case study

Arkhn is improving how hospitals and researchers safely and privately store and access data

Here’s why the French company piloted Google’s differential privacy technology.

Theo Ryffel,
Co-Founder of Arkhn

Executive Summary

➜ Arkhn is creating a safe and secure database of information from hospitals, and experimenting with Google’s open source privacy-protecting technology to help.

Theo Ryffel co-founded the French data-management company Arkhn in 2019 to help hospitals gather data in a secure, safe, and uniform way. Ryffel, 25, pictures a not-so-distant future in which artificial intelligence has revolutionized the healthcare industry: finely-tuned algorithms help pharmaceutical labs develop lifesaving medicines faster, assist overwhelmed ER staff with triage, or provide physicians with a second opinion in fractions of a second. The biggest roadblock standing in the way of that future, Ryffel believes, is not technology but data and data security. And he thinks privacy-protecting technology like Google’s differential privacy (DP) library is a key to this exciting future.

Managing the vast amounts of patient data required to train these algorithms poses daunting logistical and data privacy challenges.

“Every department in a hospital uses its

ARKHN

17

EMPLOYEES

8

HOSPITALS

FOUNDED IN PARIS IN

2019

own specific software to generate data on the patient, yet also requires data from other departments in order to function," says Ryffel, who studied data science and machine learning at Polytechnique in Paris and Imperial College in London. "This creates confusion, in that doctors need to log on to many different applications in order to retrieve the information they require to make a diagnosis." These processes are time-consuming, and it's not automatable because each system stores data in a proprietary format.

Arkhn can collect data across all departments and store it in a standard format called Fast Healthcare Interoperability Resources, or FHIR (pronounced "fire"). The result is a trove of anonymized yet searchable patient data—collected with patient consent—that researchers, developers, and programmers can use to advance science and medicine while ensuring patient privacy.

The key is finding a way to make these search results specific enough to be useful, yet not so specific that they compromise patient anonymity or raise privacy concerns. For this, Ryffel and his research team at Arkhn turned to Google. With Google's open source differential privacy library they were able to quickly deploy the same world-class anonymization technology that Google uses every day in its products like Google Maps and Assistant to introduce statistical noise into their data to prevent the identification of patients.

To explain how DP works, Ryffel proposes a hypothetical scenario in which it isn't applied. Imagine that a pharmaceutical company wants to identify and study a cohort of French diabetic women over 65 in order to develop a new diabetes drug. The company enters those parameters into Arkhn's search tool, which combs through the records of every hospital in Arkhn's network to pinpoint the exact number of patients that match the search. "Let's say every time you run the search, you get 12 people," says Ryffel. "Then, one day, you get 11. Let's say you also have access to the hospital's log—or perhaps you're simply standing outside the building watching comings and goings—and you notice that a patient has signed out. It's very likely, then, that this person is diabetic and over 65."

If DP is applied, however, that kind of triangulation isn't possible. "Instead of showing you 12 people, it might show you 10 or 14," Ryffel says. "That minimal imprecision doesn't make a difference to you as a research or pharmaceutical company. All you need to know is there is a sufficiently large cohort to justify reaching out to the hospital and going through the process of identifying patients and getting their consent to participate in a study."

Experimenting with Google's DP library was simple. "The tools Google produces are very clear and very secure," says Ryffel. "Any company can easily and effectively integrate Google's differential privacy algorithms into their own work."

While Arkhn is still in its infancy, the company sees massive potential in helping hospitals and researchers better manage their data in a safe manner that protects the privacy of individuals. The project, he says, is a first step toward the AI-enhanced future of healthcare he imagines, which is made possible with the help of Differential Privacy: since DP introduces uncertainty only at the margins, the accuracy of aggregate data isn't in question. "In a sense," says Ryffel, "we're laying the groundwork for machine learning." Research breakthroughs will follow with data that is actionable, safe, and secure.

> “The tools Google produces are very clear and very secure," says Ryffel. "Any company can easily and effectively integrate Google's differential privacy algorithms into their own work.”