

Collaborating for Cures: New Initiative Helps Cancer Data Cross Borders

Pacific NW cancer centers and Microsoft aim to accelerate research by helping data flow.

March 5, 2020 By Fred Hutch News Service

Every biopsy tells a story.

To a cancer patient, their tiny bit of tissue tells a profoundly human one. It speaks of fear and pain and uncertainty.

To a computer, the story is bits and bytes. Lots of them. In the era of DNA sequencing, a single patient can generate a terabyte or two of data. Researchers sift through that data for genetic clues that could guide the development of new treatments.

Collectively, data from every cancer patient might add up to a story of breakthroughs. All that data might hold hidden insights that could change all our lives.

But right now it can't. Because in cancer research, data doesn't move. It's siloed — divvied up in databases across research centers.

Enter the Cascadia Data Discovery Initiative, [launched by Fred Hutchinson Cancer Research Center and sponsored by Microsoft](#). CDDI aims to establish a regional data-sharing ecosystem. It brings together institutions from across the Pacific Northwest with a simple goal: to make it easier for researchers to find and share biomedical data, and to collaborate.

Improving access to data will then let researchers tap into the region's wealth of expertise in artificial intelligence and machine learning to solve challenging scientific problems, said Raphael Gottardo, PhD, scientific director of Fred Hutch's Translational Data Science Integrated Research Center.

"Unlocking and sharing data is critical to discovering new ways to treat and ultimately cure cancer," said Gottardo, who holds the J. Orin Edson Foundation Endowed Chair. "Working in close collaboration with Microsoft through its [AI for Health initiative](#), we will be able to harness new advances in AI, machine learning and cloud computing to spur innovation and open up new avenues for preventing and treating cancer and related diseases."

The potential for these powerful technologies to accelerate cancer research is huge, said CDDI collaborator [Shannon McWeeney, PhD](#), associate director for computational biomedicine in the Oregon Health & Science Knight Cancer Institute. But they are very data-hungry.

“We could have all the algorithms in the world, but if we don’t have the data we’re not going to get there,” she said. “We envision a global research community in which sharing data becomes the norm.”

Microsoft’s AI for Health initiative is a new \$40 million, five-year program to empower researchers and organizations with AI to improve the health of people and communities around the world.

Transformative tech at medicine’s doorstep

CDDI is part of the larger [Cascadia Innovation Corridor](#) initiative, an effort to link the economies of Vancouver, B.C., Seattle and Portland through strategic partnerships.

It’s no secret that technology helps power those economies. Homegrown companies like Microsoft and Amazon are advancing the fields of data science, machine learning and artificial intelligence, and that work is transforming industries worldwide. It’s why the phone in your pocket can tell you how tough the commute home will be or whether you’ll need an umbrella next Tuesday.

Big data powers it all. And it really is big. A computer needs hundreds of thousands of data points to, say, tell the difference between a picture of a cat and a dog.

Cancer research often works on a much smaller scale. “We can see how big data and AI are transforming other industries,” said [Raymond Ng, PhD](#), director of University of British Columbia’s Data Science Institute. “But the transformation hasn’t hit cancer research fully because, in health research, data doesn’t move.”

One clear challenge to data sharing in health research: Researchers often don’t know what’s out there. CDDI has been working to address this challenge by creating a searchable database with information about what data is available at other CDDI-participating institutions. As a pilot project, Fred Hutch is pulling together information about datasets from research groups at BC Cancer, Fred Hutch, UW Medicine, and the OHSU Knight Cancer Institute to develop a search platform prototype.

Think of it as a catalog that shows what multiple institutions have, said Brenda Kostelecky, PhD, director of the [Cascadia Data Alliance](#).

“We’re not asking institutions to share the contents of their library at the outset,” she said. “Rather, you could find out whether other research centers have something you’d want to check out. That can spark new collaborations and shape new research questions.”

Finding ways to enhance privacy protection also has the potential to accelerate data sharing. UBC’s Ng is helping spearhead CDDI efforts to unlock data through privacy-preserving technology,

which can allow researchers to find and share data more easily while still protecting sensitive patient information. Microsoft has been leading the charge on developing an open-source platform for one such technology called differential privacy which, while broadly used in other industries, has not yet been applied to health research. The AI for Health collaboration will provide an opportunity to pair health research experts with data scientists to build an understanding of where privacy-preserving methods could be applied in health research.

Tearing down data silos

Not all the challenges can be solved by technology. Researchers often navigate a dizzying array of legal and regulatory challenges before the data starts flowing.

Lawyers and ethics experts at each CDDI-participating institution have developed their own requirements for how to share data that complies with federal, state, and provincial laws and regulations. That means data-use agreements often need to be hammered out each and every time researchers want to share data. Building on standardization work being done by the [National Center for Data to Health](#) and the [Federal Demonstration Project](#), CDDI is developing a common approach to data-use agreements between participating organizations. Such an approach could reduce the time it takes for researchers to negotiate access to data at other CDDI institutions, which today can stretch to half a year or longer.

Biomedical data might be ones and zeros to a machine. But researchers know every tissue sample tells a human story. They feel a responsibility to the patient it came from. They don't want it to just sit in a freezer. They hope it becomes part of a much larger story of finding cures.

"It's very impersonal when you talk about data because there is a real person on the other side of this who has suffered and chosen to give their sample so that others don't have to suffer," said [Aline Talhouk, PhD](#), a UBC assistant professor of obstetrics and gynecology and a CDDI collaborator on the search platform pilot. "It's unethical not to maximize the use of these samples," she said. "They shouldn't just sit in a freezer or a database. They should drive research forward."

[This article](#) was originally published on January 29, 2020, by Hutch News. It is republished with permission.