LIBRARYJOURNAL

Thomson Reuters Data Citation Index

Thomson Reuters wokinfo.com/products_tools/multidisciplinary/dci/; to see a video of this product at work, visit: wokinfo.com/media/vid/dci-video.html

■ *By Bonnie Swoger*

CONTENT Over the last few years, researchers have increasingly been asked to share the data behind their discoveries. Sometimes these data are incorporated into journal articles as tables. Sometimes they are included as separate data files on a publisher's website as a part of the "supplementary material." Increasingly, the data are also being shared in institutional or disciplinary repositories, separate from the journal articles they supported. While this material is more accessible than ever, the research community faces several related challenges including how to discover available data and how to cite it.

Thomson Reuters' Data Citation Index seeks to answer both questions by providing a discovery platform for research data, suggested citation formats, and citation links between data sets and the journal articles that use (and reuse) them.

The Data Citation Index lives on the Web of Knowledge platform used by Thomson Reuters for other databases including the Web of Science databases Science Citation Index, Social Science Citation Index, and Arts and Humanities Citation Index. The Data Citation Index is integrated with the other products, allowing users to easily go back and forth between the scholarly literature and the cited material.

Data are selected for the Index through editorial review of the repositories in which they are housed. These include government sources (NCBI, NOAA), disciplinary repositories (PANGAEA, Archaeology Data Service), and institutional and interinstitutional repositories (Inter-University Consortium for Political and Social Research). All data from selected repositories are included.

Repository selection is based on several factors outlined in the Thomson Reuters white paper, "Repository Evaluation, Selection and Coverage Policies." Review centers on several factors, including the persistence and stability of the repository, peer review, links to the research literature, the age of the material, inclusion of funding statements, and the language of the data and metadata. Thomson invites suggestions for repositories to add to their offerings.

Records in the Data Citation Index fall into three main categories: repositories, data studies, and data sets. Repositories records include a description of the types of data included, contact information for the editor (either a person or an organization) and links to the repository on the web. The data studies records include both studies and experiments that are stored in the data repositories, typically linking to the data sets and offering a brief abstract, the types of associated data, and contact information for the study authors. Links to the Data Citation Index records for associated data are included, as well as direct links to external websites. Data set records include the titles (some of which are more descriptive than others), authors, a descriptive abstract, and a link to the data.

USABILITY Users familiar with the Web of Knowledge platform will easily be able to dive into the Data Citation Index.

The default search screen allows for three fields, with the option to add additional ones. The Web of Knowledge platform is designed for thoughtful, structured searches and does not offer a Google-like single search box. Users can limit their query by year, language, or document type (repository, data study, or data set). Standard searches by author, keyword, title, and subject descriptors are also available.

Results lists are sorted by date, but users can switch to sorting by relevance, times cited, source, or author. The Web of Knowledge platform offers powerful filtering tools after a search is completed, allowing users to exclude or limit the results based on any of the metadata fields.

At the time of this review, most data sets in the database had not been cited very often: the most cited items for a search for "gene" were the Pathogen Host Interactions Database and TreeBASE, both of which are repositories. The most-cited set in that search was cited nine times. As institutions and companies like Thomson Reuters continue to make data more discoverable and offer concrete suggestions for how to cite it, the number of citations to resources such as these will likely increase.

Users easily track citations by clicking the "Times Cited link" that appears in the results list as well as in the record for the data set, data study, or repository. This brings up a list of resources from multiple Web of Knowledge databases that cite the item of interest, including other data sets or articles.

Each data record includes several useful tools. For students and researchers new to citing this type of material, Thomson Reuters includes a suggestion based on the recommendations of the free data resource DataCite.org. Records can also be easily saved to Endnote or exported to other reference management software via BibTeX files. Users can set up citation alerts for any data set, data study, or repository and can suggest corrections to item records.

PRICING The Data Citation Index is a subscription product that follows a lease model, providing access to the entire data file. Factors such as the size of the subscribing institution and its current Web of Science holdings are considered in their subscription fee, which can range from 14%–28% of the institution's overall Web of Science contract.

VERDICT Most of the repositories included here can be searched for free on their own websites, and resources like DataCite.org offer free discovery services. Nonetheless, this product is a great addition to data discovery services and its integration with Web of Science puts it right where researchers are likely to be looking. Features such as citation tracking make this product highly anticipated by those who study the use and reuse of research data. While Data Citation Index is an expensive database and free data discovery services can suffice for cash-strapped institutions, this is a superior tool that should be considered by academic libraries that can afford it.

Bonnie J.M. Swoger is the Science and Technology Librarian at SUNY Geneseo's Milne Library and the author of the Undergraduate Science Librarian blog, undergraduatesciencelibrarian.org. Readers can contact her at swoger@geneseo.edu