

CASE STUDY

Washington University School of Medicine in St. Louis

Synthetic data mimics real healthcare data
without patient-privacy concerns

MDCLONE

 Washington University in St. Louis
SCHOOL OF MEDICINE

Institute for Informatics (I²)

Washington University in St. Louis

School of Medicine, Institute of Informatics (I2)



Institute for Informatics (I²)

LOCATIONS

St. Louis, Missouri

SIZE

15,539 students enrolled

2,040 beds at the medical center

16,950 total employees

One of the largest academic clinical practices in the United States

AREAS OF FOCUS

Medical Research

Environmental and Energy Research

Innovation and Entrepreneurial Research

Plant Science Research

WEBSITE

wustl.edu

As one of the top medical schools in the country, Washington University School of Medicine in St. Louis, Missouri, is highly regarded for excellence in research, teaching, and patient care. To support the growing need to collect, manage and harness big data in support of its mission, the School of Medicine created the Institute for Informatics (I2). The institute provides a home for biomedical informatics and data science research, practice, and educations spanning the entire School, with the goal of driving precision medicine approaches that can improve healthcare and public health.

Given these objectives, I2 engages in innovative research, workforce development, and informatics service delivery, targeting a variety of critical areas of need including:

- + The integration and dissemination of heterogeneous data, information, and knowledge resources
- + Computational approaches to the interpretation of bio-molecular, image, and clinical phenotypes to inform precision medicine
- + The acceleration of clinical and translational research through the systematic management of study protocols, data resources, and analytical pipelines
- + The creation of learning healthcare systems in which cyclical evidence generation and application become integral to care delivery
- + The use of ubiquitous computing and sensing technologies capable of facilitating population health monitoring and intervention strategies
- + Methodological and technical approaches to enable and enhance research reproducibility and rigor

Challenges

Like many academic health centers, attempting to integrate and analyze patient data for research studies, Washington University has faced challenges.

One of the biggest hurdles is making data available to researchers “at scale” while simultaneously protecting the privacy and confidentiality of the patients from whom that data is generated.

This hurdle is amplified by a number of related challenges, including:

DISORGANIZED DATA:

Data collected in electronic health records and other systems tends to be organized to be ideal for patient care or individual encounters, but when conducting research, the data needs to be reformatted to make it useful – it needs to be organized over a longitudinal patient time frame.

REGULATORY REQUIREMENTS:

Investigators often wait weeks or months to begin research projects involving real patients.

LIMITED RESOURCES:

Data change over time for individual patients and for populations, often requiring data to be reformatted to make it useful. Once the data are available and organized, there may be additional delays while researchers consult technology professionals, informaticians, and data scientists.

In response to these issues, I2 has undertaken a number of efforts to make data more diverse and larger-scale data accessible to researchers at the School of Medicine than ever before – while maintaining the highest levels of patient privacy and data security. Ultimately, making such data available empowers faculty, staff, and trainees to discover patterns in those data, with the goal of discovering new knowledge and improving patient outcomes.

Solution

The School of Medicine needed a way to enable researchers to access and analyze complex patient data, while protecting patient privacy.

In 2018, Washington University in St. Louis became the first medical school in the United States to

implement MDClone's platform, after attending the GlobalSTL Health Innovation Summit and connecting with the company as it was seeking collaborators in North America.



"There have been many proposed solutions to make data more accessible to researchers, including data de-identification, all of which have sought protect patients' privacy and confidentiality while maintaining the fidelity of source data. Unfortunately, traditional data de-identification methods can remove critical data elements, negatively impacting ensuing research projects."

Dr. Philip R.O. Payne, Ph.D., FACMI, FAMIA
Director of I2 and Chief Data Scientist, Washington University School of Medicine

By utilizing MDClone, the School of Medicine is able to provide higher fidelity data to researchers than traditional de-identification methods would allow for, while complying with the strictest patient privacy and confidentiality requirements.

THE SOLUTION: MDCLONE AND SYNTHETIC DATA

MDClone, through software it developed, generates synthetic data that mimic real patient populations, and those synthetic data sets can then be made available to School of Medicine researchers quickly and efficiently. Synthetic data can be analyzed as if they were original data, while addressing important privacy concerns. To do so, MDClone creates a synthetic analogue of healthcare data collected from actual patient populations. While the synthetic data set is similar in overall characteristics to the original data, there is no identifying information that can be traced back to individual patients or their identities. MDClone's software as well as the computational and network environments where the software is used, have been designed to comply with state-of-the-art cybersecurity and privacy controls.

"The beauty of synthetic data is that they allow us to quickly create data sets that look and feel just like the real data that are generated every time we interact with patients, which results in studies that arrive at conclusions comparable to those conducted with such source data, while greatly improving our ability to protect and maintain the confidentiality and privacy of the patients and communities we serve" said Dr. Payne.

BENEFITS

To demonstrate the power of synthetic data, the School of Medicine has conducted three pilot projects to ensure both data accuracy and patient privacy, focusing on:

- + Prediction of head trauma severity
- + Sepsis prediction
- + Sexually transmitted infections

The school's researchers were able to use synthetic data just as they would with real patient data to not only describe populations but also produce and validate advanced Machine Learning (ML) models.



“Despite coming from very different domains, our three use cases all demonstrated that the synthetic data were scientifically valid representations of the original data. We showed that a ML model trained on the synthetic data performed well when applied to the original data.”

Randi Foraker, Ph.D., M.A., FAHA, FAMIA

Director of the Center for Population Health Informatics at I2, Director of the Center for Administrative Data Research, and the Director of the Public Health Data & Training Center at the Institute for Public Health, Washington University in St. Louis

This speed and agility in terms of accessing data have created an ideal environment for researchers at the School of Medicine to accelerate scientific discovery. In addition, as a founding member of MDClone's Global Network, I2 has been able to collaborate with academic and health systems throughout the globe, instantly enabling high impact, data-driven research and innovation projects that reach across institutional boundaries.

With the MDClone platform, researchers who have completed a training curriculum and signed a data use agreement that ensures they will use synthetic data responsibly and for scientific purposes only, can log in and conduct queries in real-time. Instead of waiting weeks, months, or years to gain access to information, users can get access to data in real-time in order to ask and answer important questions about the data produced in the healthcare environment. In addition, the ability of MDClone users to rapidly iterate and refine their data-driven projects, leveraging the sophisticated MDClone query and data analysis tools, means that projects can be quickly designed and executed, especially when using emergent computational methods such as ML and cognitive computing.

Results

Making data safe, accurate, and accessible while still protecting patient privacy and confidentiality has stumped healthcare providers for years. MDClone's solution is changing the way data is accessed, analyzed and shared globally, bringing us closer to solutions to such challenges, and accelerate the pace of data-driven research as a result.

Today, the School of Medicine utilizes MDClone's synthetic data generation capabilities to:

- + Do preliminary data analysis for grant applications
- + Conduct analyses to include in abstracts and submissions for research conferences
- + Streamline the process for preparing submissions to scientific meetings and journals

Dr. Foraker is using synthetic data to support the development of new cardiovascular drugs. Synthetic data enables a faster research process, with less cost and more flexibility.

In a paper published in the peer-reviewed journal *Frontiers in Digital Health*, Dr. Foraker and her colleagues used synthetic data generated by MDClone to predict the risk of death in patients with heart failure within one year of their diagnosis: “ML models have considerable potential to improve accuracy in mortality prediction, such that high-risk surgical intervention can be applied only in those patients who stand to benefit from it.” The authors concluded that using synthetic data “uniquely allows a broader application of our results by enabling the sharing of data without risk of exposure” of individual patients’ information.

“Every time we interact with patients, we have an opportunity to learn from the experience and to improve their care, the care their family receives, and the care their community receives. If you know more, you can do more.”

Dr. Philip R.O. Payne, Ph.D., FACMI, FAMIA
Director of I2 and Chief Data Scientist at Washington University School of Medicine



TAKE THE FIRST STEP

Get Started Today

**Unlock healthcare data.
Transform care.**

sales@mdclone.com
mdclone.com/see-a-demo

SEE A DEMO

MDCLONE

