Extending an I2B2-based Clinical Data Repository with the R Statistical Platform

Daniel W. Connolly, Lemuel R. Waitman
University of Kansas Medical Center, Kansas City, Kansas

HERON[1] includes the usual i2b2 cohort query facilities on a largely self-service basis as well as bulk export of data for off-line analysis with approval by an oversight committee. We aim to support richer analysis without the need for bulk export.

We have adapted the survival analysis plug-ins from the **R Engine Cell** work[2] to address scalability limitations and provide secure extensibility. Our approach, **rgate**[3], has one privileged database access module, which uses R and the ROracle package to run SQL queries directly against our data warehouse, along with any number of unprivileged analysis scripts. The analysis scripts for our survival plug-ins use the R survival package for plotting.

RStudio Server[4] provides an integrated development environment (IDE) for R over the web. HERON integrates with RStudio via an **R Data Builder** plug-in (Figure 1). Like the i2b2 timeline, this plug-in takes a patient set and a set of concepts as input (Figure 2). The patient set identifier and concept paths, along with a filename of the user's choosing, are sent via rgate to the R Data Builder analysis script, which queries the data warehouse and saves the results on the server. Then the user logs in to RStudio Server and uses the R readRDS() function to load the data for further analysis (Figure 3). Like i2b2, and unlike bulk export, this approach is designed to keep the data on the server in our data center.

Applications

The survival analysis plug-ins are available to the entire HERON user-base. In addition to typical applications such as 5-year cancer survival plots, students have used it to visualize 28-day sepsis outcomes.

The HERON study team has piloted the R Data Builder and RStudio Server in work-in-progress on a quality-improvement investigation into medication timing.

We have not established governance policies for use of RStudio Server by the general HERON user-base. While HERON users execute a system access agreement that prohibits screenshots, printing, and other data export, the risks associated with RStudio Server are significantly higher than with I2B2 plugins. For example, R includes facilities to email a data set, and we have not yet put in place any technical limitation on such facilities.

---

[1] Waitman, Lemuel R et al. "Expressing Observations from Electronic Medical Record Flowsheets in an i2b2 based Clinical Data Repository to Support Research and Quality Improvement." *AMIA Annual Symposium Proceedings* 2011: 1454.

[2] Segagni, Daniele et al. "R Engine Cell: integrating R into the i2b2 software infrastructure." *Journal of the American Medical Informatics Association* 18.3 (2011): 314-317.

[3] Connolly, Daniel W et al. "Integrating R efficiently to allow secure, interactive analysis within a clinical data warehouse." R User Conference, Nashville TN, June 2012

[4] Racine, Jeffrey S. "RStudio: A Platform-Independent IDE for R and Sweave." *Journal of Applied Econometrics* 27.1 (2012): 167-172.

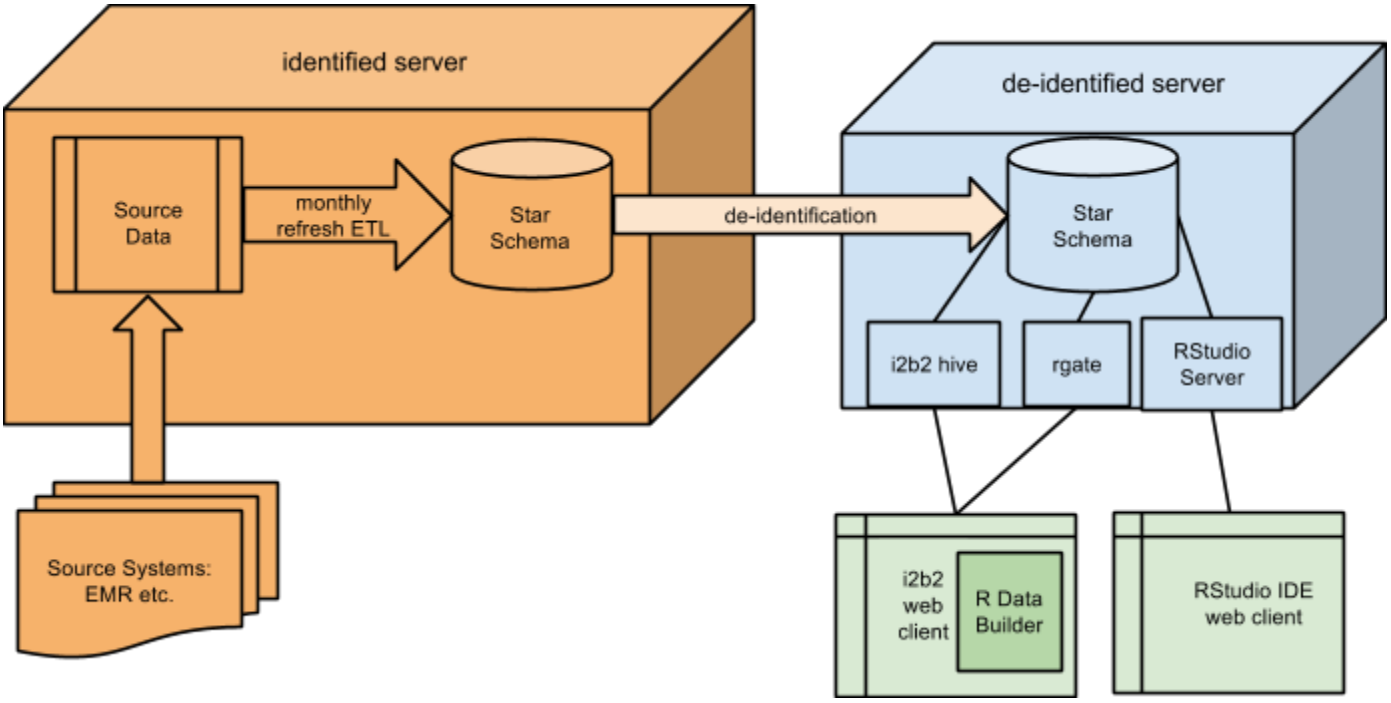

Figure 1: HERON Architecture.

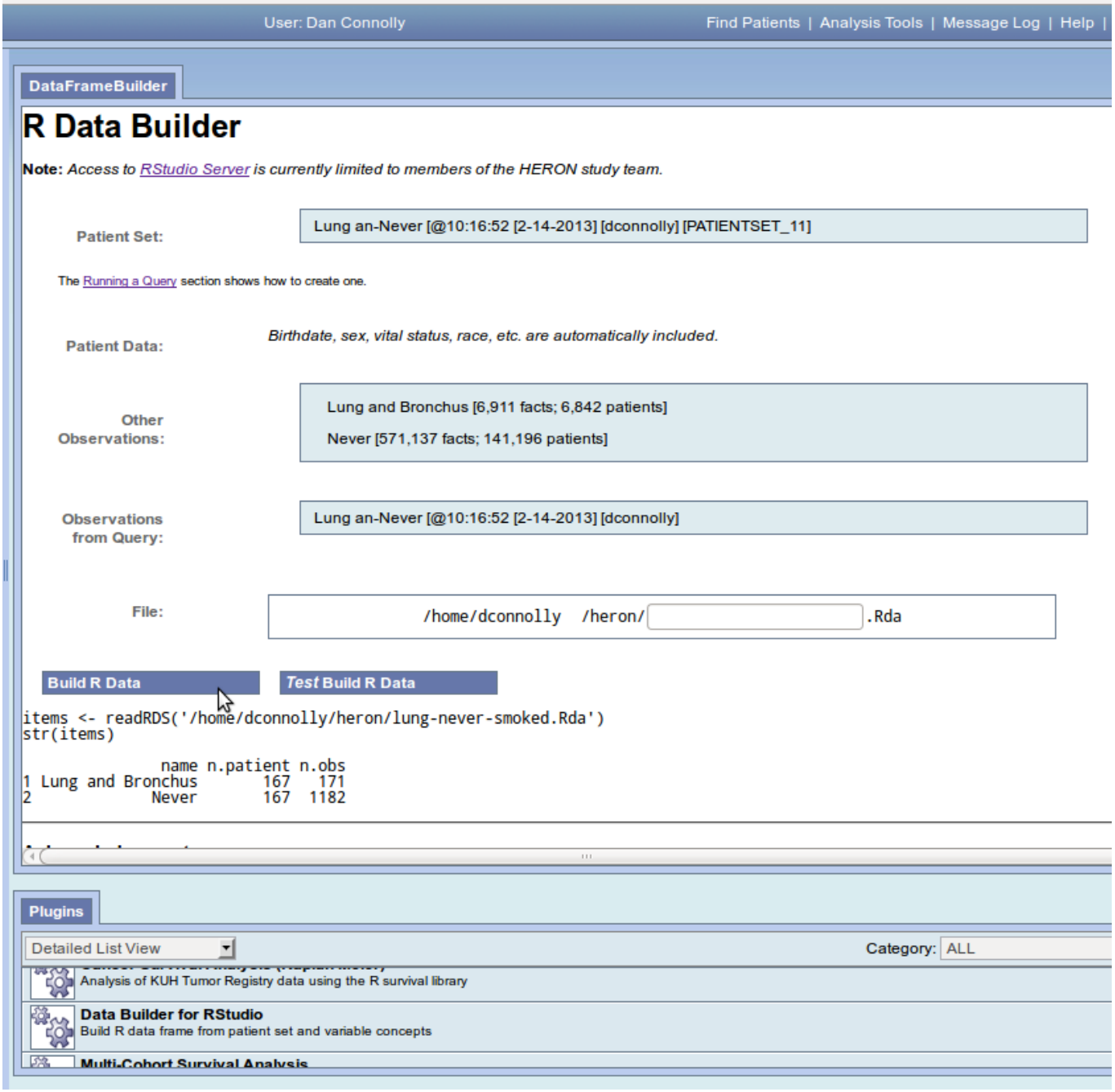

Figure 2: R Data Builder Plug-in User Interface

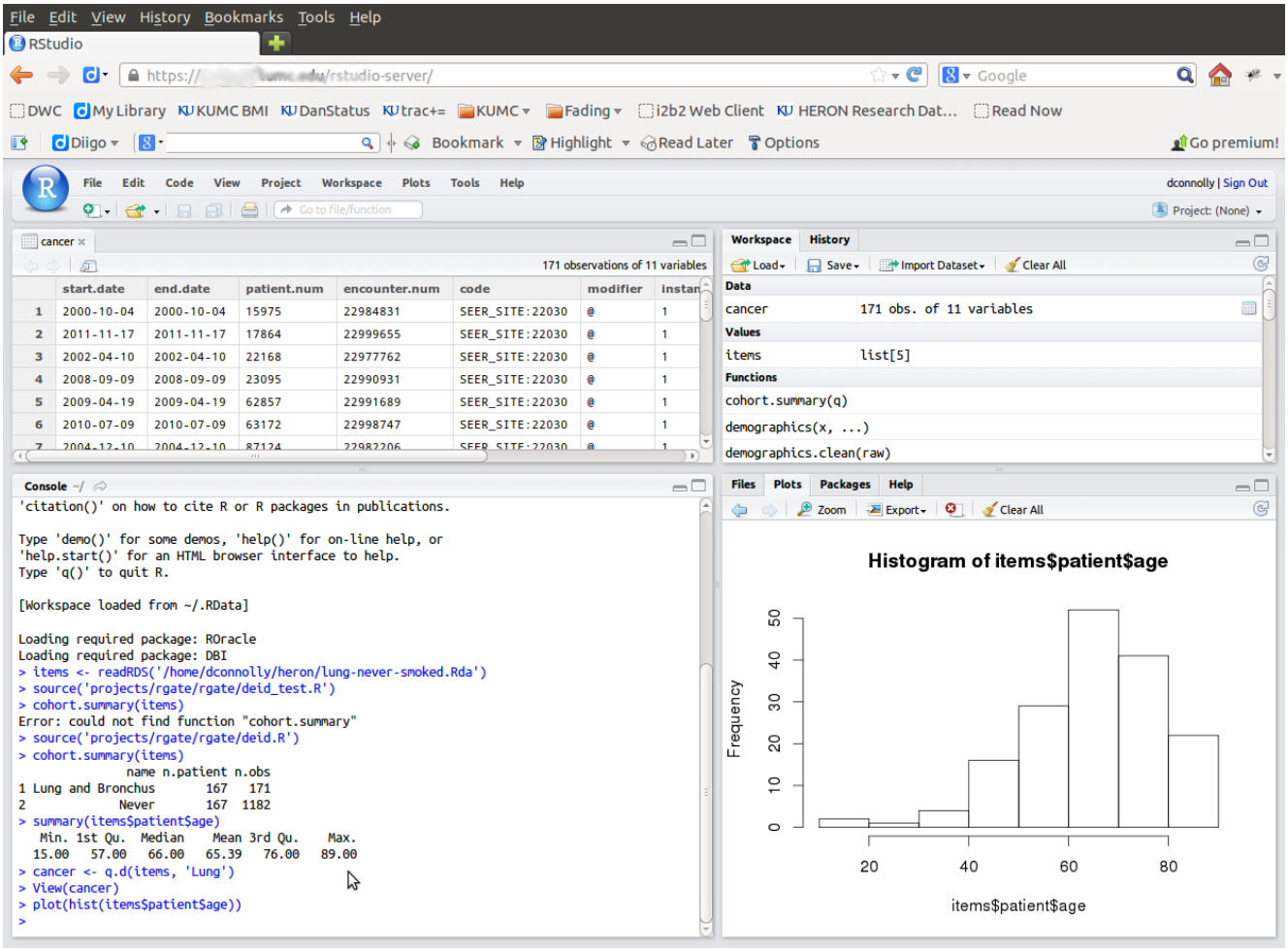

Figure 3: Analyzing data from I2B2 in RStudio.