"The convergence of health-related data and associated applications demands computing and analytics at large scales."

Gina Tourassi
Director, ORNL Health Data Sciences Institute

Health Data Sciences at Oak Ridge National Laboratory

The increasingly large, complex, and diverse datasets being generated across the scientific spectrum show great promise in revolutionizing our understanding of many of today's most pressing scientific challenges, including healthcare.

Truly achieving an enhanced understanding of health drivers and outcomes requires parsing and making sense of today's vast health datasets, a process that in turn requires a research environment with world-class facilities; expertise across computing, data, and medical science; and an infrastructure capable of protecting sensitive health information. ORNL has all three. In recognition of its unique capabilities in advancing national health missions, the Laboratory established the Health Data Sciences Institute (HDSI) in 2013. Specifically, the Institute is tasked with three primary objectives:

- Delivering scientific innovation, informatics tools, and a computing infrastructure to enable the effective use of data for public benefit
- Advancing a broad range of sponsor and health policy priorities and the health research community at large
- Growing the health data science community via a user facility for collaboration, education, and training

HDSI includes biomedical researchers, system architects, data scientists, computer scientists, mathematicians, IT services, and high-performance computing experts. Via HDSI and related efforts, ORNL aspires to become the primary hub for national health data R&D.

2017
HPCwire Award for Best Use of AI for CANDLE

>22 million
patient records

85 HDSI
publications

>100
ORNL staff participating in Deep Learning User Group

ORNL researchers are utilizing advanced artificial intelligence (AI) and the fastest supercomputer on Earth to understand the effectiveness of cancer treatments and key drivers of cancer outcomes at the population level.

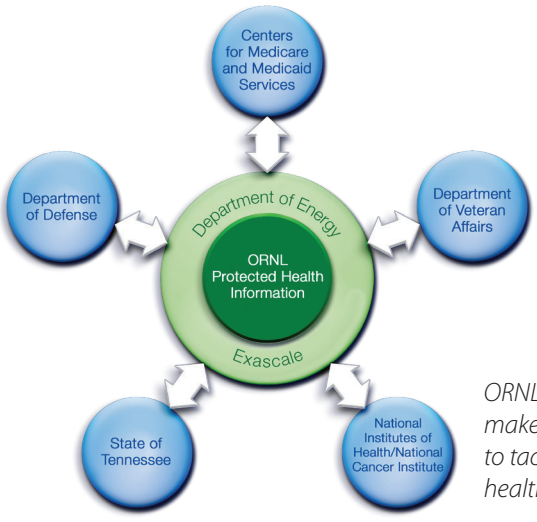
Why ORNL?

- ORNL offers unparalleled user facilities such as the Oak Ridge Leadership Computing Facility, Spallation Neutron Source, and Manufacturing Demonstration Facility, all of which produce enormous datasets and require ORNL to maintain a robust expertise in data science.
- The Laboratory's Summit supercomputer, with a peak speed of 200 petaflops, is currently the world's most powerful system for science and is specifically designed to train AI software (or algorithms) to tackle large scientific datasets.
- ORNL's Spallation Neutron Source enables the study of cancer initiation events at atomistic scales.
- The CADES–Knowledge Discovery Infrastructure (CADES-KDI) houses nearly 23 million individual Veterans Affairs (VA) patient records, the most comprehensive set of health and genomic data on the planet.
- ORNL staff possess unprecedented expertise across a wide range of computing and data science, including large-scale computation, graph analytics, and AI, and proven success utilizing population-scale medical data to enable translational "quantitative biology" science.
- The Laboratory is tasked with pushing the boundaries of privacy-aware computing for scalable, sensitive data.

Partnerships and Collaborations

HDSI is currently assisting numerous and varied critical health efforts at both the state and national levels, including the following:

- **Joint Design of Advanced Computing Solutions for Cancer** – DOE has partnered with the National Cancer Institute for an enhanced understanding of cancer biology and its application in facilitating new drug discovery and development and in applying big data analytics to a suite of cancer databases to optimize therapies. ORNL researchers are involved in three pilots, simultaneously pushing the boundaries of both computational science and our understanding of cancer.
- **Cancer Distributed Learning Environment (CANDLE)** – CANDLE is a scalable deep neural network code that addresses three top challenges of the National Cancer Institute: understanding the molecular basis of key protein interactions, developing predictive models for drug response, and automating the analysis and extraction of information from millions of cancer patient records to determine optimal cancer treatment strategies. The CANDLE project is also helping to inform the next generation of supercomputing architectures.



- **Tennessee Department of Health** – HDSI is partnering with the TDH to harness the power of data analytics in understanding what areas of the state are at the most risk for increased addiction rates and how to best allocate resources.
- **MVP-CHAMPION** – Because ORNL securely hosts millions of VA health records, HDSI researchers are perfectly positioned to assist clinical investigators in advancing clinical outcomes for both veterans and the larger population.

ORNL's unique suite of capabilities makes it an ideal partner in the quest to tackle today's most data-centric health challenges.

Contact: Gina Tourassi
Director, ORNL Health Data Sciences Institute
tourassig@ornl.gov, 865-576-4829
One Bethel Valley Road, Oak Ridge, TN 37830