# Electronic Discovery:

## *Applicable to More Than Just E-mail*

Craig S. Mullins,
Corporate Technologist, *NEON Enterprise Software, Inc.*

white paper

NEON
Enterprise Software, Inc.

A recent survey of corporate counsel found that the typical U.S. company faces an average of 305 lawsuits and spends $12 million a year on litigation alone, not including settlements or judgments.[1] In such a litigious environment, the wise corporation will ensure that it is adequately prepared for the inevitable legal battles it will face. One aspect of this preparation must be the proper retention, care, and management of corporate data.

In this technological age, businesses have migrated to storing most corporate data purely electronically. As such, there has been a strong focus on electronic discovery (or e-discovery), led by the recent changes in the Federal Rules of Civil Procedure (FRCP). Electronic discovery refers to any process in which electronic data is sought, located, secured, and searched with the intent of using it as evidence in a civil or criminal legal case. Indeed, according to Gartner, Inc., electronic evidence is the predominant form of discovery today.[2]

Companies understand the importance of data and the need to preserve it, however many are not properly retaining and securing structured data for business and legal protection. Most organizations focus on enacting policies and procedures to retain their e-mail documents, because many of the highly publicized e-discovery cases have relied on information discovered in e-mail communications. As such, it is wise to tackle e-mail first, but it is critical not to stop there. Electronic evidence could encompass any type of data stored anywhere (ref. Section V of the FRCP).

An e-mail archiving and retention program is important, but is insufficient on its own to ensure your company is prepared for an e-discovery request. You should consider data located on the network, on various servers, hard drives, shared drives, laptops and PDAs, smart phones, and even on backup tapes.

With this in mind, consider for a moment what type of data is likely to be the most important to the on-going business of your company – the transactional data stored in database management systems. After all, this is the data that drives your company's business. It is this data that must be readily accessible in the event of an e-discovery request. Unfortunately, little focus is given to database data, at least in terms of preparation for e-discovery .The data in corporate databases is mission critical, it is literally used to run the business – from tracking insurance claims and managing brokerage trades, to handling deposits and withdrawals from bank accounts.

Indeed, lawsuits can hinge on the accuracy of this data. It is not sufficient to save statements and reports generated from transaction data, because the accuracy of the data on the statement can be questioned by the lawsuit. It is imperative that additional steps be taken to archive and retain transaction data for the same reasons we archive and retain e-mail documents.

Of course, the amount of transaction data to be managed can be quite voluminous. Maintaining data in operational databases over long periods of time creates problems— transactional performance problems as data volumes expand, and data authenticity problems for data that must be maintained for legal purposes. As such, data must be periodically archived from operational databases.

But what do we mean by the term "database archiving?" Database archiving is the process of removing selected data records from operational databases when it is likely they won't be referenced again, and placing them into a secure archive data store where they can be retrieved if needed.

As simple as this may sound, there are many significant challenges and requirements posed by database archiving. Perhaps the most important consideration is that archived data must be hardware and software independent. When data retention requirements span over decades, the production system from which the data was archived may no longer exist – at least not in the same form, and perhaps not at all. For example, data, or music on an eight track would be nearly unusable today. Many people would have issues even accessing data on a cassette tape. Or to put it in IT terms, consider the ever changing storage formats that come in and out of vogue. Could you read a punched card, which was the popular storage mechanism for large-scale computing in the 1960s and 1970s? What about something of more recent vintage, like 5 1/4 inch floppy disks?

[1] *According to a survey of 422 in-house lawyers, conducted by the New York law firm of Fulbright and Jaworski, LLP, as cited in CFO Magazine, October 13, 2006.*

[2] *Gartner, Inc. Research Note G00136366*

The archive also must be able to store a large amount of data. As we store more data, we will archive more data, but combining this with long mandated data retention periods provides some significant challenges.

One potentially devastating implication is that the archive for certain types of large volume applications may not fit into existing DBMS solutions. Every DBMS product has architectural limitations regarding the size of the databases it can manage. Database systems were not designed for multi-decade data archival, but to support daily business operations such as transactions and reporting. And though the limits change over time, organizations cannot rely on those limits changing in time to meet their requirements for long-term data retention.

The archive also must be able to manage data for very long time periods. Long-term data retention requirements are expanding. The SNIA Data Management Forum recently published its 100 Year Archive Requirements Survey which indicates that long-term generally means greater than 10 to 15 years, but more than 38% indicate that long-term means greater than 100 years. As a result of these extended data retention requirements , the archived data may outlive the systems and the programmers that generated them. Because data structures change over time, the archive must be able to support multiple variations of the data structure as it changes. Your operational database schema is not static -- database column sizes and definitions change, new data is added to the database, and so on. This means that the archive should be able to support multiple variations in the database schema over time.

And importantly, data must remain unchanged once it is archived. The archive must be able to protect against data modification. Only ‘‘read’’ access should be available to the archived data, and the data must be guaranteed to be authentic. Mechanisms to prevent surreptitious modification are necessary too.

Additionally, the archive requires metadata to be useful. The ability to understand the type of data being managed by the archive is vitally important because the types of data that could be archived are as varied as the number of individual businesses and applications in use. Furthermore, chain of evidence metadata is required to prove the authenticity of the data in the archive.

An archive must also support the ability to discard data exactly when it must be removed. For example, if the legal requirement is to retain the data for 35 years, then 35 years + 1 day is not acceptable! The data is there to support e-discovery requests made during lawsuits against your company. If there is no legal mandate to keep the data, then why go to the expense of keeping it? And if you do keep data past its legal life, it becomes discoverable and cannot be discarded if you are subsequently served with a lawsuit. So discarding from the archive is an important practice to implement appropriately and in accordance with regulations.

And finally, we create the archive for our structured data because we may someday need to access it for discovery. Culling data for e-discovery can be very challenging when trying to locate just a small subset of the applicable data among gigabytes, or even petabytes of information. The archive solution must allow for the review of each type of business transaction in context with its metadata.

Taking all of these considerations into account, a secure, durable archive data store must be used to retain data that is no longer needed for operational purposes, and it must enable query retrieval of the archived data in order to cull transactions for discovery.

## About the Author

Craig S. Mullins is a data management strategist for NEON Enterprise Software, Inc.. He has extensive experience in the field of database management having worked in various capacities with multiple database management systems. Craig is also the author of two books on data management. He can be contacted at craig.mullins@neonesoft.com or via his web site at http://www.craigsmullins.com

## About NEON Enterprise Software

NEON Enterprise Software is a leading technology provider of data management solutions that allow organizations to preserve, protect, retain and cost effectively manage their database data for its legal life. In an ever-changing world, our solutions allow you to manage your data with confidence and minimize business risk. Founded in 1995, NEON Enterprise Software is headquartered in Sugar Land, Texas, and serves customers worldwide with its dedicated team of industry experts.

For more information, visit www.neonesoft.com or call 281.491.6366 or 888.338.6366.