Recap of MAVERIC Wiki Changes

Andy Zimolzak, MD, MMSc

November 19, 2014

The MAVERIC wiki[1] was created in December 2010. It was evidently used a fair amount through October 2011[2], but with virtually no edits in 2012–2013. Starting May 2014, usage has increased.[3]

Problem/niche

The Precision Oncology Project and the Point of Care Trial project rely in part on a workflow that goes roughly as follows. *First,* the subject matter expert knows that we want to track the effect of drug x on (for example) serum sodium. The SME attempts to operationalize the concept of serum sodium, suggesting a list of terms to be searched for in a field (e.g. `LabChemTestName`) of CDW. *Second,* the data pull team returns a list of lab tests matching these search terms.[4] *Third,* the SME reviews the list of tests and decides (adjudicates) which items truly match the concept. The SME returns either a refined general rule to specify which items match the concept, or an enumerated list of which items match the concept.[5] *Fourth,* the data pull team creates a query that represents this adjudication.[6] The clinical SME may give formal approval or "sign off" on the data pull at step 3, or possibly on the results of step 4.

The data pull and NLP teams intend to use the wiki to capture this workflow so that these deliverables are *available for viewing* by study coordinators or others, with "provenance" (responsible SME and data extractor) clearly noted; *available for reuse* in a library, for the purposes of other projects; and *available for renewal,* with their date of creation clearly noted, because CDW elements may change over time. The rationale for the SME including/excluding certain tests can also be documented.[7]

Recent wiki changes

October 23–29   I created several standard templates that can be used in multiple places, to avoid duplicating the same wikitext code. These are mainly for Data Pull and NLP subsections of the wiki. Also organized the structure of some wiki pages.

October 30   First actual use of the wiki to store adjudications— specifically to store "POP Additional Documents" (cancer-related

[1] http://vhabosapp8/wiki

[2] It was used for things like GenISIS meeting notes, documenting GenISIS & HPC server configurations, and documenting server setup procedures. Top contributors were Phsieh, Lalitha.viswanath, and Mmcduffie.

[3] Now used for DCP, POP, data pull, "phenotype models," and NLP. Top contributors: Rpourali, Nilla, Beth, Andy Z. Now there are 54 content pages, and 11 users who have been active in the last 30 days.

[4] The list sent to the SME is often an Excel file. Data extractor probably generates SQL code in the process.

[5] So far I return this adjudication to Beth in the form of an edited Excel file.
[6] probably generating refined SQL code.

[7] The basic idea is that the methods of science should be documented well enough to be reproducible—in other words, "show your work." For our type of science, *code is method.* See also http://sciencecodemanifesto.org

radiology notes) and sodium.

November 5   Shadow wiki created for testing, using MediaWiki version 1.23.6 instead of 1.16.0. This will allow us to use new MediaWiki extensions to do two things: write wikicode templates that use conditional logic,[8] and do syntax highlighting of source code snippets. Ned and I are also given shell accounts on `vhabosapp8` thanks to Paul.

November 6–10   Magnesium and ALT lab adjudications uploaded.

November 12   Shadow wiki becomes real wiki.

November 12–14   POP surgical procedures and AST adjudications uploaded.

Suggestions for you all in future

Look at the wiki.[9] Make a user account. Edit things (be bold). Ask if you need help.[10] Don't upload enormous things. Use talk pages.

Questions for the group

- What should we call the deliverable that the data pull expert and the SME create in collaboration?[11]

- Not much complete SQL is on the wiki currently. Is the wiki the right place for SQL code? Or should the wiki merely have a pointer to a source control repository?

- What training on the wiki do you want?

- How does wiki fit in with data pull request tracker?

- Should SMEs deliver general rules or enumerations?

- Do you want an e-mail every time a new algorithm is uploaded?

- What should be captured about each algorithm?[12]

Further reading about MediaWiki and our server

"A wiki is a Web application which allows people to add, modify, or delete content in collaboration with others."[13] MediaWiki is the software that makes Wikipedia run, but anyone can install it to their Web server to create their own wiki. MediaWiki uses its own markup language called wikitext.

[8] An example of invoking such a template, used on the NLP dashboard page, looks like this: `{{Phase|Requirements}}`. This expands to a nice looking table with one cell highlighted. Elsewhere this template is invoked as `{{Phase|Development}}`, which expands to the same table with a different cell highlighted. This avoids a *lot* of copy-pasting of wikicode tables.

[9] Reminder! http://vhabosapp8/wiki

[10] Andrew.Zimolzak@va.gov

[11] Thoughts: data model, concept model, algorithm, adjudication, library, phenotype, phenotype model, study data dictionary.

[12] Currently caputured on basic algorithm pages on wiki: algorithm version, SQL statement, SME name, approval date. On the more detailed wiki form we have the following fields: lab name, sample source (body site), date created, data source, SQL search terms used, clinical adjudication list (link to Excel file), adjudicator name, SQL data extractor name, extracted data location, content of extracted data, extraction verified by:, SQL extraction code.

[13] http://en.wikipedia.org/wiki/Wiki

The `vhabosapp8` server runs Red Hat Enterprise Linux, Apache, MySQL, and PHP. It also hosts Subversion repositories and the Jira issue and project tracker. Currently 13 out of 74 GB of space is free on `vhabosapp8`.