



## HOW XPATTERNS SAVES MORE THAN 4,000 HOURS IN DATA PREPARATION EVERY MONTH

Investments in John Snow Labs accelerate data operations and increase quality



Data scientists and the teams they serve can easily spend [50 percent to 80 percent of their time](#) preparing and standardizing data for analysis. In the healthcare industry, this includes regularly updating medical indexes and dictionaries to comply with federal mandates, as well as curating and verifying lists of known cybersecurity threats. This is called **data operations (“DataOps”)**, or among more beleaguered data scientists: “data wrangling.”

For [Atigeo](#), a technology company whose flagship big data analytics product, [xPatterns](#), generates knowledge from all available data to deliver insights, predict outcomes, and mitigate risks, DataOps is a necessary priority—but not its first. To reduce the time spent finding, cleaning, formatting, updating, and publishing data for analysis, the team of healthcare data analysts at Atigeo partnered with [John Snow Labs](#) (JSL), a DataOps company whose mission is to “accelerate the use of data to improve human well-being.” On both a practical and ideological level, JSL was an apt fit for Atigeo, whose own charge in healthcare is to help organizations leverage its technology to drive operational efficiency, improve clinical and financial performance, and provide the most compassionate, patient-centric care.

Working with JSL and its data libraries—which include healthcare provider indexes, billing codes, and sets of approved drugs and medical terminology—Atigeo saves an estimated 4,096 hours preparing data every month. “Updating the Master Provider Index [MPI] is a ten-minute process with JSL,” says Claudiu Branzan, Principal Lead of Software Engineering at Atigeo. “We’ve dramatically accelerated our ability to update the MPI using JSL’s curated datasets, replacing a process that used to be two days of work.” This means Branzan and his team can update the MPI at a rate **96 times faster** than before.

### Inside the numbers

xPatterns must ingest approximately...

- 200 frequently changing datasets (monthly on average) like the Master Provider Index
- 100 complex datasets—twice-yearly updates that now take half a day instead of a week
- 250 infrequently changing datasets—one update per year, on average

Now that both frequently and infrequently changing datasets take about ten minutes instead of two days to update, and complex datasets take about half a day instead of one week to update, Atigeo saves an estimated...

- 3,166 man-hours making monthly updates to frequently changing datasets
- 600 man-hours making twice-yearly updates to complex datasets,
- 330 man-hours making yearly updates to infrequently changing datasets,

...for a grand total of **4,096 man-hours saved on average, every month.**

This equates to 24 man-months, or **24 full-time employees** updating datasets every hour of every month.

## Domain expertise delivers quality data for wiser decisions

Domain expertise drives the high quality and corresponding ease of use of JSL datasets. As a company specializing in DataOps for healthcare analytics, its teams contain real doctors, among them data researchers with medical degrees from Harvard and doctorates in cognitive psychology. [Experts of this caliber](#) act as...

- **librarians**, dissecting broad healthcare inquiries to find data scientists the exact information they need,
- **domain consultants**, directing data scientists to the most relevant datasets for a given healthcare problem,
- **alchemists**, researching and simulating data to fill privacy-related gaps in datasets,
- and **engineers**, recommending the right tools for specific datasets.

High-quality medical data requires high-quality protection. At JSL, medical professionals work alongside hackers, Certified Ethical Hackers trained to detect system vulnerabilities using the same knowledge and tools as bad actors—but with lawful methods and legitimate intentions.

## Realized benefits of higher quality data

The JSL combination of industry expertise and cybersecurity intelligence helped Atigeo identify **4.5 times more cyber threats** for one customer. Over a period of four months, Atigeo analyzed traffic at its enterprise client using two different datasets: a JSL blacklist of known cyber threats containing 5.3 million Internet Protocol (IP) addresses, and an aggregate blacklist of three commercial datasets containing 1.3 million unique IPs. Using the JSL list, Atigeo identified 19,258 blacklisted IPs in the organization's cyber traffic. Using the aggregate list, Atigeo identified 4,234 IPs—3,335 of which were also on the JSL blacklist. With 15,024—4.5 times—more blacklisted IPs identified through JSL, the evidence was clear: a longer list is a safer one.

Today, Atigeo has more time to deliver specialized, actionable knowledge in real time to strengthen sound judgments for providers, administrators, and patients alike. For caregivers, xPatterns brings informed decision-making at the point of care and predicts risks after care; for administrators, it unifies data across existing tools, extracting information from structured and unstructured notes. And for patients, this brings outcomes such as fewer avoidable readmissions, timelier interventions, and better protection. Free from the manual burden of finding, cleaning, formatting, updating, and publishing data for analysis, Atigeo can focus on what matters most in healthcare: health, and the data science that advances it.