THE UNIVERSITY of NEW MEXICO

Does the VIVO Harvester Populate Bibliographic References for an Institutional Instance of VIVO Effectively?

Jonathan Eldredge, Philip Kroth, Gale Hannigan, Cristina Murray-Krezan, Richard Carr, Edward Weagel, and Chad Hantak

EBLIP 7 Saskatoon, Saskatchewan. July 16, 2013

OBSERVATION

THE UNIVERSITY of NEW MEXICO

For as long as libraries have existed, librarians have tried to connect people with similar interests:

- Traditionally, through linking learners and scholars to recorded ideas and information;
- Less conventionally, by linking scholars with one another, directly and interpersonally

OBJECTIVE

To test the accuracy of the VIVO Harvester for identifying publications indexed in PubMed™ by research faculty authors linked to an institutional instance of VIVO.

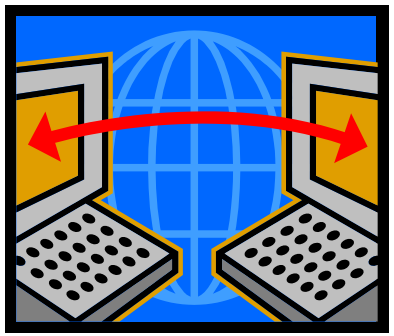
BACKGROUND

THE UNIVERSITY of NEW MEXICO

The VIVO Collaboration seeks to build an:

- Authoritative
- Open-source
- Dynamic
- Web-based
- Social/collaborative

network of research scientists.



The VIVO network will enable scientists to identify colleagues with similar interests or complementary skills in a manner never before possible in order to speed new scientific discoveries.

Busy scientists do not have the time to assemble all authenticated information on their publications or grants to populate their individual profiles manually at their institutions

Critical to VIVO's success will be building the capacity for automatic data ingestion from a variety of authoritative data sources such as the NIH Reporter and PubMed

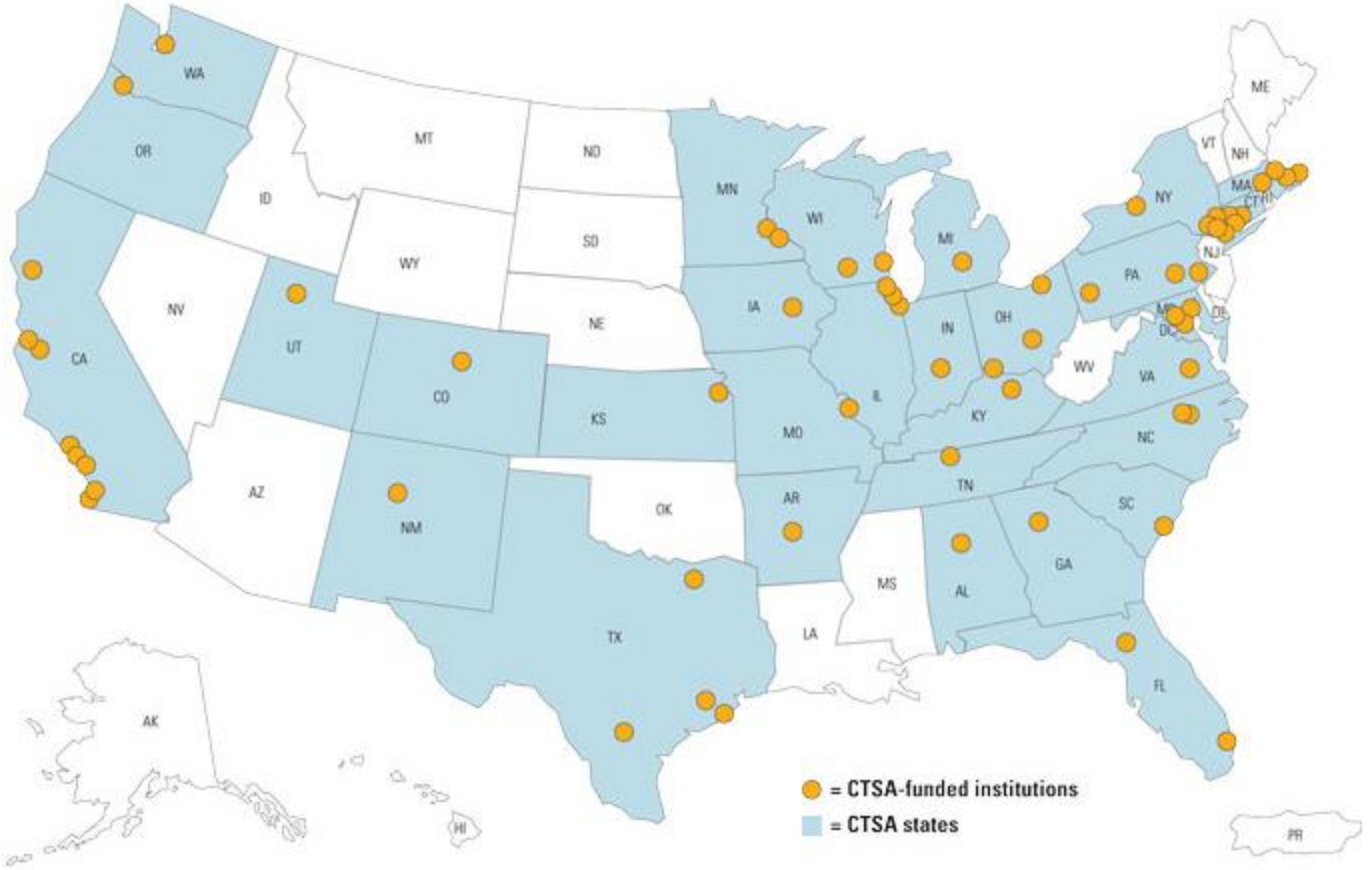
The VIVO Harvester seeks to generate authoritative lists of faculty publications automatically from PubMed.

METHODS

- Comparison Study
- Population: Faculty members at the University of New Mexico with a formal affiliation with the UNM Clinical and Translational Sciences Center

- The Center represents one of 60 such centers in the US funded by the National Institutes of Health

Map of CTSA-Funded Institutions



Population

- The definition “formal affiliation” seemed precise by the standards of most research studies involving human populations
- This aspect turned out to be a major challenge since all faculty members already had formal, primary affiliations with other academic departments at UNM instead with the Center
- After several iterations, we developed a list to identify and recruit participants

- From a list of 108 affiliated faculty members
- We stratified members first by:
  - Basic sciences
  - Clinical
  - Pharmacy
  - All other disciplines

- We then randomly selected proportionate numbers of faculty members from each stratification category
- The study population consisted of 25 faculty members

Gold Standard Searching

- Faculty librarians searched for each selected research faculty member for the inclusive years of 2001-2011.
- Faculty librarians consulted research faculty member CVs and verified questionable references or discrepancies with the faculty members.

- The list of verified publications became the “Gold Standard” publications list for the project.
- One author ran the VIVO Harvester against a non-public instance of VIVO in a naïve mode using a list of faculty members’ names in PubMed author format variations and the text words “New Mexico.”

- Team members compared the Gold Standard lists against the VIVO results for each of the 25 faculty members
- Team members carefully documented when the VIVO Harvester produced either false positives or false negatives

RESULTS

THE UNIVERSITY of NEW MEXICO

- The VIVO Harvester demonstrated high specificity but low sensitivity in identifying authors' publications from PubMed

THE UNIVERSITY of NEW MEXICO

Gold Standard

Harvester Results	(+)	(-)
(+)	419	4
(-)	222	7,417,645
	641	7,417, 649

- 7,417,290 is the total number of articles indexed in PubMed from 2000-2011 (inclusive) taken from : http://www.nlm.nih.gov/bsd/licensee/2013_stats/2013_Totals.html.

Statistical Analysis

THE UNIVERSITY of NEW MEXICO

- A paired sign-rank test performed by the team biostatistician (CM-K) compared the results from the Gold Standard to the results from the VIVO Harvester. The Gold Standard results were significantly different from the VIVO Harvester results ($p < 0.0001$) .

DISCUSSION

THE UNIVERSITY of NEW MEXICO

- VIVO allows faculty members to add missing publications from PubMed, but this option creates extra work that many faculty members may not want to lend to populating their complete publication records.

- VIVO's accuracy likely will depend on local experts who are familiar with designing PubMed queries based on a specific institution's characteristics and who can integrate these queries into their VIVO Harvester configuration.

CONCLUSION

THE UNIVERSITY of NEW MEXICO

The VIVO Harvester demonstrated a limited ability to accurately identify all publications by authors linked to an institutional VIVO instance using a naïve configuration of the software.

THE UNIVERSITY of NEW MEXICO


CLINICAL & TRANSLATIONAL
SCIENCE CENTER

ACKNOWLEDGEMENTS

- The authors thank Jason Bengtson for his searching assistance with this project.

- This project was supported, in part, by the National Center for Research Resources and the National Center for Advancing Translational Sciences of the National Institutes of Health through Grant Number 8UL1TR000041, The University of New Mexico Clinical and Translational Science Center.