and citations.	Opportunities: • COVID-19 revealed a need to share citations as evidence emerges and a possible approach leverages FHIR with a Citation Resource.[23] The FHIR Citation Resource supports interoperable exchange between MEDLINE and citations for many knowledge representations not found in MEDLINE such as preprints, websites, and computable resources. With substantial attention to attribution of scholarly contribution the Citation Resource includes author data, a contributor element and a taxonomy (credit taxonomy – 14 different kinds of attributions). Some article pathways ask for contributor statements – if you apply citations to knowledge artifacts, that allows the expression of contribution to all types of artifacts.[23],[24] • Software code is emerging as an important output of biomedical discovery and the adoption and reuse of free and open source software necessitates tracking and referencing these outputs. The NLM can emerge as a leader for the curation of software code as scholarly output, including identifying evaluation rubrics and training to promote established best practices with respect to open source software. Furthermore, the NLM may lead in establishing useful metrics for software performance and relevance – such as code downloads, citations, or external contributions. • Computable notebooks – such as Jupyter notebooks to publish digital content with associated data/metadata/code – provide additional sources of important research content that should help researchers advance their careers.

f. Perspectives, practices, and policies (including those related to open science, the need for diversity, equity, and inclusion in research, algorithmic bias, expectations of reproducibility of research, etc.)	
Challenge: • Algorithmic bias and the inability to reproduce research results threatens to undermine scientific advances.	Opportunities: • In addition to what we mention in section (b) we note that risk of bias in algorithms can result when biases are part of the model of analysis. The NLM should support development of structured methods to identify and mitigate internal bias – such as those found in code systems. We note that bias needs to be considered in the design of programs and initiatives, similarly to how cybersecurity should be part of medical device design. • Where algorithms and software code are included as part of the scientific output, it is important for the peer review process to include the opportunity to review and attempt to reproduce the established findings. This can be done through establishing expectations through adoption of open source – which can allow the source code to be made available, while not always requiring the author to make it publicly and/or freely available (although additional work to move towards open science can and should be encouraged). This review process can strengthen the scientific field, and may additionally open a new avenue by which issues such as algorithmic bias can be identified. • Many biased algorithms stem from the use of biased or incomplete data sets and sources in development and training. Such biases may not be documented, or in some cases, biased data may be the only data available. The NLM can support efforts to define standards for algorithmic training data sets that would minimize known biases, as well as to define necessary documentation of algorithmic source data that can be used to identify known or potential biases and shortcomings that can be taken in consideration when applying such algorithms to new data sets. • NLM should continue its work with AIBLE, but also explore use of synthetic data creation and testing to help researchers understand the impact of data quality on their algorithms and approaches.
Challenge:	Opportunities:

• Explainability for how AI/ML-driven applications in healthcare is challenging.	• Explainability of AI algorithms also presents a challenge and opportunity for NLM. The NLM can support efforts to establish standards for the level and type of information about how individual algorithms function that is required by the health-care community to understand the basis and potential applicability of derived results. • As algorithms are implemented, there needs to be a way to visualize/communication their mechanism of action to aid explainability. • Explainability should always include discussion of how bias / diversity was factored in - or not – so that the end-user can appreciate the potential limits of applicability of such systems to their target populations.
Challenge: • Citizen science should be included as part of this section's issues.	Opportunities: • Citizen science – not just education of informaticians, but education of the broader society on public health and science more generally is certainly a core information need that NLM can serve. • Some of the consumer-oriented material on NLM's web site were clearly focused on providing accessible information to the general public. Rather than discontinuing such services on 10/1/20 as indicated, we would urge NLM to not only re-instate but expand them so that consumers have a truly user-friendly authoritative resource to turn to.
g. Workforce needs (including data science competencies, effective strategies for recruitment and retention of underrepresented minorities, opportunities for training and continuing education for middle- and late-career researchers and librarians, etc.)	
Challenge: • There is a deficit of professionals with skills and knowledge necessary for data curation, retrieval, and stakeholder communication regarding data access, and data literacy education.	Opportunity: • Workforce opportunity for health sciences and data librarians: There's an intersection between these librarian roles and the role for which there is growing need in research analytics / data science / research informatics / data warehousing infrastructures, e.g., data concierge services or data support analysts. There is a need to have greater workforce in this space to help manage and curate data, participate in data–governance and request

	structures, and communicate with and educate research data requesters about access and data literacy issues,. This is a cross-over opportunity for which relatively little additional training would be required for librarians across a range of career stages; same potential for graduates from master's degree programs in informatics programs. This need was expressed and discussed during the 2019 Healthcare Data Analytics Association conference. • Additional recommendations include: ○ Additional emphasis on just-in-time training for mid and late career professionals ○ Explicit efforts to diversify the field of informatics by bringing Under Respresented Minorities (URMs) into training programs. ○ Acknowledgement that PhD-level informaticians are needed in industry and government, as well as academia. ○ Continued support for T15 training programs.

[i] Vulcan HL7 FHIR Accelerator: http://www.hl7.org/vulcan/
[ii] Mobilizing Computable Biomedical Knowledge: https://mobilizecbk.med.umich.edu/about
[iii] http://hl7.org/fhir/uv/cpg/2019Sep/
[iv] https://confluence.hl7.org/display/CDS/EBMonFHIR
[5] Confluence. EBMonFHIR clinical decision support. 2019. https://confluence.hl7.org/display/CDS/EBMonFHIR Accessed September 11, 2020.
[6] Ibid, http://build.fhir.org/evidence.html
[7] Ibid, http://build.fhir.org/evidencevariable.html
[8] Ibid, http://build.fhir.org/statistic.html
[9] Ibid, http://build.fhir.org/ordereddistribution.html

[10] AMIA Recommendations to NIH Regarding: Use of the Health Level Seven International (HL7®) Fast Healthcare Interoperability Resources (FHIR) for Capturing and Sharing Clinical Data for Research Purposes. https://www.amia.org/sites/default/files/AMIA-Response-to-NIH-RFI-on-FHIR.pdf. Accessed September 15, 2020.
[11] NEST Coordinating Center Data Quality Framework: https://nestcc.org/data-quality-and-methods/
[12] Healthcare Analytics Adoption Model: https://www.healthcatalyst.com/healthcare-analytics-adoption-model/
[13] Institute of Medicine Roundtable on Evidence-Based M. The National Academies Collection: Reports funded by National Institutes of Health. In: Olsen L, Aisner D, McGinnis JM, editors. The Learning Healthcare System: Workshop Summary. Washington (DC): National Academies Press (US) Copyright 2007, National Academy of Sciences.; 2007
[14] [12] DeCamp M, Lindvall C. Latent bias and the implementation of artificial intelligence in medicine. J Am Med Inform Assoc. 2020; 27(12):
[15] McCradden MD, Joshi S, Anderson JA, et al. Patient safety and quality improvement: ethical principles for a regulatory approach to bias in healthcare machine learning. J Am Med Inform Assoc. 2020; 27(12):
[16] Ferryman. Health Disparities in the FDA's AI and Machine Learning Regulatory Framework. J Am Med Inform Assoc. 2020; 27(12):
[17] Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019; 366(6464):447-53.
[18] Petersen C, Smith J, Freimuth R, et al. Recommendations for the Safe, Effective Use of Adaptive CDS in the US Healthcare System. J Am Med Inform Assoc. 2020; 27(12):
[19] https://vsac.nlm.nih.gov/
[20]MEDMorph Project. CDC. Available at: https://www.cdc.gov/csels/phio/making-ehr-data-more-available.html
[21] Reportable Conditions Knowledge Management System (RCKMS). CSTE. Available at: https://www.rckms.org/about-rckms/
[22] Digital Bridge Project. Available at: https://digitalbridge.us/
[23] Ibid, http://build.fhir.org/citation.html
[23]http://build.fhir.org/citation.html
[24] https://docs.google.com/spreadsheets/d/1HAskxADsbNsiQLAOksDkg2inj3-52zl3IGpTlIvV0bc