In 2011, the FDA issued a black box warning on Celexa, one of the most widely prescribed antidepressants in the US. At higher doses, the drug had been linked to potentially dangerous changes in the electrical activity of the heart (a so-called prolonged QT interval). As a result, physicians began switching patients to a similar drug called Lexapro. "It was absurd because Lexapro and Celexa are almost exactly the same chemically," says Shawn Murphy, MD, PhD, associate professor of neurology at Massachusetts General Hospital. "But because the FDA did not conduct a study of patients taking Lexapro, it didn’t have a black box warning."

Rather than wait a few years for the FDA to gather new data about Lexapro, Murphy and his colleagues decided to look at data that already existed in electronic medical records. And to do that, they turned to the i2b2 open-source software suite. Developed by i2b2 (Informatics for Integrating Biology and the Bedside), a National Center for Biomedical Computing (of which Murphy is a part), the i2b2 platform enables researchers to use existing clinical data for discovery research.

To look at Lexapro’s side effects, the team mined electronic medical records from more than 38,000 patients with both antidepressant prescribing data and heart monitoring data (EKGs) available. "And we could actually show—just based on these previously acquired EKGs—that Lexapro was not only causing a prolonged QT interval, it was even a little bit worse than Celexa," Murphy says. The study was published in BMJ in 2013 and the results are being reviewed by the FDA. "That was an adverse event that we could detect just by going back and mining the electronic data that had been collected in the normal course of clinical care on our patients," Murphy says.

Institutions are moving to i2b2 in large numbers because of its proven track record; large user group; and open-source license. "There’s not a general competitor for the i2b2 platform only because the terms of i2b2 are so liberal," Murphy says. "There’s not a general competitor for the i2b2 platform only because the terms of i2b2 are so liberal," Murphy explains. "The most rewarding and unique thing about i2b2 is that it really is just the starting point for people who want to develop their own tools at their own hospital." Cancer centers, for example, might use their own data to develop their own survival plots—graphs showing the long-term benefit of various treatments or the prognosis for people with different cancer types—while children’s hospitals might want to build growth curves. "They’re totally enabled to develop it in ways they see fit for their use cases," Murphy says.

To capture a snapshot of i2b2’s impact, informatics leaders at two hospitals that have been active in the i2b2 community:...
University of Kansas (KU) Medical Center and the Cincinnati Children’s Hospital.

Kansas: Extending i2b2 and Connecting to Other Institutions

In 2010, when Russ Waitman, PhD, was recruited to be KU’s director of medical informatics, he made the critical decision to build the university’s clinical research informatics infrastructure on i2b2. Given a limited budget, he says, “I couldn’t afford to sit here and be yet one more informatics shop that’s going to reinvent the wheel.” Instead, he decided to “take the wheel from Harvard and build a car.”

i2b2 allows investigators to identify cohorts of patients for research studies in a self-serve manner. Researchers can do limited analyses on the de-identified data and if the results are promising can then request approval from their institutional review board to obtain the full dataset. In 2013, the KU system served 142 users with 4,751 queries; and fulfilled 53 further requests for full datasets. Waitman’s team has also expanded on what i2b2 can do. For example, they can now integrate hospital data with data from cancer registries; and statistical analysis plugins they created allow users to analyze datasets on the server, without having to download them to their computers (thus protecting the data from potential release or loss of privacy protections).

Waitman’s team is leading an effort to share data across multiple hospitals using the i2b2 platform. They recently received a Clinical Data Research Network contract ([http://www.pcornet.org/clinical-data-research-networks/](http://www.pcornet.org/clinical-data-research-networks/)) to link ten health systems that are already using i2b2 in Kansas, Iowa, Missouri, Wisconsin, Minnesota, Texas, and Nebraska (the Greater Plains Collaborative). Waitman is excited about this, he says, because researchers will be able to perform queries across records for the 6 million patients at these institutions.

Using the i2b2 platform, electronic medical records can also be linked to biological samples collected during routine hospital and clinic visits. Initially, Waitman’s team hopes to work with surgical pathologists to determine whether samples held by their hospitals could be useful for research. Waitman’s team is also looking at ways to use the data stored in i2b2 to drive hospital quality improvement. For example, data from electronic medical records (EMR) could reveal how well different doctors are meeting goals for diabetic control. “Because that’s what I think is going to make it really sustainable,” Waitman says. “You have your core clinical research covered with your EMR data and then you want to blend downstream to the biological and then you want to blend upstream to get the health systems benefiting from it.”

Cincinnati: Boosting i2b2 User-friendliness and Building Disease Registries

“We’ve had our hands dirty with i2b2 for quite a while now,” says Keith Marsolo, PhD, associate professor of pediatrics. When he was recruited to Cincinnati Children’s in 2007 to build a research data warehouse, he surveyed available platforms. Some institutional software cTAKES (clinical Text Analysis and Knowledge Extraction System). Much of the relevant data in electronic medical records is locked away in narrative notes. But open-source text-mining tools such as cTAKES can extract critical information from free text.

And Cincinnati investigators are already using the system to make novel discoveries. In a December 2013 paper in the *Journal of Pediatric Gastroenterology and Nutrition*, Marsolo and colleagues showed, for the first time, that two rare childhood diseases—eosinophilic gastrointestinal disorders (EGID) and pediatric PTEN hamartoma tumor syndromes (PHTS)—are strongly associated. By querying their i2b2 data warehouse and a local eosinophilic database, they were able to search through more than one million patients to find eight with confirmed PHTS, five of whom also had EGID.

Marsolo’s team has also focused on creating multi-center disease registries using i2b2’s data sharing software SHRINE (Shared Health Research Information Network). Chronic childhood diseases are often hard to study due to the small number of patients at any single institution. But SHRINE allows institutions to share data while protecting patient privacy. Marsolo’s team helped create the 60-site CARRAnet (Childhood Arthritis and Rheumatological Research Alliance); with more than 5,000 patients, this registry is the largest available for pediatric patients with rheumatic disease.

Cincinnati Children’s is also one of two children’s hospitals involved in the eMERGE (electronic Medical Records and Genomics) network, which links electronic medical record data with genetic samples (and also uses i2b2). Using data from this network, Marsolo and colleagues performed a genome-wide association study looking for genetic variants associated with obesity in children. Their results confirmed previous results from adult studies and also identified novel variants in kids. The results were published in *Frontiers in Genetics* in December 2013.

The future of i2b2 may depend on how electronic medical record companies evolve their tools going forward, Marsolo says. For example, EPIC, a medical software company, has recently come out with a tool for identifying cohorts of non-anonymized patients. Even if a de-identified version of that tool “might change the value proposition and primary use case for i2b2,” Marsolo says, i2b2 will likely endure because it is customizable and fosters data sharing across multiple institutions.