

: What are sources of data for building a data-driven DSS?

by Dan Power

Editor, DSSResources.com

Managers and IT staff who are thinking about opportunities for building innovative, data-driven decision support and predictive analytic systems should be examining this question. Too often the question of data sources is answered hurriedly and without reflection and the standard answer is to use data from a transactional system, a data warehouse, an accounting system, an operational data store or "big" data. Many internal and external data sources are possible sources for meeting a specific decision support need. In some cases, important internal source data is not being recorded or captured and external data sources may be expensive or require new data collection methods. Machine data and social media data are data of convenience like transaction data that may or may not be useful. Despite difficulties, data-driven decision support systems should be built using the most appropriate data for a specific decision task. Using convenient, readily available data is often a poor choice. Find the data that is useful for meeting a decision support need.

Building a data-driven DSS with easily accessible data usually seriously limits the usefulness of the system. That approach reminds me of the story of a man searching for a lost gold piece on a dark night. A passerby sees the man searching on his hands and knees by a street light and asks if he can help. The searcher says sure and tells of the lost gold piece. The passerby gets on his knees and starts looking and eventually gets frustrated. He asks "When and how did you lose this gold piece?" The searcher relates that he was walking down this sidewalk an hour ago and he dropped the coin accidentally and it rolled. Our patient passerby then asks "Why are we searching here?" Calmly the searcher explains, "There is a street light here."

Just because we have the data and just because it's in a data warehouse or Hadoop file system doesn't mean that by using that data we can find the answers for our decision support questions. So what sources should managers consider? If we think a source may be useful, how can we obtain and process the data?

In general, lists of categories can be checked off while investigating many topics and that is the case with data sources. So what are some potential sources of useful decision support data? Keep in mind all data sources have limitations.

1) Commercial data sources. Often for a fee, organizations can purchase data on current and potential customers, suppliers or products. This data may need to be merged and sorted with

: *What are sources of data for building a data-driven DSS?*

internal data to make it useful for decision support. Privacy, licensing and copyright issues need to be evaluated when considering commercial data sources. Much more data will become available from commercial sources in future years, the problem is that competitors can also obtain this data.

2) Customer/stakeholder surveys and questionnaires. Many firms will plan and conduct surveys using web forms, telephone or mail interview protocols. Randomly offering customers a chance to respond to an automated telephone survey and receive a reward of some type can be a quick, systematic way to gather customer satisfaction data. Similar approaches can be used with employees and other stakeholders.

3) Direct observation and data capture. Organization members and paid observers can regularly capture and record data on customer or competitor behaviors of interest to managers. A web form can record the observations and make the data available for decision support.

4) Government data sources. Local, state, federal, and international government agencies are major suppliers of data. Finding and organizing the data can be a major task. In the U.S., start with the FedWorld.gov web site.

5) Passive electronic data capture. Radio frequency identifiers, bar code readers and other data capture technologies can be implemented to gather innovative data. Privacy concerns may be an issue, but full disclosure and consent forms can help deal with such concerns. Affinity cards for customers or identification badges can be integrated into the data collection system.

6) Transaction data. The record keeping systems in organizations have extensive useful data. In some cases the systems need to capture additional data at the point of sale or when a transaction occurs to really provide good data for decision support. Also, legacy systems may have data quality problems that must be corrected.

7) Machine data and log files. Computers generate operating and log data. The Internet of Things is increasing the available machine data. Examples of machine data include: application logs, clickstream data, sensor data and Web access logs. Machine data is helping to create the "big data" phenomenon for decision support and analytics.

8) Social media data. Data includes mentions and hashtag counts and much more unstructured data. Companies like Gnip (<http://gnip.com/>) provides access to the full archive of historical Twitter

: What are sources of data for building a data-driven DSS?

data. Gnip provides access to "Foursquare, Tumblr, WordPress, and many others. We also offer managed access to the public APIs of Facebook, YouTube, Instagram, Google+, Flickr and others."

We have too much data and not enough data, that is the the good news and the bad news of data sources. We want quality data that is relevant to meeting our decision support needs. Please explore all of the potential sources of data when you identify a decision support need, don't get talked into searching under the street light.

References

B&E DataLinks, Business and Economics (B&E) Statistics Section of the American Statistical Association (ASA), at URL <http://www.econ-datalinks.org/> .

Climate Modeling and Diagnostics Group, Lamont-Doherty Earth Observatory of Columbia University at URL <http://rainbow.ideo.columbia.edu/> .

Foote, K. E. and M. Lynch, "Data Sources for GIS," The Geographer's Craft Project, Department of Geography, The University of Colorado at Boulder, at URL http://www.colorado.edu/geography/gcraft/notes/sources/sources_f.html .

Lechner, F., The Globalization Website, at URL <http://www.sociology.emory.edu/globalization/data.html> .

Power, D.J. "What is machine data?" Decision Support News, Vol. 14, No. 02, January 20, 2013 at URL <http://dssresources.com/newsletters/201.php>

Scorecard Data Sources, at URL <http://www.scorecard.org/> .

: What are sources of data for building a data-driven DSS?

Originally published Power, D.J. "What are sources of data for building a data-driven DSS?" DSS News, Vol. 8, No. 21, October 21, 2007 at URL <http://dssresources.com/newsletters/201.php>

Author: Daniel Power

Last update: 2014-06-20 04:29