

# 2023 Will Ring in a New Year of Tension Between Security Needs and (Post-Pandemic) Demand for Remote Work

The Covid pandemic may or may not be over, depending on who you ask. Regardless, the pandemic has shifted workplace expectations, possibly forever. To attract top talent in the new year, enterprises must enable remote work on an ongoing basis. The question is how to give remote employees the data they need to operate in a way that also satisfies the current world's heightened security concerns. Below are a few options to consider for 2023:

**Option 1.** The first option is to let remote employees carry around necessary data on laptops or other portable devices. Of course, this portable approach requires hard disk encryption and the like. But even with all that, do enterprises really want a copy of their crown data jewels going to the beach or the ski slopes? And is toting all of this around even practical? Critical information requires critical updates and sometimes on a minute-by-minute basis.

**Option 2.** The second option is a secure web-based repository for the data. The data itself can reside on-premises or in the cloud. Updating the data is a lot easier when everything is in an online repository versus replicated on various devices around the globe. For security, the enterprise can set up any level of log-in authentication prior to letting the remote worker connect to the online data.

Of course, whether the data follows the remote worker or the remote worker logs into the data, access alone to the data is not sufficient for productive work. Like an in-office employee, a remote employee needs precision text retrieval capabilities to instantly find necessary information and leverage that to get work done. To find something across the Internet, the answer is an Internet search application like Google. In contrast, concurrent instant search across terabytes of enterprise data requires a text retrieval program like dtSearch®.

**How a text retrieval program works.** A text retrieval program needs to first index the data before it can instantly search terabytes. An index holds each specific word and number in the data, along with the position of each in the data. Indexing

Article contributed  
by [dtSearch®](#)



**To attract top talent  
in the new year,  
enterprises must  
enable remote work  
on an ongoing basis.**

is easy: just point to the folders and the like to index and the text retrieval program can take it from there. A single index can hold a terabyte of text and there are no limits on the number of indexes that the text retrieval program can build and the end-user can cover in a single integrated search.

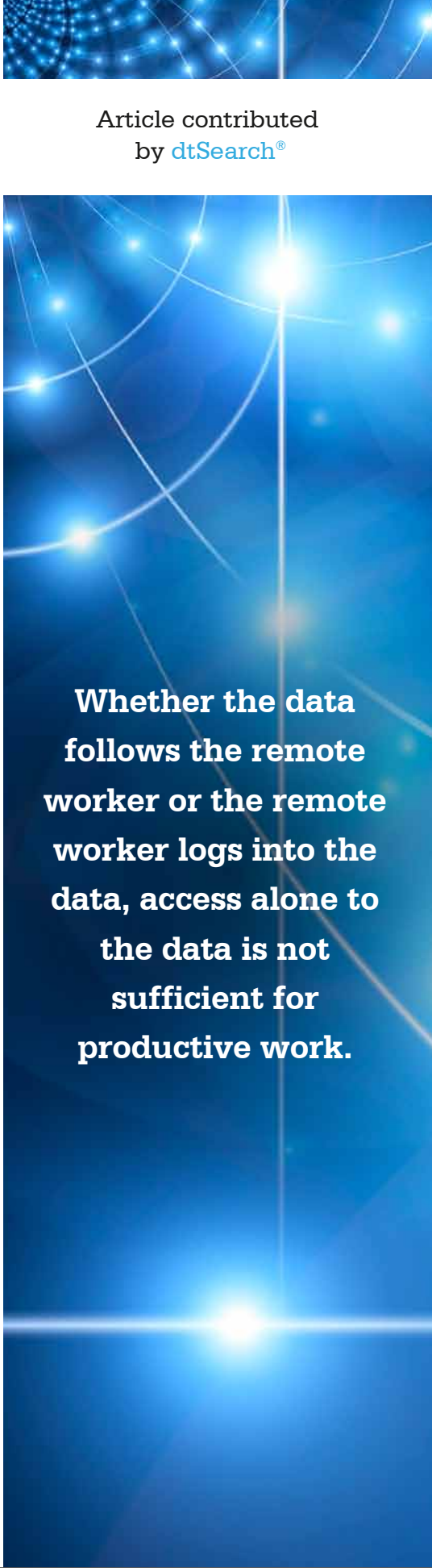
For indexing efficiency, a text retrieval program does not pull up each file in its associated application. Rather, the text retrieval program iterates over the binary version of each file. Parsing specifications for binary formats can be hundreds of pages long and differ vastly among different data formats. The first step for the text retrieval application is to figure out the correct parsing specification to apply.

Luckily, just from the binary format itself, the text retrieval program can figure out whether that file is a PDF, a web-based format, a Microsoft Office format such as Outlook/Exchange, Word, PowerPoint, Excel, Access, OneNote, etc. As a technical matter, the text retrieval program can figure out the applicable format without any regard to the file extension. That is a good thing, as it is all too easy to save a PowerPoint with a .PDF extension or a PDF with a .DOCX extension.

It is important to understand just how comprehensive a window into data the binary format provides to the text retrieval program and correspondingly to the end-user for search.

- ◆ Metadata that may be relatively hidden inside an associated application view of a file (like looking at a word processing document in Microsoft Word) is "front and center" in the binary format.
- ◆ Text that blends in with the background in an associated application view of the file, like white text against a white background or black text against a black background, is just like any other text inside a binary format.
- ◆ Binary formats enable the text retrieval program to identify potentially problematic files such as "image only" PDFs that require OCR prior to full-text search.
- ◆ Any tracked changes that remain in a document will appear in the binary format.
- ◆ Finally, the binary format offers ready access to multilevel nested file structures. If there is an email with a ZIP or RAR attachment, and inside of that is a PowerPoint with an embedded Word document, all that will be available to the text retrieval program.

Article contributed  
by [dtSearch®](#)



**Whether the data  
follows the remote  
worker or the remote  
worker logs into the  
data, access alone to  
the data is not  
sufficient for  
productive work.**



**Option 1 plus option 2.** As noted above, the text retrieval program can build and simultaneously search multiple indexes. Since the remote worker can check off multiple indexes to search, a single search can cover not only the shared online repository, but also any individual indexes the remote worker may create from that worker's own local files and emails.

To reflect data updates, the text retrieval program can update indexes automatically on a schedule. The update can pick up new files, file deletions and file modifications without the need to reindex everything else "from scratch." Crucially, updating indexes does not block out searching, so there is no data down time. Online search can proceed in a stateless manner, with no limit on the number of instant search threads, even while indexes update.

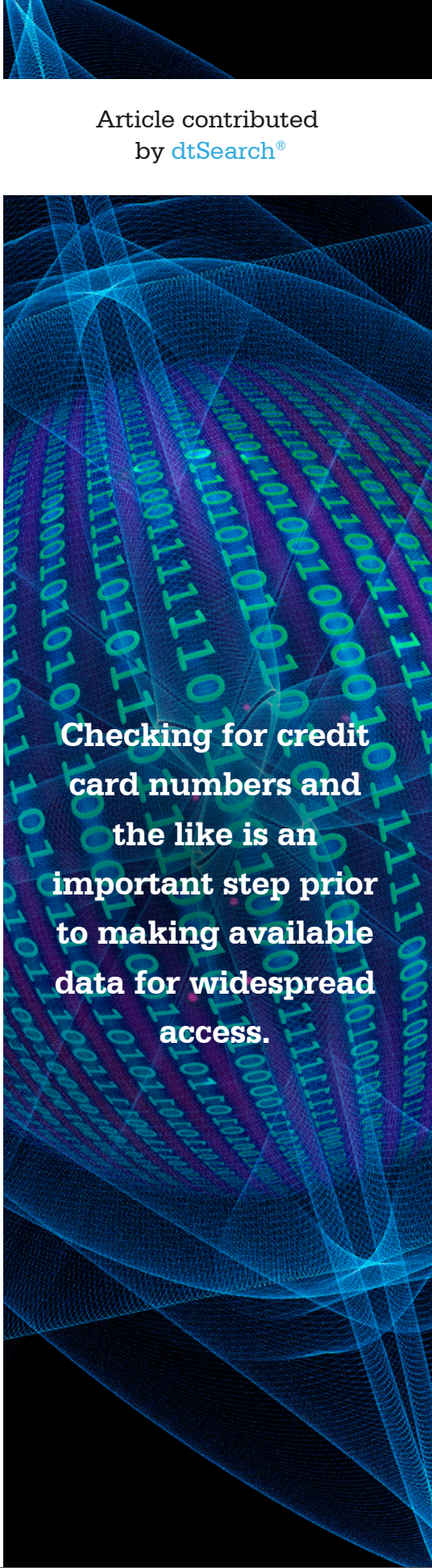
After a search, the remote (or in-office) worker can browse a complete copy of retrieved items with highlighted hits. In doing so, the text retrieval program will by default return to the original file to display that in full with added hit highlights. But what if the original file itself is slow to access or otherwise unavailable? This contingency is particularly of concern with shared web-based indexes that may reside separately from the original file repositories.

Caching, however, can store a full copy of files inside the index. The disadvantage of caching is that the index will be much bigger - about the size of the original data. But with caching, browsing search results with highlighted hits will remain "snappy" regardless of the status of the source data.

Note on Personal Identifiable Information (PII). Details on the over 25 full-text and metadata search options that a text retrieval program can offer are beyond the scope of this piece. But one option I want to mention is that the text retrieval program can search for personal information like identifying credit card numbers that may have found their way into data. Checking for credit card numbers and the like is an important step prior to making available data for widespread access.

In sum, for 2023, expect top talent to continue to insist on remote work. And the enterprise can answer: "no problem."

Article contributed  
by [dtSearch®](#)

An abstract graphic featuring a dark background with glowing blue and purple lines that resemble binary code or data streams. The lines are curved and layered, creating a sense of depth and movement. The overall aesthetic is high-tech and digital.

**Checking for credit  
card numbers and  
the like is an  
important step prior  
to making available  
data for widespread  
access.**