Sharing Healthcare Data
CRA-I Workshop Report

CRA-I

Computing Research Association
Industry

Sharing Healthcare Data
CRA-I Workshop Report

Workshop Organizers
Jessica Pater (Parkview Health)
Helen Wright (CRA-I)
Tammy Toscos (CRA-I, Parkview Health)

Workshop Advisory Committee
Shion Guha (University of Toronto)
Fayika Farhat Nova (Parkview Health)
Divesh Srivastava (CRA-I, AT&T)

All Organizers and Committee members served as authors on this report

CRA-I
Computing Research Association
Industry

Executive Summary

The Sharing Healthcare Data workshop, sponsored by the Computing Research Association's Industry committee (CRA-I), was held on October 17th and 18th, 2024 in Washington D.C. A total of 35 representatives from industry/non-profit (15), academia (14), and government (6) participated, representing a cross-section of roles that connect with healthcare data (see Appendix for the full Workshop Participant list).

Workshop Goals: The main impetus of this workshop was to convene a cross-section of researchers, practitioners, innovators, and staff of key federal agencies to discuss the state-of-the-art as it relates to the sharing of healthcare data and the potential barriers and opportunities we have on the horizon. The key goals were (1) discuss the challenges of sharing healthcare in different settings, (2) identify current best practices and opportunities for sharing healthcare data, and (3) outline an actionable roadmap for next steps.

Workshop Activities: To achieve the goals of the workshop we organized a set of four panels which each had representatives from our different stakeholder groups. The panels focused on barriers to sharing data, the integration of AI and connection of health models, ethics of sharing healthcare data, and the current and emerging regulatory landscape with a focus on AI compliance. Additionally, we had two keynotes on the first day which galvanized the attendees to think audaciously and actionably about healthcare data sharing. Finally, we organized breakout sessions where participants were given prompts as a call-to-action for next steps.

Emergent Themes: Based on the workshop activities, the primary theme that emerged was that the increasingly disruptive role that AI plays within the healthcare space has led to misalignments that are not fully understood by key stakeholders. Additional themes included agency and control of healthcare data, validating broken or misaligned trust, and the impact of proposed legislation and regulation. What was evident is that stakeholder groups often do not recognize the breadth and depth of activity within this domain, as external visibility is often very difficult without some form of shared goal or incentive. To this point, a shared vision was provided for how the government agencies present linked together to form critical infrastructure for industry and academia to navigate and advance research and innovation.

Recommended Next Steps: Continued engagement is essential due to the state of flux related to healthcare AI policies and regulations that are emerging from the newly established Assistant Secretary for Technology Policy and the Office of the National Coordinator for Health Information Technology (ASTP/ONC). Participants outlined a set of potential future roundtables and/or workshops that explicitly address each previously identified misalignment and missing stakeholders whose voices need to be part of this ongoing conversation.

1. Introduction & Background

The importance of sharing healthcare data is well established. It contributes to improved accuracy in research [1], accelerating impacts of research [2,3], supporting health decision-making [4], improving efficiency and reducing healthcare costs [5], and restoring trust in healthcare [6]. After many decades of discussing the importance of sharing healthcare data [7], several impediments and challenges remain in creating systems that safely support the data sharing [4] and incentivize the sharing of this data [8]. Additionally, there are moral and ethical arguments associated with transparency and access of healthcare data [9,10]. Moreover, the influx of AI into the healthcare domain has made the sharing of health data even more urgent as massive data are key to their development and ongoing support of AI-driven solutions [11] but may unintentionally identify patients from deidentified data [12]. Policy plays a key role in this discussion as existing approaches to data sharing often restrict access [13], thus incentivizing the sharing of this data is critical [8,12].

1.1 Workshop Goals

The main goal of the workshop was to discuss current opportunities and risks associated with health data sharing, explore associated challenges of data ownership, access, and control, and identify technological innovations that can overcome these barriers. In addition to fostering collaboration and discussion between participants, key objectives were to advance the conversation on ethical and regulatory considerations in AI-driven healthcare and generate actionable insights that can inform future research and policy.

1.2 How we Assembled

The Computing Research Association's Industry committee (CRA-I) hosted a virtual roundtable on December 6, 2023. The roundtable panel brought together diverse industry perspectives including a community-based health system, a state-based health information exchange, and an applied research center. The conversation highlighted the transformative potential of healthcare data sharing in today's rapidly evolving healthcare landscape. As a force with the capacity to revolutionize patient care, research, and policymaking, it has become a crucial endeavor. However, the discussion acknowledged the myriad challenges, ranging from privacy concerns to interoperability issues.

The CRA-I council decided to build on the success of the roundtable and expand the conversation to a 2-day workshop. The workshop organizers and advisory committee structured the workshop structure around core concepts that emerged from the roundtable. Due to the importance of including representatives from government agencies with healthcare data initiatives, it was decided to convene the workshop in Washington D.C. The workshop took place on October 17-18, 2024. A total of 35 participants came together with almost even distribution from academia (14 participants), industry /nonprofit (15), and government (6 participants).

1.3 Workshop Structure

The workshop organizers set the stage that all comments and discussions that took place in the workshop would not be directly identified in the workshop report – that no remarks would be directly attributed to individual participants.

Day 1 began with a discussion of the workshop goals, expectations for participation, and the agenda for the day. The day started off with a keynote from Deborah Estrin (Cornell Tech), "Patient-generated data sharing: advancing hybrid, longitudinal, patient care with digital biomarkers and therapeutics (DBx, DTx)". We then had three panels: Barriers to Healthcare Data, InterAI: Connecting Health Models/AI, and Ethics in Health Data Sharing. The first day ended with a keynote from Tom Kalil (Renaissance Philanthropy) which focused on "Can AI Save Lives?".

Day 2 began with a review of key themes from the previous day and the agenda for the day. This was followed by the final panel for the workshop: Navigating the Regulatory Landscape: AI Compliance in Health Data Sharing. Participants then broke into groups to address the following prompts:

- Who is not at the table that needs to be here and why;
- What are the issues that still need further definition/ exploration; and

- Where are the opportunities for innovation.

Finally, the groups reconvened for a report out and the workshop was adjourned.

2. Workshop Activities

2.1 Keynotes

The morning keynote speaker Deborah Estrin focused her discussion on "Patient-Generated Data Sharing: Advancing Hybrid, Longitudinal Patient Care with Digital Biomarkers and Therapeutics (DBx, DTx)". She emphasized the potential of patient-generated data in shaping comprehensive care. A key element was how do we think about data within different contexts, including the contexts in which the data are collected (which may affect their quality and utility as well as what sharing is appropriate) and contexts of use. Data without a sense of appropriate contexts will do more harm than help. Furthermore, digital therapeutics can scaffold caregivers in addition to producing valuable data. This would necessitate a paradigm shift from the standard transactional approach to a more systems approach. Finally, governance is critical and still needs standards, incentives, better infrastructure, and considerations of individual and collective privacy.

The afternoon keynote speaker Tom Kalil focused his discussion on "Can AI Save Lives?". He talked about the power of connecting longitudinal data with high dimensional data, asking the audience about what data do we need, if that data exists, and who owns it or has access to it. When thinking about AI and health data, it is critical to think about prioritization. The discussion included aspects of asking what data is missing that could make healthcare better? Is there data within the government that we could engage at a deeper level? The issue with "missing data" is there is often misaligned or missing incentives to share this data. Finally, participants were given the charge to be more concrete and action oriented.

2.2 Panel: Barriers to Sharing Healthcare Data

This panel brought together a diverse and expert group to discuss the multi-layered challenges surrounding healthcare data sharing. The panelist engaged in conversation that explored structural, technological, and socio-cultural dimensions of data sharing, focusing on historical distrust [14], incomplete data [15], and inequities in access [16], and technology adoption [17] in healthcare domain.

Trust as the Foundation of Data Sharing

One of the key themes that emerged was the role of **trust** in building effective data-sharing mechanisms in healthcare [18,19]. Trust isn't just about being transparent; it requires intentional, sustained efforts, the building of long-term relationships, and culturally sensitive communication strategies. Panelists shared experiences working with marginalized communities, emphasizing that historically underserved groups—particularly communities of color—face systemic inequities in healthcare [20]. These inequities have led to widespread mistrust in health systems and efforts related to data sharing. Panelists characterized trust as a relational asset—developed through transparency, respect, and consistently protecting sensitive information over time.

Furthermore, **trust must be earned**, not assumed. This is especially salient with respect to data repositories like Health Information Exchanges (HIEs) as data breaches can significantly damage trust if transparency and accountability are not prioritized [21]. The conversation emphasized that building trust goes far beyond meeting mere compliance requirements. True trust-building involves proactive strategies such as meaningful community engagement, shared decision-making, and ensuring the voices of diverse populations are heard. Solutions discussed included fostering partnerships with community organizations like faith-based groups, local libraries, and non-profits to increase grassroots health literacy and address misinformation, thus strengthening trust from the ground up.

Additionally, panelists and participants discussed the importance of educating stakeholders, including patients, about the various entities that govern the data ecosystem, along with their roles, policies, and mechanisms to ensure data safety, quality, and ethical use. Understanding this taxonomy of entities is crucial for building trust, as it helps clarify where accountability lies and promotes confidence in data-sharing practices.

Lack of Ownership, Consent, and Health Literacy as Barriers

Patient confusion about **ownership and control of their data** is another major barrier to data sharing [22]. Panelists and participants discussed that many patients still lack clarity about *who owns their data*, how it is shared, and the potential risks and benefits of participating in research or broader data sharing. Lack of clarity around data ownership, bonded with distrust around how data will be handled, further intensify the challenges of quality data collection and sharing across health system.

Panelists emphasized that patient **health and technological literacy** are critical in addressing these challenges. There is a need for educational and communication efforts that articulate not just *why* data sharing matters but *how* it can directly improve individual and community health outcomes. Transparency about data use, as well as clear communication about safeguards and ethical practices can help patients feel empowered rather than vulnerable when engaging with healthcare data systems.

Workshop panelists advocated for developing clear use cases on how and why data will be collected and used to improve patient outcomes to address trust issues. Additionally, establishing a shared vocabulary and standardized frameworks like the Office of the National Coordinator's **Trusted Exchange Framework and Common Agreement**[1] were discussed as ways to reduce misunderstandings, clarify consent processes, and streamline compliance pathways, ultimately promoting patient confidence and trust.

Incomplete Data and System Design as a Barrier to Innovation

The panel also focused on the **incompleteness of data** as a major systemic challenge in barriers of sharing healthcare data. Both people and organizations can intentionally or unintentionally withhold data, and technical limits may also lead to incompleteness. Incomplete data can lead to incorrect interpretations–at either the individual or population level–which may limit, or–even worse–misdirect innovation and investment, especially if those using the data believe it to be more complete than it is. Disparities in how healthcare institutions collect and analyze data contribute to variation in data quality and completeness, leading to downstream inequities in innovation and care delivery. For instance, panelists noted that smaller healthcare systems or providers from rural areas struggle with limited funding and technical barriers, slowing the adoption of standards that could improve data quality and interoperability.

The panelists mentioned regulatory frameworks like **Electronic Clinical Quality Measures (eCQMs)**[2] as an important standardized mechanism for improving data quality and patient outcomes. However, the panelists also shared concern on whether current practices adequately capture the full scope of patient data, and whether the data sets used for eCQMs are sufficiently representative of diverse patient populations. Addressing these structural barriers may involve rethinking incentives for stakeholders (health systems, patients, researchers) and adopting standardized measures that prioritize shared goals, equitable access, and innovation.

The discussion took a human-centered turn by exploring the role of technology, AI, and machine learning in addressing data inconsistencies and disparities. Panelists strongly advocated for a human-centered, **culturally competent approach** [23] where technologies used and adopted in health systems should be developed and deployed in ways that respect the values, histories, and lived experiences of marginalized communities. This includes recognizing **rural disparities**, addressing access to care barriers, and when **AI-driven solutions** are developed, prioritizing fairness, accountability, and representativeness. Panelists and workshop participants discussed the need to engage communities within the decision making and use of the systems, highlighting the All of Us project as a successful example.

---

[1] https://www.healthit.gov/topic/interoperability/policy/trusted-exchange-framework-and-common-agreement-tefca

[2] https://www.cms.gov/medicare/regulations-guidance/promoting-interoperability-programs/electronic-clinical-quality-measures-basics

2.3 Panel: InterAI-Connecting Health Models

This panel explored the integration of AI models into healthcare, focusing on challenges in data sharing, ethical concerns, and opportunities to improve patient care. Discussions ranged from the technical aspects of AI interoperability to human-centered considerations in healthcare. Specifically, we have organized the summary of this panel along the following dimensions and expanded them into takeaways for clinicians, policymakers, AI developers and other stakeholders.

The Role of Communication in Connecting AI Models for Better Healthcare Outcomes

Panelists noted the critical role of provider-patient communication. Studies have shown that patient behavior is influenced most by communication from a physician when compared with nursing or other ancillary healthcare providers [24]. This underscores the need for AI tools that enhance rather than replace critical conversations with physicians, to ensure patients receive clear and actionable information directly from trusted sources. AI models that augment rather than automate clinical decision-making was the overall takeaway from this discussion. As such, when AI models are connected to improve external validation, we need to make sure that they understand local decision-making contexts [25].

Data Sharing and Standardization Challenges

Panelists also raised the issue of standardizing data sharing and improving the efficiency of AI-generated systems, particularly in formatting medical notes. Currently popular AI technical advances such as RAG (retrieval-augmented generation) [26], which integrates embeddings (numeric, semantic representations) from images into conversations, were discussed as promising but still imperfect. While embeddings offer advantages over simple image descriptions (which often result in the loss of critical context) they are still susceptible to bias, including representational and emergent biases [27].

The panel also touched on the challenges of creating interoperability across AI models. While the focus of the discussion was on the current generation of AI models that use word, text and image embeddings to derive contextual meaning from non-quantitative data, there was also general consensus on the need to create interoperability across all kinds of AI models as organizations, communities, systems and individuals may have different understandings of AI models. Specific to the current generation of AI models, each new model release necessitates rebuilding embeddings, which hampers seamless integration and slows progress in adopting these technologies. Guidance from organizations such as NIST[3] were discussed as means to create standards that can influence the design and development of AI tools to meet specific business needs and afford interoperability.

Ethical and Human -Centered Concerns in AI Integration

The ethical implications of integrating AI tools and models in healthcare were a major point of discussion amongst the panelist and there were different perspectives about the extent to which ethical and human-centered considerations [28] need to be factored into AI models that connect and enhance healthcare decision-making. Key questions that were raised included:

- **Where is the human in the loop?** While AI can assist with insights and decision-making, it should complement rather than replace human judgment, ensuring care remains patient-centered [29].
- **Can AI understand daily variations and values?** The concept of "normal" varies for every patient, and AI must account for latent signals captured through sensors or other data streams to reinforce individual values and preferences rather than just completing discrete tasks [30].

Additionally, panelists proposed the idea of **"AI scaffolding"** [31], where AI acts as part of the care network, enhancing the capabilities of caregivers through insights and learning support rather than supplanting their roles. This approach involves designing multi-stakeholder engagements to create dynamic care systems that adapt to diverse needs. Additionally, we may need to develop new methods for

---

[3] https://sibr.nist.gov/record/health-data-technology-and-interoperability-certification-program-updates-algorithm-29

assessing the systems we integrate pre-deployment, given the current stat of these system to "hallucinate" or produce false data.

Complex Systems and Trust in Dynamic Healthcare Networks

Healthcare data exists within overlapping spheres, and understanding how these spheres interact is essential for designing dynamic networks of care [32]. Panelists explored how to enable effective data handoffs between stakeholders and systems, focusing on trust and selective disclosure. Intelligent systems that accommodate dynamic environments where caregivers and patients come and go and create a seamless experience are needed. The panelists also recognized the importance of including all stakeholders, not just clinicians, in the design and implementation of AI models. The ability to question and argue with AI outputs was deemed crucial to align these systems with human values and manage risks effectively.

Equity, Safety, and Accountability in AI-Driven Healthcare

Concerns about digital equity were raised, particularly regarding the private sector's role in creating disparities. For instance, patients or organizations that cannot afford to access or share data may face inequities in care. The panel referenced frameworks like AHRQ's Digital Health Equity Framework[4] to guide the equitable deployment of AI in healthcare.

Safety assessments for new technologies were another priority. Panelists emphasized the need for innovative methods to evaluate the risks of integrating AI tools. Suggestions included funding audits, integrating third-party assessments, and designing systems to reduce automation risks by complementing human care rather than replicating existing inefficiencies.

2.4 Panel: Navigating the Regulatory Landscape: AI Compliance in Health Data Sharing

This panel focused on reviewing the various federal landscape as it relates to health data and sharing. The discussion was anchored in a metaphor put forth by the panelists to help participants understand the individual agency roles and responsibilities in addition to how they "fit together".

Metaphor: Highway Transportation.

The metaphor provided was that of a car traveling on a road. The first thing needed is the actual design and engineering of the highway. Consider NIST as the engineers who measure the roads before they are built. They specify the technical framework—interoperability, security, and standards—that permits unrestricted data flow[5]. NIST collaborates with other entities like ASTP/ONC to develop and maintain technical specifications like FHIR and other standards to ensure different systems and networks can securely communicate. They are the core foundation of the data sharing ecosystem, ensuring stability, trust, and scalability.

Once the highway is defined, it needs to be built. The ASTP/ONC creates and maintains the infrastructure-interoperable systems and networks-that connects all stakeholders. They are the conduit that connects hospitals, research institutions, government agencies and private entities and allows data to move securely and efficiently between different stakeholders[6]. They bridge gaps and bring everyone into the same network, while promoting innovation, collaboration, and patient empowerment. An example of this is their pivotal role in TEFCA[7].

After the highway is built, policies are needed to regulate activity to ensure safe use. As the group responsible for quality and safety, AHRQ ensures data sharing is evidence-based, supported by research, and compliant with ethical

---

[4] https://digital.ahrq.gov/ahrq-funded-projects/creating-digital-healthcare-equity-framework-accompanying-guide-its-use/practical-guide

[5] https://www.nist.gov/itl/products-and-services/healthcare-standards-testing

[6] https://www.healthit.gov/topic/about-astponc

[7] https://www.healthit.gov/topic/interoperability/policy/trusted-exchange-framework-and-common-agreement-tefca

principles. They act as the regulator of "speed", enabling progress while preventing harm. AHRQ's strategic position allows it to act as a balance between innovation and caution[8]. While allowing the rapid exchanges of data to promote innovation, AHRQ sets clear boundaries that protect against misuse, bias, and unintended consequences. For example, AHRQ's Patient Safety indicators help hospitals identify preventable complications, allow them to identify areas for interventions and align shared data for quality monitoring[9]. AHRQ ensures that data sharing does not just move quickly but moves wisely, based on proven evidence and policy.

Finally, with the proper infrastructure in place different entities/groups can now utilize the infrastructure, thus representing drivers on this highway. Drivers can include an expansive group of stakeholders, including policymakers, researchers, administrators, caregivers, technology developers, and many other participants who rely on or contribute to the flow of data. Drivers' feedback and experiences fuel continuous improvements, ensuring the system remains relevant and effective. Just as drivers expect and depend on clear roadways, traffic signals, and safety laws to commute from point A to point B, stakeholders depend on trusted frameworks, guidelines, and technical solutions to access, share, and act on healthcare data without compromising ethics, privacy, or quality. However, a limitation of this metaphor was that it depends on roads that are already built, thus requiring vision for new possible pathways.

Data Quality

To realize data-dependent advances in healthcare (e.g., precision medicine [1]), data quality is of paramount importance. Panelists discussed the importance of using data that is representative. Additionally, we often use data to train models that were not created for that model. Most patient data we are using were created for billing purposes but are now being used to identify symptomology. Picking up a theme from the previous day, humans in the loop will continue to be critical.

Assessment of AI

Assessing the quality of AI outputs and the potential biases will be critical moving forward. Assessment will be a key feature of forthcoming policy and guidance. ASTP is coordinating HHS's strategy for data and AI, with guidance being published in 2025 [33]. There is a need for transparency **and continuous assessment** of AI as new inputs could have serious consequences on biases and outputs. It is important that community stakeholders (e.g., patients, providers) have visibility not only within the development and implementation of AI systems, but also in their ongoing assessments.

2.5 Panel: Ethics in Healthcare Data

This panel explored various ethical considerations and implications of sharing traditional and non-traditional healthcare data or pervasive data [34]. Specifically, patient generated health data from wearables like fitness trackers and social media were discussed.

Sharing data at the individual level: Advancing Individual Health

Digital traces from wearables and social media can provide key insights into individual health status and needs that are often not collected or discrete within the electronic health record [28]. However, there are ethical tensions around if and when these data should be integrated into health records and if they are not, how to systematically collect these data. Additionally, the objectivity of patient generated health data was discussed and the tension between giving additional context and insight and the challenges of linking health data from outside formal channels (e.g. EHR) with traditional health data.

Innovative Methods Increasing Representation: Advancing Science & Population Health

Traditional methods of data collection carry with them known biases. Innovative methods can help broaden our understanding of health needs and lived experiences of minority or historically unheard voices [35, 36]. Integrating community-based participatory research and other

---

[8] https://www.ahrq.gov/cpi/about/index.html

[9] https://qualityindicators.ahrq.gov/measures/psi_resources

community-driven health informatics approaches can improve the engagement of underserved populations [37,38]. Data donation is another avenue to collect non-traditional health data. Digital traces of our health are present via our phones [39] and connected devices [40], and by online engagement [41]. To better understand how people behave online, the National Internet Observatory is collecting data from people living in the U.S. [42].

The All of Us program is an innovative federal program working to improve healthcare through research by building a diverse database that is accessible[10]. There are currently over 800,000 unique participants of which nearly 550,000 having completed the initial steps of the program. By using a cloud-centric approach to data, the program facilitates collaboration by bringing researchers to the data, unlike the traditional approaches of bringing data to the researchers [43,44]. This program also addresses a key issue of systematically excluded data. By giving individuals the ability to contribute their own healthcare data, a wider representation and diversification within data may be achieved.

Provocations

The panelist provided several provocations to the participants. First, we assume we know what the data gaps are, but how do we know? Engaging communities is one way to validate our understanding, however deep qualitative/ethnographic work is needed before we can claim success in our knowledge. Additionally, computer scientists often use clinical lenses for sensemaking within digital or online spaces. However, this knowledge could potentially be limited or biased based on who is not represented in the data [45,46].

3. Charge to Action

3.1 Shared Understanding and Vision of the Community and How We Fit Together

To wrap our workshop, we imagine a healthcare data-sharing ecosystem where everyone, from patients to policymakers, play a critical role in moving us closer to a seamless collaboration, equitable access, and transformative healthcare outcomes. Achieving this harmony in data-sharing ecosystem starts with building a shared understanding and vision: a roadmap that connects diverse stakeholders under a common purpose. This shared goal is not just about interoperability; it is about establishing collaboration and ensuring every stakeholder understands their role in the ecosystem.

A critical question was also brought up as to: ***who is missing in our conversation?*** We need health system Chief Information Officers, Chief AI Trust Officers at federal agencies, consumer protection groups, health agencies not represented (e.g., HRSA, FQHCs, CHCs), accreditation bodies (e.g., Joint Commission), patient voices (e.g., trans youth, rare diseases), standards bodies (e.g., HL7), bioethicists, community organizations that support care delivery and community engaged researchers, and insurance companies. Future engagements could work to expand representation to these identified groups.

3.2 Increase crossover between the Healthcare Industry and Academic Computing Researchers

Currently, a gap exists between academic computer science researchers and the health systems that need their innovations. While the use of technology is prevalent in health care organizations, most health systems do not employ computing research professionals as part of their operations. Computing researchers who seek to enhance the efficacy and efficiency in healthcare through the application of algorithms to big data, for example, lack access to large health data sets. Moreover, these scientists also need to be able to connect to health care providers who can provide context around these data. Additionally, computing researchers may be interested in working in

---

[10] https://allofus.nih.gov/

a healthcare context but are disconnected from national initiatives (e.g., The Trusted Exchange Framework and Common Agreement-TEFCA[11]).

Computing research has a long history of translating into new product and innovation in various industries [47, 48]. The lack of healthcare data access has potentially slowed these types of rapid data-oriented advancements in medicine (i.e., in contrast to generative AI). Although generative AI holds potential to speed up healthcare data sharing, the quality and bias of these tools introduce new risks for patient safety and the need for further evaluation and development of guidelines [49,50].

4.0 Key Takeaways & Recommendations

4.1 Trust

4.1.1 Establishing appropriate **trust is foundational** to encourage data sharing, particularly with historically marginalized populations. A possible strategy is to partner with trusted local organizations (e.g., faith-based groups, libraries, and community centers) who act as intermediaries in connecting patients to healthcare information and technologies, and who can build patients confidence in sharing health information with their healthcare provider. Giving patients confidence in the data ecosystem involves educating relevant stakeholders on their roles and responsibilities and the related policies and mechanisms for safe data sharing practices.

4.1.2 Developing **human-centered, culturally competent data-sharing strategies** is crucial for establishing trust within healthcare systems. This approach can help to acknowledge and address historical distrust, misinformation, and security concerns, giving patients the confidence to share data that can be used to inform their care.

4.1.3 Enhancing physician-patient communication is critical for maintaining trust. Direct communication between physicians and patients is more effective in influencing behavior and ensuring better healthcare outcomes. AI tools should prioritize supporting and augmenting patient-provider interactions rather than replacing them, thus maintaining the physician's central role in care.

4.2 Stakeholder Engagement and Roles

4.2.1 **Collaborations** across the traditional and non-traditional healthcare ecosystem, including healthcare providers, payers, patients, and community partnerships in the design and implementation of data-sharing systems are needed to bring together stakeholders' varying interests in data sharing.

4.2.2 **Building stakeholders' capacity** is essential for establishing confidence in safe data sharing practices. Strategies may include educating stakeholders on their roles and responsibilities, making policy language accessible and understandable, and establishing mechanisms and processes that ensure patient safety

4.3 Data Access, Standardization, and Quality

4.3.1 **Data democratization** by ensuring better access to healthcare data can enable patients to easily access their health information. Moreover, when patients can readily access and share their health data, it simplifies the process of transferring information between healthcare providers, reducing the risk of errors, delays, or gaps in care.

4.3.2 **Standardization** of data **and interoperability** of data systems across the healthcare and community organizations is crucial for tracking upstream and downstream effects. Standardization becomes even more challenging when non—traditional health data is incorporated into the data ecosystem (e.g., social media data, patient generated data, digital therapeutics). The lack of standardization in data sharing and interoperability between AI models is a significant barrier. Continuous advancements, such as retrieval-augmented generation (RAG) and embedding integration, offer promise but require ongoing work to ensure compatibility across models and systems.

4.3.3 Tackling the challenge of **incomplete data** by focusing on targeted use cases, establishing **strategic partnerships**, and leveraging AI/ML tools to fill data gaps is important to promote innovation. This will ensure that all healthcare

---

[11] https://www.healthit.gov/topic/interoperability/policy/trusted-exchange-framework-and-common-agreement-tefca

providers, regardless of size, can share and access consistent, more complete data with the understanding of what is missing for better care coordination, especially during care transitions.

4.3.4 Designing dynamic networks with **safety, accountability, and dynamic care** in mind. These type of networks of care requires understanding overlapping data systems, enabling trust, selective data disclosure, and effective handoffs. Moreover, new safety assessment methods, audits, and accountability mechanisms are needed to mitigate risks associated with AI in healthcare. There also needs to be ongoing assessment on the impact of data-sharing policies on patient safety and outcomes.

4.3.5 Health data is **context dependent**. Time needs to be invested with communities to understand certain types of health data (e.g., social determinants of health). Our approaches should have a way of providing patient voice into the contextualization and understanding of this data.

4.3.6 **New methods** of data collection and sharing are critical moving forward. NIH's All of Us program serve as an exemplar of how data sharing can accelerate health research and medical breakthroughs. Other models include data donation which has shown promise within the citizen science health research domain.

4.4 Ethical Considerations

4.4.1 **Equitable access** of data must be addressed. The private sector's role in healthcare AI risks exacerbating inequities, as access to data and tools may be limited by cost. Frameworks like AHRQ's Digital Health Equity Framework can guide the development of more inclusive and accessible systems.

4.4.2 AI systems must be **ethically designed to include human oversight** ("human-in-the-loop"), capturing individual values and daily variations rather than simply automating tasks. AI should act as a scaffold for healthcare, supporting care networks without replacing human involvement and decision-making.

4.5 Regulatory Landscape

4.5.1 **Addressing regulatory and legal challenges** by aligning data collection/sharing practices and promoting the use of a **common vocabulary** with the latest standards can improve data interoperability. Using regulatory frameworks like ONC's **TEFCA** can help to streamline communication between stakeholders and ensure interoperability across health systems, especially for smaller organizations who may face technological and financial constraints.

4.5.2 Creating **feedback mechanisms** to emerging ATSP/ONC AI policies in addition to the Center for Devices and Radiological Health (CDRH) regulation of AI-enabled medical devices. Researchers, providers, and patient voices are critical and should have a formal channel that is accessible.

5.0 Conclusion & Next steps

All participants agreed that it was essential to continue this conversation. All participant groups contributed that additional workshops were key to continuing to build consensus and momentum in this space. These could focus on where healthcare needs to go and what can future computing innovation do to get us there and bring together the missing voices that were articulated by workshop participants. The regulatory landscape is vital to any activities within this space and several inter-organizational efforts are underway, specifically focused on the role of AI. As these begin to emerge in the coming months, the guidelines and frameworks outlined within this report will be critical for this community to move forward in a more integrated and harmonious fashion.

ACKNOWLEDGEMENTS

The workshop organizers extend their sincere appreciation to the dedicated CRA staff members who played a crucial role in the success of this workshop. We would like to thank Nene Bundu, Peter Harsha, and Brian Mosley for their invaluable support in handling logistics, managing last-minute changes, and ensuring a smooth and productive event.

REFERENCES

[1] Castaneda, C., Nalley, K., Mannion, C., Bhattacharyya, P., Blake, P., Pecora, A., Goy, A., & Suh, K. S. (2015). Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. Journal of clinical bioinformatics, 5, 1-16. https://doi.org/10.1186/s13336-015-0019-3

[2] Groves, P., Kayyali, B., Knott, D., & Kuiken, S. V. (2013). The 'big data' revolution in healthcare: Accelerating value and innovation. McKinsey.

[3] Whicher, D., Ahmed, M., Siddiqui, S., Adams, I., Grossman, C., & Carman, K. (2020). Health data sharing to support better outcomes. *Washington, DC: National Academy of Medicine*.

[4] Martin, L. T., Nelson, C., Yeung, D., Acosta, J. D., Qureshi, N., Blagg, T., & Chandra, A. (2022). The issues of interoperability and data connectedness for public health. Big Data, 10(S1), S19-S24. https://doi.org/10.1089/big.2022.0207

[5] Ait Abdelouahid, R., Debauche, O., Mahmoudi, S., & Marzak, A. (2023). Literature Review: Clinical Data Interoperability Models. Information, 14(7), 364. https://doi.org/10.3390/info14070364

[6] Sheehan, M., Friesen, P., Balmer, A., Cheeks, C., Davidson, S., Devereux, J., Findlay, D., Keats-Rohan, K., Lawrence, R., & Shafiq, K. (2021). Trust, trustworthiness and sharing patient data for research. Journal of medical ethics, 47(12), e26-e26. https://doi.org/10.1136/medethics-2019-106048

[7] Institute of Medicine (US). Sharing Clinical Research Data: Workshop Summary. Washington (DC): National Academies Press (US); 2013 Mar 29. 2, The Benefits of Data Sharing. Available from: https://www.ncbi.nlm.nih.gov/books/NBK137823/

[8] Panagopoulos, A., Minssen, T., Sideri, K., Yu, H., & Compagnucci, M. C. (2022). Incentivizing the sharing of healthcare data in the AI Era. *Computer Law & Security Review, 45*, 105670. https://doi.org/10.1016/j.clsr.2022.105670

[9] Porsdam Mann, S., Savulescu, J., & Sahakian, B. J. (2016). Facilitating the ethical use of health data for the benefit of society: electronic health records, consent and the duty of easy rescue. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2083). https://doi.org/10.1098/rsta.2016.0130

[10] Jackson, B. R., Kaplan, B., Schreiber, R., DeMuro, P. R., Nichols-Johnson, V., Ozeran, L., Solomonides, A., & Koppel, R. (2024). Ethical dimensions of clinical data sharing by US healthcare organizations for purposes beyond direct patient care: Interviews with healthcare leaders. Applied Clinical Informatics. https://doi.org/10.1055/a-2432-0329#info

[11] Pereira, T., Morgado, J., Silva, F., Pelter, M. M., Dias, V. R., Barros, R., Freitas, Cl, Negrao, E., Flor de Lima, B., Correia da Silva, M. & Madureira, A.J. (2021, June). Sharing biomedical data: Strengthening ai development in healthcare. In *Healthcare* (Vol. 9, No. 7, p. 827). MDPI. https://doi.org/10.3390/healthcare9070827

[12] Kancherla, J. (2020). Re-identification of health data through machine learning. https://doi.org/10.2139/ssrn.3794927

[13] Van Panhuis, W. G., Paul, P., Emerson, C., Grefenstette, J., Wilder, R., Herbst, A. J., Heyman, D., & Burke, D. S. (2014). A systematic review of barriers to data sharing in public health. *BMC public health, 14*, 1-9. https://doi.org/10.1186/1471-2458-14-1144

[14] Corbie-Smith, G., Thomas, S. B., & George, D. M. M. S. (2002). Distrust, race, and research. *Archives of internal medicine, 162*(21), 2458-2463. https://doi.org/10.1001/archinte.162.21.2458

[15] Gurupur, V. P., Abedin, P., Hooshmand, S., & Shelleh, M. (2022). Analyzing the Data Completeness of Patients' Records Using a Random Variable Approach to Predict the Incompleteness of Electronic Health Records. Applied Sciences, 12(21), 10746. https://doi.org/10.3390/app122110746

[16] Bak, M., Madai, V. I., Fritzsche, M. C., Mayrhofer, M. T., & McLennan, S. (2022). You can't have AI both ways: balancing health data privacy and access fairly. Frontiers in genetics, 13, 929453. https://doi.org/10.3389/fgene.2022.929453

[17] Sun, J., & Qu, Z. (2015). Understanding health information technology adoption: A synthesis of literature from an

activity perspective. Information Systems Frontiers, 17(5), 1177-1190. https://doi.org/10.1007/s10796-014-9497-2

[18] Holm, S., Kristiansen, T. B., & Ploug, T. (2021). Control, trust and the sharing of health information: the limits of trust. Journal of medical ethics, 47(12), e35-e35. https://doi.org/10.1136/medethics-2019-105887

[19] Taylor, L. A., Nong, P.A., and Platt, J.O. (2023). Fifty years of trust research in health care: a synthetic review. The Milbank Quarterly, 101(1), 126-178. https://doi.org/10.1111/1468-0009.12598

[20] Santos, P., Faughnan, K., Prost, C., & Tschampl, C. A. (2023). Systemic barriers to care coordination for marginalized and vulnerable populations. Journal of Social Distress and Homelessness, 32(2), 234-247. https://doi.org/10.1080/10530789.2021.2021361

[21] Yeo, L. H., & Banfield, J. (2022). Human factors in electronic health records cybersecurity breach: an exploratory analysis. *Perspectives in health information management, 19*(Spring). PMID: 35692854.

[22] Kostkova, P., Brewer, H., De Lusignan, S., Fottrell, E., Goldacre, B., Hart, G., Koczan, P., Knight, P., Marsolier, C., McKendry, R.A., and Ross, E. (2016). Who owns the data? Open data for healthcare. Frontiers in public health, 4, 7. https://doi.org/10.3389/fpubh.2016.00007

[23] Truong, M., Paradies, Y., & Priest, N. (2014). Interventions to improve cultural competency in healthcare: a systematic review of reviews. BMC health services research, 14, 1-17. https://doi.org/10.1186/1472-6963-14-99

[24] Collins, Sarah. "Explanations in consultations: the combined effectiveness of doctors' and nurses' communication with patients." Medical education 39, no. 8 (2005): 785-796. https://doi.org/10.1111/j.1365-2929.2005.02222.x

[25] Chui, V., Pater, J., Toscos, T., & Guha, S. (2023, January). Applying Human-Centered Data Science to Healthcare: Hyperlocal Modeling of COVID-19 Hospitalizations. In *Companion Proceedings of the 2023 ACM International Conference on Supporting Group Work* (pp. 24-26). https://doi.org/10.1145/3565967.3570979

[26] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.T., Rocktäschel, T. and Riedel, S., 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. Advances in Neural Information Processing Systems, 33, pp.9459-9474.

[27] Li, S., Koh, P. W., & Du, S. S. (2024). On Erroneous Agreements of CLIP Image Embeddings. arXiv preprint arXiv:2411.05195.

[28] Pater, J. A., Coupe, A., Nova, F. F., Pfafman, R., Carroll, J., Brouwer, A., Bohn, C., Li, J., Todd, N., Chang, F.L. & Guha, S. (2023). Social Media is Not a Health Proxy: Differences Between Social Media and Electronic Health Record Reports of Post-COVID Symptoms. *Proceedings of the ACM on Human-Computer Interaction, 7*(CSCW1), 1-25. https://doi.org/10.1145/3579624

[29] Aragon, Cecilia, Shion Guha, Marina Kogan, Michael Muller, and Gina Neff. Human-centered data science: an introduction. MIT Press, 2022.

[30] Bohr, Adam, and Kaveh Memarzadeh. "The rise of artificial intelligence in healthcare applications." In Artificial Intelligence in healthcare, pp. 25-60. Academic Press, 2020. https://doi.org/10.1016/B978-0-12-818438-7.00002-2

[31] Sætra, Henrik Skaug. "Scaffolding human champions: AI as a more competent other." Human Arenas (2022): 1-23. https://doi.org/10.1007/s42087-022-00304-8

[32] Abowd, G. D., & Mynatt, E. D. (2000). Charting past, present, and future research in ubiquitous computing. *ACM Transactions on Computer-Human Interaction (TOCHI), 7*(1), 29-58. https://doi.org/10.1145/344949.344988

[33] US Health and Human Services. (2025, January 10). "HHS Releases Strategic Plan for the Use of Artificial intelligence to Enhance and Protect the Health and Well-Being of Americans." Retrieved from: https://www.hhs.gov/about/news/2025/01/10/hhs-releases-strategic-plan-use-artificial-intelligence-enhance-protect-health-well-being-americans.html

[34] Sawyer, S. (2011). NSF Award #1133934: RCN: Digital Society and Technologies Research Coordination Network. Retrieved from https://www.nsf.gov/awardsearch/showAward?AWD_ID=1144934

[35] Maestre, J. F., MacLeod, H., Connelly, C. L., Dunbar, J. C., Beck, J., Siek, K. A., & Shih, P. C. (2018, April). Defining through expansion: conducting asynchronous remote communities (arc) research with stigmatized groups. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (pp. 1-13). https://doi.org/10.1145/3173574.3174131

[36] Mbunge, E., Muchemwa, B., & Batani, J. (2021). Sensors and healthcare 5.0: transformative shift in virtual care through emerging digital health technologies. Global Health Journal, 5(4), 169-177. https://doi.org/10.1016/j.glohj.2021.11.008

[37] Unertl, K. M., Schaefbauer, C. L., Campbell, T. R., Senteio, C., Siek, K. A., Bakken, S., & Veinot, T. C. (2016). Integrating community-based participatory research and informatics approaches to improve the engagement and health of underserved populations. *Journal of the American Medical Informatics Association, 23*(1), 60-73. https://doi.org/10.1093/jamia/ocv094

[38] Veinot, T. C., Clarke, P. J., Romero, D. M., Buis, L. R., Dillahunt, T. R., Vydiswaran, V. V., ... & Antonio, M. G. (2022). Equitable research PRAXIS: A framework for health informatics methods. *Yearbook of Medical Informatics, 31*(01), 307-316.

[39] Carroll, J. K., Moorhead, A., Bond, R., LeBlanc, W. G., Petrella, R. J., & Fiscella, K. (2017). Who uses mobile phone health apps and does use matter? A secondary data analytics approach. Journal of medical Internet research, 19(4), e125. https://doi.org/10.2196/jmir.5604

[40] Schroeder, J., Karkar, R., Murinova, N., Fogarty, J., & Munson, S. A. (2019). Examining opportunities for goal-directed self-tracking to support chronic condition management. Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies, 3(4), 1-26. https://doi.org/10.1145/3369809

[41] Alshaikh, F., Ramzan, F., Rawaf, S., & Majeed, A. (2014). Social network sites as a mode to collect health data: a systematic review. Journal of medical Internet research, 16(7), e3050. https://doi.org/10.2196/jmir.3050

[42] Feal, A., Gleason, J., Goel, P., Radford, J., Yang, K. C., Basl, J., Meyer, M., Choffines, D. Wilson, C. & Lazer, D. (2024). Introduction to National Internet Observatory.

[43] All of Us Research Program Investigators. (2019). The "All of Us" research program. New England Journal of Medicine, 381(7), 668-676. https://doi.org/10.1056/NEJMsr1809937

[44] Ramirez, A. H., Gebo, K. A., & Harris, P. A. (2021). Progress with the All of Us Research Program: opening access for researchers. Jama, 325(24), 2441-2442. https://doi.org/10.1001/jama.2021.7702

[45] Toscos, T., Drouin, M., Pater, J., Flanagan, M., Pfafman, R., & Mirro, M. J. (2019). Selection biases in technology-based intervention research: patients' technology use relates to both demographic and health-related inequities. Journal of the American Medical Informatics Association, 26(8-9), 835-839. https://doi.org/10.1093/jamia/ocz058

[46] Pater, J. A., Reining, L. E., Miller, A. D., Toscos, T., & Mynatt, E. D. (2019, May). " Notjustgirls" Exploring Male-related Eating Disordered Content across Social Media Platforms. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1-13). https://doi.org/10.1145/3290605.3300881

[47] National Academies of Sciences, Engineering, and Medicine. (2012). *Continuing Innovation in Information Technology: Interactive Timeline*. Retrieved December 15, 2024, from https://nap.nationalacademies.org/visualizations/continuing-innovation/

[48] National Research Council, *Continuing Innovation in Information Technology: Workshop Report*. Washington, DC: The National Academies Press. (2012). Retrieved from https://nap.nationalacademies.org/catalog/23393/continuing-innovation-in-information-technology-workshop-report

[49] Howell, M. D. (2024). Generative artificial intelligence, patient safety and healthcare quality: a review. BMJ Quality & Safety, 33(11), 748-754. https://doi.org/10.1136/bmjqs-2023-016690

[50] Rauh, M., Marchal, N., Manzini, A., Hendricks, L. A., Comanescu, R., Akbulut, C., ... & Weidinger, L. (2024, October). Gaps in the Safety Evaluation of Generative AI. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (Vol. 7, pp. 1200-1217). https://doi.org/10.1609/aies.v7i1.31717

Workshop Participants

First Name	Last Name	Affiliation
Marcy	Antonio	University of Michigan
Tracy	Camp	CRA, Colorado School of Mines
Kevin	Chaney	AHRQ
Ronald	Emeni	CRISP-DC
Deborah	Estrin	Cornell Tech
John	Garguilo	NIST
Margarita	Gonzalez	Georgia Tech Research Institute
Shion	Guha	University of Toronto
Douglas	Horner	Medical Informatics Engineering
Peter	Margolis	Cincinnati Children's
Endrit	Meta	Center for Health Affairs
Sean	Munson	University of Washington
Beth	Mynatt	Northeastern University
Fayika Farhat	Nova	Parkview Health
Jessica	Pater	Parkview Health
Ranjani	Ramamurthy	Global Health Labs
Andrea	Ramirez	*All of Us* Research Program
Vikram	Ravindra	University of Cincinnati
David	Schlueter	University of Toronto
Vivek	Sarkar	Georgia Tech
Jeffery	Smith	ASTP/ONC
Katie	Siek	Indiana University
Anna	Squicciarini	National Science Foundation
Ram	Sriram	NIST
Divesh	Srivastava	CRA-I, AT&T
Tong	Wu	Rutgers University
Helen	Wright	CRA

NOTES

CRA-I
Computing Research Association
Industry

1828 L Street, NW, Suite 800
Washington, DC 20036
www.cra.org/industry/