Data Sharing in the Era of COVID-19: Transforming Public Health
Randi E. Foraker
Washington University in St. Louis School of Medicine

Public health experts and health systems agree that data sharing in the era of COVID-19 is critical for understanding the *reach* and *impact* of the virus. Reach can be defined geographically, in terms of city, county, or state jurisdictions – or in terms of demographics: age, sex, race – or even clinically, in terms of the virus' ability to affect those patients with pre-existing conditions to a greater extent as compared to those who are healthier prior to infection. Meanwhile, impact can be quantified in terms of the *magnitude* of the virus' effect on certain geographic locales or "hotspots"; as we have seen among patients of advanced age, male sex, or black race; and on those who have diabetes, hypertension, or who are immunocompromised.

My colleagues and I in the [Institute for Informatics](#) at Washington University in St. Louis have implemented a data-driven response to COVID-19 which has been iteratively shaped by the data management, sharing, and analytic challenges we have faced since the pandemic began. In a [recent article](#), we uncovered the challenges of data access, integration, interoperability, and sharing across healthcare provider organizations and make recommendations for overcoming these barriers in the short- and long-term. We are working to surmount similar hurdles as we seek to expand the data sharing network and the utility of these data for a broader regional population health data solution.

A comprehensive, robust, and sustainable public health data infrastructure is needed to respond to COVID-19 and other pressing public health issues, such as sexually transmitted infections and substance use disorders. Typically these data are collected by jurisdiction (i.e., city or county) and are not easily integrated across geopolitical boundaries to provide comprehensive regional public health insights. Therefore, building upon our success with data sharing across healthcare provider organizations in our region, we seek to align these data with that from public health departments, laboratories, case management systems, mobile technologies, and other sources in support of a sustainable public health data infrastructure.

Optimizing the *mechanics* of data sharing is no small feat. Not only are data "siloed" from one another, housed in distinct jurisdictions or organizations, they exist on different platforms and in different formats. In order to address the practical issues of exchanging and integrating data across entities, we recently published the following [guidelines](#): 1) utilize existing tools and technologies to fill urgent data infrastructure needs, as developing "ideal" systems can be time-consuming; 2) collaboratively design the mechanisms and formats of data transmission, as these may vary across data sources; 3) organize data into a centralized format for analysis; and 4) discuss legal and research compliance issues with general counsel and institutional review boards to determine the best path forward.

Once the logistical barriers of data sharing are overcome, and the *mechanics* of data sharing are in place, the *implications* of data sharing can be addressed. Ideally, the entities sharing data can

convene to determine how the data resource can be used to benefit the broader good. A data trust can be employed to collect and store the data, and enforce rules regarding how the data are accessed, used, and shared. Data trusts can negotiate on the behalf of stakeholders and allow the collective to have a say in who has access to the data and for what purpose. Data trusts also can enforce an accountable and transparent governing structure for the resulting public health data ecosystem.

We propose to create a regional partnership to equip public health departments, elected officials, healthcare professionals, and citizens with the data and insights needed to guide decision-making. Ultimately, our vision for a public health data infrastructure is to enable local public health departments to conduct their important work of case identification and contact tracing more efficiently, while also deriving insights from the data asset to inform the general public about the reach and impact of the pandemic.

We hope to use the COVID-19 pandemic as a pilot for creating a prototype public health data infrastructure as we continue to build toward a more robust and sustainable data asset in support of public health practice and regional data insights. As a result, the data trust described above may need to remain agile in the context of a relatively nascent, and rapidly evolving, public health data infrastructure. At the outset, we may not understand the future scope of data to be collected, who should have access, and who can use the data. As this vision takes shape, we continue to engage in a robust community-wide dialogue with stakeholders in support of the health, wellness, and productivity of the region.