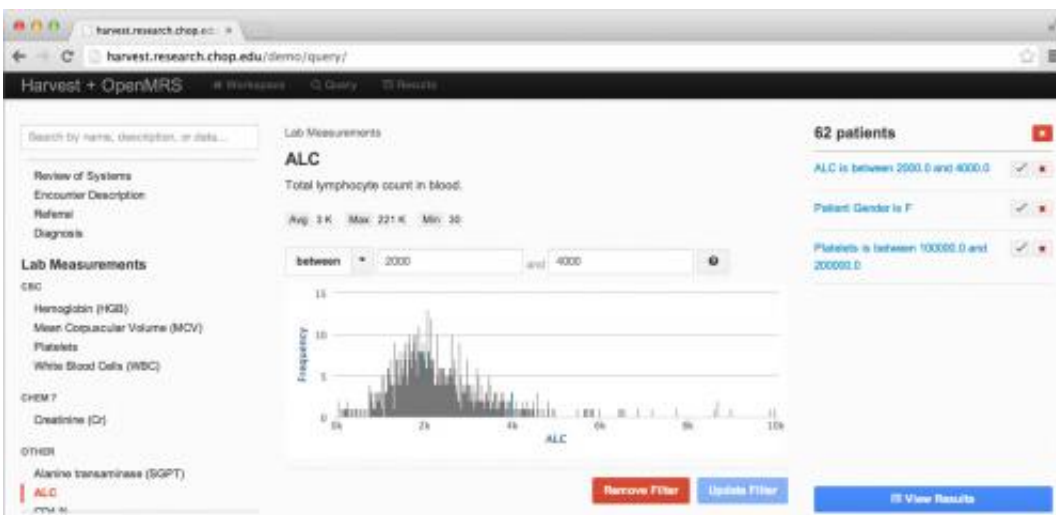


Harvest toolkit offers innovative data discovery resource for biomedical researchers

October 17 2013



Harvest users can query sets of data such as blood test results to easily display results by a variety of criteria, e.g., to create subgroups by gender or by those within a specific range of platelet counts or leukocyte counts. Credit: The Children's Hospital of Philadelphia

Biomedical researchers often confront large quantities of information that may be amassed in many forms: vital signs, blood cell counts, lengthy DNA sequences, bar graphs, MRIs, patient demographics, and so much more. How do researchers assemble, access and analyze all that data without having to become specialized database technicians themselves?

A team of informatics experts and [biomedical researchers](#) at The Children's Hospital of Philadelphia (CHOP) offers a new software toolkit to help researchers wrangle [data](#). Their open-source, highly interactive framework called [Harvest](#) is designed to let users to navigate quickly among different types and levels of data.

"We want to help researchers explore their data, not their database," said Byron Ruth, lead developer of Harvest at CHOP's Center for Biomedical Informatics (CBMi). Ruth is one of the co-authors of a paper introducing CBMi's Harvest framework, appearing online Oct. 16 in the *Journal of the American Medical Informatics Association*.

Co-author Michael J. Italia added, "Research institutions typically work through their information technology staff to provide a single data warehouse that may be too general-purpose for all its projects, or develop one-off solutions on a case-by-case basis for each project."

"Our approach in Harvest is different," said Italia, the CBMi's manager of Applications Research. "We decided to focus on end-users, generalizing the toolkit for application to any biomedical study with multiple collaborators, but also allowing individual software developers and data managers to customize the software for specific projects."

Harvest, said Italia, "isn't just shrink-wrapped, ready-to-go software." He estimates that Harvest typically provides 80 percent of the work, leaving it to any institution's [software developer](#) to adapt the framework to a project's needs, in collaboration with each project's principal investigator. Harvest is open-source, so users are free to see bug reports, check software patches, and share fixes and customizations with a wider community of users.

A key feature of Harvest is the ability to maneuver smoothly among various levels of data, from individual patient records to aggregated

reports of all patients in a database, and to subpopulations in between. Users can construct queries to slice and dice data—grouping subjects, for instance, by age or ethnicity, calling up individual blood test results or MRIs, or including or excluding specific diagnoses.

One advantage of Harvest is that it provides transparency and visibility to data in a manner that is familiar to a researcher who is invested in a particular disease or project. "Harvest adopts convenient and clear interfaces to view and explore data that are increasingly used in other industries, such as social media," said senior author and CBMi director Peter S. White, Ph.D. "We have found that this often helps users to quickly familiarize themselves with data they are seeing, which increases the likelihood that they will trust the resource, and even be incentivized to contribute to the project as it develops."

The Harvest developers say their tool reflects the growing complexity of research in the Big Data era of [electronic health records](#) and genomic technology. In the 1980s and 1990s, much federal research followed a hypothesis-driven model, focusing on predefined measurements within a patient population. Many current databases collect vast amounts of many data types with fewer preconceived notions of what is significant. "Harvest allows users to formulate and refine questions and even explore a different view of data that wasn't apparent or important to them a few minutes before," said Italia.

The CBMi team originally designed Harvest to manage data from AudGenDB, an audiology database funded by the National Institute on Deafness and Other Communications Disorders.

In the current paper, the study team evaluated Harvest by running it through its paces with data from two other collections: CardioDB, which stores clinical data from 47,000 of CHOP's pediatric cardiology patients, and OpenMRS, a public dataset holding lab results, infection status and

other clinical results from electronic health records.

CBMi is now applying Harvest to more collections, including the Longitudinal Pediatric Data Resource, a long-term informatics system being created to store clinical data derived from thousands of children with conditions detected in national newborn screening programs.

"Although these collections contain very different types of data, the Harvest toolkit is flexible enough to adapt to these different scenarios, while at the same time maintaining a consistent set of basic components," added Ruth.

Harvest has been focused primarily on the end-user up to this point. Future work includes a bigger focus on software developers and applications support personnel who are responsible for setting up and maintaining Harvest applications. "However, as we think about the life cycle of a Harvest application, we want it to be as easy as possible for administrators and [software](#) developers to get Harvest into the hands of researchers," Italia said.

More information: "Harvest: An Open Platform for Developing Web-Based Biomedical Data Discovery and Reporting Applications," Journal of the American Medical Informatics Association, published online Oct. 16, 2013. jamia.bmj.com/content/early/2013/10/16/amiajnl-2013-001825

Provided by Children's Hospital of Philadelphia

Citation: Harvest toolkit offers innovative data discovery resource for biomedical researchers (2013, October 17) retrieved 26 April 2024 from <https://medicalxpress.com/news/2013-10-harvest-toolkit-discovery-resource-biomedical.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private

study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.