

Rapidly Retrieve Clinical Information across Immense Data Sources: Introducing Reveal

John Leveille, d-Wise Technologies, Inc., Raleigh, NC, USA

ABSTRACT

The vast and disparate nature of clinical data challenges our ability to search and analyze. Using Reveal, you can now answer precise questions gleaning data from spanning disparate sources in seconds rather than hours. Questions like: Where is GI Bleeding? What hypertension events were seen across this compound? Reveal's search technology eliminates the need for multiple manual, time-intensive searches through various clinical data systems and uniquely reaches sources previously impossible to search. It bridges the gap between data and analytic tools so users avoid spending time hunting for the information and, instead, focus on the analysis to answer key scientific questions.

SHINE A LIGHT ON CLINICAL DATA

Pharmaceutical companies are awash in data, and it isn't getting easier. Increased regulatory requirements, deepening medical sciences, and mergers and acquisitions are adding to ever-growing mountains of data. You know how to store your data, but how do you find it again? Days later months later years later ?



SEARCH FAR AND WIDE

Reveal is designed to perform a federated search across different repositories. You can build a single search index which includes multiple, heterogeneous data sources. Even if your data resides in different physical locations, Reveal can consolidate the information into one searchable body of knowledge.

EASY SEARCH

To use Reveal, you simply type in clinical terms to find your data. There is no need for complicated query languages like SQL.

The screenshot shows the Reveal web interface. At the top, there's a search bar with the query "blood AND pneumonia" and a "Help me" button. Below the search bar, it says "Found 43 hits in 0.215 secs." A table displays the search results with columns: Folder, Name, Type, Size, Modified, Relative Score, and Notes. The first three rows are visible, showing files from "studies/study8", "studies/study6", and "studies/study6". The "Notes" column provides a summary of matches for each file. At the bottom, there's a pagination bar showing "Page 1 of 2" and "Page Size: 25".

Folder	Name	Type	Size	Modified	Relative Sco	Notes
studies/study8	mh.sas7bdat	sasDataTabl	216 KB	08/18/11 3:23:02 PM		2 matches for 'blood'; 30 matches for 'pneumonia'
studies/study6	suppae.sas7bdat	sasDataTabl	152 KB	08/18/11 3:19:54 PM		33 matches for 'blood'; 2 matches for 'pneumonia'
studies/study6	mh.sas7bdat	sasDataTabl	168 KB	08/18/11 3:19:54 PM		1 match for 'blood'; 20 matches for 'pneumonia'

This picture shows the search results screen from Reveal after the user has entered the search

blood AND pneumonia

Reveal displays all of the files where both of the words blood and pneumonia occur. Search results appear in 1 or 2 seconds along with a relative ranking and file information.

PhUSE 2011

NARROW THE FOCUS

Using metadata-powered search terms you can narrow your search to a study path.

The screenshot shows the Reveal search interface. At the top, the Reveal logo is on the left, and the user is logged in as 'Reveal Administrator'. The search bar contains the query '\$path:*camd*'. Below the search bar, it says 'Found 160 hits in 0.219 secs.' A table displays the search results with columns: Folder, Name, Type, Size, Modified, Relative Score, and Notes. The results are filtered to show only datasets within the 'camd/1055' folder. The table lists 10 datasets with their names, sizes, and modification dates. At the bottom, there are navigation controls for page 1 of 7 and a page size of 25.

Folder	Name	Type	Size	Modified	Relative Score	Notes
camd/1055	sv.sas7bdat	dataset	145 KB	03/01/11 3:15:01 PM		
camd/1055	suppmh.sas7bdat	dataset	177 KB	03/01/11 3:14:59 PM		
camd/1055	qs.sas7bdat	dataset	1,953 KB	03/01/11 3:14:57 PM		
camd/1055	mh.sas7bdat	dataset	769 KB	03/01/11 3:14:49 PM		
camd/1055	lb.sas7bdat	dataset	7,985 KB	03/01/11 3:15:01 PM		
camd/1055	ds.sas7bdat	dataset	81 KB	03/01/11 3:14:46 PM		
camd/1055	dm.sas7bdat	dataset	49 KB	03/01/11 3:14:45 PM		
camd/1055	cm.sas7bdat	dataset	4,289 KB	03/01/11 3:14:47 PM		
camd/1055	cm.sas7bdat	dataset	1,185 KB	03/01/11 3:14:43 PM		
camd/1055	ae.sas7bdat	dataset	257 KB	03/01/11 3:14:46 PM		

Here we see the search results from the query

\$path: *camd*

Using table and column names you can narrow your search to have precise clinical meaning

The screenshot shows the Reveal search interface with the query 'lbtestcd:ALB'. It reports 'Found 19 hits in 0.191 secs.' The table shows results for various studies, listing the folder, name, type, size, modified date, and the number of matches for 'alb' in the 'lbtestcd' column. The results are sorted by the number of matches in descending order. The table lists 12 results, with the first 11 being 'sasdat' files and the last two being 'dataset' files. At the bottom, there are navigation controls for page 1 of 1 and a page size of 25.

Folder	Name	Type	Size	Modified	Relative Score	Notes
studies/study8	lb.sas7bdat	sasdat	3,224 KB	08/18/11 3		351 matches for 'alb' in lbtestcd
studies/study7	lb.sas7bdat	sasdat	896 KB	08/18/11 3		100 matches for 'alb' in lbtestcd
studies/study6	lb.sas7bdat	sasdat	2,312 KB	08/18/11 3		223 matches for 'alb' in lbtestcd
studies/study5	lb.sas7bdat	sasdat	920 KB	08/18/11 3		102 matches for 'alb' in lbtestcd
studies/study4	lb.sas7bdat	sasdat	152 KB	08/18/11 3		12 matches for 'alb' in lbtestcd
studies/study3	lb.sas7bdat	sasdat	1,184 KB	08/18/11 3		112 matches for 'alb' in lbtestcd
studies/study2	lb.sas7bdat	sasdat	608 KB	08/18/11 3		61 matches for 'alb' in lbtestcd
studies/study1	lb.sas7bdat	sasdat	1,088 KB	08/18/11 3		104 matches for 'alb' in lbtestcd
camd/1054	lb.sas7bdat	dataset	23,649 KB	03/01/11 3		786 matches for 'alb' in lbtestcd
camd/1053	lb.sas7bdat	dataset	87,873 KB	03/01/11 3		2,942 matches for 'alb' in lbtestcd

Here we see an example of searching for laboratory results of tests for albumin using the query

lbtestcd:ALB

Reveal can search documents just like other enterprise search technology. However, Reveal is unique in that it recognizes data sets which are organized into rows and columns. The data set metadata is extracted by Reveal along with the rows of data. Both of these are indexed and available for searching. This powerful combination of searchable content enables you to find just what you are looking for.

CASE STUDIES

CASE STUDY 1

PhUSE 2011

Problem: Post approval safety issues are being reported across a number of the company's drugs within Alzheimer's. The team needs to be able to search across studies to find data that might contain an increase in anxiety events and a potential interaction with Memantine, a common medication for Alzheimer's.

Search: A Reveal search:

Memantine AND Anxiety

displays results indicating that studies 1013 and 1014 have significant number of hits

The screenshot shows the Reveal search interface. At the top, there's a search bar with the query 'Memantine AND Anxiety' and a 'Help me' button. Below the search bar, it says 'Found 10 hits in 0.206 secs.' A table displays the search results:

Folder	Name	Type	Size	Modified	Relative S	Notes
camd/1014	cm.sas7bdat	Dataset	10,593 KB	03/01/11 3:28:51 PM		354 matches for 'memantine'; 154 matches for 'anxiety'
camd/1013	cm.sas7bdat	Dataset	10,961 KB	03/01/11 3:25:53 PM		227 matches for 'memantine'; 167 matches for 'anxiety'
camd/1052	cm.sas7bdat	Dataset	3,201 KB	03/01/11 3:06:43 PM		19 matches for 'memantine'; 47 matches for 'anxiety'

At the bottom, there's a pagination bar showing 'Page 1 of 1' and 'Page Size: 25'. It also indicates 'Displaying hits 1 - 10 of 10'.

Results: The search finds a number of medication data sources that indicate both memantine and Anxiety are present. Next, the team extracts those data sources into an analytical tool to perform further research and look for trends.

CASE STUDY 2

Problem: An entire submission has been outsourced to a CRO and that organization has delivered the data. All 14 studies in our submission should collect the ADAS-COG questionnaire. The CRO should have verified this result.

Search: Reveal is used to search for the questionnaire across studies. The search

`$path:*camd* $name:qs* ADAS-COG`

finds that only 13 study data sets have ADAS-COG. An additional search:

`$path:*camd* $name:qs* NOT (ADAS-COG)`

identifies the one questionnaire data set that does not include ADAS-COG.


Welcome: **Reveal Administrator** | Logout
Administration

Found 1 hit in 0.263 secs.

Folder	Name	Type	Size	Modified	Relative S ▼	Notes
camd/1055	qs.sas7bdat	Dataset	1,953 KB	03/01/11 3:14:57 PM		

Page 1 of 1
Page Size: 25
Displaying hits 1 - 1 of 1

Results: The data managers notify the CRO immediately to determine why they this study did not include ADAS-COG in the delivered data sets.

CONCLUSION

Data is growing rapidly in both size and complexity presenting a daily challenge for pharmaceutical companies. Innovative tools are necessary in order to locate the data you need for the task at hand. Reveal is a fast, scalable search solution that can help you with these challenges. Reveal searches terabytes of data in just a few seconds, reducing the time it takes to answer questions for regulators, reviewers, and your entire organization.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

John Leveille
d-Wise Technologies, Inc.
4020 Westchase Blvd, Suite 527
Raleigh, NC, USA
Work Phone: +1 919 600 6236
Email: jleveille@d-wise.com
Web: <http://www.d-wise.com>

Brand and product names are trademarks of their respective companies.