# Discover business databases

**Brent Babcock** and **Jeff Johnson** urge you to discover the potential benefits of conducting business database discovery

he problems of trying to locate specific information from collections of text documents are well known. The main discovery problem is that the information of interest can be conveyed in many different ways with a myriad of terms used by multiple individuals or groups of individuals. For example, the meaning of a date can only be properly understood from the context of the document containing that date (for example, in a patent infringement case, was the date mentioned in a document the date the design was created? Approved? Changed? Or was this the date the allegedly infringing product was sold to a customer?). Furthermore, mathematical operations, like totalling specific amounts or graphing results, are virtually impossible to perform when the data reside in unstructured text.

#### Control the chaos

Many of the difficulties associated with obtaining useful information from unstructured data are eliminated, or at least greatly reduced, by obtaining discovery of structured business databases. Unlike unstructured content, in which the authors are free to choose or create their own terms and format them as they see fit, structured database information is stored in carefully defined tables and columns. Only certain types of data elements are placed in specific fields, oftentimes with a controlled set of choices or vocabulary for each field.

Unstructured content can typically be characterised as including anything anywhere, whereas with well-designed databases, there is a place for everything and everything in its place. For example, in a structured database, the dates that a product was manufactured, shipped, and received will reside in three separate and readily identifiable locations. Comparable information in word processing files may appear almost anywhere and be stored in a variety of date formats.

Further, while there may be many different terms that employees use to refer to a particular product in e-mails or other documents, the so-called product ID in an inventory control database will be a unique and unambiguous term permitting precise and complete retrieval.

Many of an organisation's core commercial operations are initiated, tracked, and recorded almost exclusively in its databases, including purchasing, sales, billing, inventory, customer contacts, resource scheduling, claims call centres and human resources systems. While most databases are customised for individual organisations, the underlying software of these databases is usually widely recognised (for instance, SAP, SalesForce, Oracle, Sybase and DB2).

Not only can structured databases provide for more precise, targeted searching, but the search results can also be far more useful because they can be sorted, summarised, and presented in an unlimited number of ways. For example, even if thousands of sales were documented in thousands of customer order confirmation emails, it could be difficult to provide total sales by product, by region, or by sales agent, without re-inputting that data into a re-created database. Not so if queries could be run on the underlying database that generated and sent those e-mails.

#### Craft your case carefully

When developing a litigation strategy, careful consideration should be given to the nature of the information sought through discovery. Obtaining e-discovery of structured databases can greatly facilitate the review and analysis of large volumes of information, potentially saving counsel and expert witnesses significant time (and money). Indeed, as discussed above, relevant discoverable information may only reside in computer databases.

Be prepared, however, to encounter resistance from your adversary when seeking

### One-minute read



The traditional focus of ediscovery has been to manage unstructured data, such as electronic documents, e-mails, presentations, and spreadsheets.

Today, courts and attorneys increasingly recognise that business databases - the informational heart of many organisations - represent a valuable but often untapped discovery resource. In fact, business databases may be the only place where certain information may exist. Obtaining the most meaningful data from databases requires specialised knowledge and tools. However, early assessment of the data potentially contained in such databases can help focus discovery efforts and thereby reduce litigation costs. IP owners should be aware of the advantages of obtaining discovery of business databases compared to conducting traditional e-discovery of unstructured records.

### Case study: Key prior art found in archived database

A recent patent dispute between two electronics manufacturers reveals how effective database discovery can be, and further demonstrates the value of taking a step-at-a-time approach to ediscovery of business databases. In that dispute, an important contested issue was whether a certain product had been sold prior to a specific critical date. If such a sale had been made prior to that date, the product would become prior art to an asserted patent and thus could be used to challenge that patent's validity. Apparently, all paper records of the product's sales prior to the critical date no longer existed. However, the electronic database records for those sales had been retained in the company's old accounting system on backup magnetic tapes.

The archived sales data had been maintained in a proprietary Unix system that was no longer used. Further, the company that had originally published the software was no longer in business. The sales database was produced in the form of a voluminous raw system tape backup. The proprietary system backup did not permit access to any of the critical schema files, which define the database's tables, fields of data, and the relationships among them, nor was there other database documentation on the tape.

While it is possible to install, and configure a working database and to research the structure and content of the data contained in it based on empirical observation and investigation alone, that approach can be guite timeconsuming and far less efficient than would be necessary if the database documentation were available. We elected to analyse the database in stages to ascertain the likelihood that the potential database content justified expending further efforts to restore the database.

The data files themselves were contained in a legacy standard database format, and those files were not compressed or encrypted. Thus, the raw data was nominally readable as ASCII text files. However without any schema documentation or the ability to view the database in a few such records, and while we were uncertain whether the records were relevant by simply examining the raw data, the initial results looked sufficiently promising to continue our efforts to restore the database.

By further examination of the system files in the tape backup, we were able to determine the original publisher and version of the database management system. We contacted the company that had acquired the original publisher and were able to locate a programmer who was

1654654: 199615139520003410CFG01400-023 000002000000006540000FCC98-00170000088800000000022000000035000000000361000 4546647: 199615139530000310CWP01400-023 000000000000006540000FOC98-001700000888000000000000000003500000000036100 6546516: 199719142000078210AWPPOLES-10176 000118700p00000000600CRA-01 0000000060000000712r000000006000000000000 5456445; 199201011000000410CFG01400-023 000002000000003350000CRA92-0001000003350000000005200p0000036100000000335000 4654441: 199301011010000110CWP01400-023 000000000000003350000CRA92-000100000335000000000000000000003610000000033500 5655822: 199401011000004210CFG01420-240 00000000000000003950000UlCRA96-0001000003950000000000000000000425000000003950 

Example of unrestored database backup file viewed in Text Editor

the form of relational data, the task of finding pertinent records was daunting to say the least. The following figure is an example (modified) of how a raw database file appears in a text editor:

We decided to first examine the raw data files using a Java program that performed a raw byte scan of the data and searched for records containing the pertinent product model number and a date within the relevant sales date range. Our search identified

knowledgeable about the database software. With the assistance of this third-party programmer, we were able to restore the relevant portions of the database and continue our examination. Fortunately, with that assistance, we were able to find evidence of a sale of the product before the critical date. Having the date of the sale transaction also provided guidance on how to focus further discovery efforts within the separately produced unstructured data.

such electronically stored information, often considered the company's crown jewels. Many counsel - at the direction of their clients – will only produce so-called static paper printouts of existing database information, arguing that the printouts contain all of the relevant information and the receiving party can therefore reconstruct its own database if it so desires. Intervention from the court may thus be necessary.

In balancing the burdens and costs of production, courts often do require the production of computerised data, particularly where the requester can demonstrate the inadequacy of paper production. See Anti-Monopoly, Inc v Hasbro, Inc, SDNY 1995 - "The law is clear that data in computerized form is discoverable even if paper 'hard copies' of the information have been produced." - Dahl v Bain Capital Partners, LLC, D Ma 2009 - requiring production of spreadsheets in native format to preserve search capabilities and other attributes - and Covad Communications Co v Revonet, Inc., DDC 2009 - requiring reproduction in native file format of records previously produced in paper; responding parties not free to select form that is more difficult or burdensome for requesting party. Consequently, if obtaining your adversary's structured database through e-discovery is an important component of your litigation strategy, be prepared to carefully articulate why the production of simple paper printouts would be inadequate and unfair.

When possible, make narrowly focused, temporally bounded, specific database discovery requests. While responding parties invariably claim that broad requests (for example, the X database or all records relating to X) are overly broad and unduly burdensome, the court is able to discern a broad request from a rifle-shot request. Leverage the meet and confer, depositions, and a database expert to learn as much about the responding party's database design and structure, then make rifle-shot requests: specific tables, fields, records, actors (both human and organisational), dates, places, concepts & key words. Discovering this information will often open up sharp avenues for discovery.

### **Reciprocity required**

Also, be prepared to reciprocate with your company's own structured databases. When a litigant requests its adversary's structured database, a reciprocal request will almost certainly be forthcoming. While the parties' postures in the case are likely different, it could be awkward to explain to a court why the production of your opponent's electronic database is essential, while the withholding of your company's database is reasonable. In an effort to minimise potential disputes in this regard, consider including a discussion of structured database production in the parties' proposed discovery plan and case schedule, typically submitted to the court early in the case.

Finally, retaining an e-discovery expert early in the case is generally a good practice. Too often counsel wait until the discovery deadlines are fast approaching before realising that key information has not been produced. Obtaining a structured database requires considerable coordination and lead time, particularly if court intervention is necessary. Extracting, deci-

## Database discovery in Microsoft v i4i

The recent high profile decision in *Microsoft v* i4i (Supreme Court, June 2011) provides an example of when database-type discovery could be crucial. In that case, Microsoft sought to demonstrate that i4i's earlier software product (called S4) was prior art and anticipated i4i's asserted patent. However, when the development of S4 ceased many years before the litigation, i4i destroyed the source code and "no copies were available", according to case documents. As explained by the Federal Circuit: "In evaluating the evidence, the jury was free to disbelieve Microsoft's expert, who relied on the

S4 user manual, and credit i4i's expert, who opined that it was impossible to know whether the [patent] claim limitation was met without looking at S4's source code. Although the absence of the source code is not Microsoft's fault, the burden was still on Microsoft to show by clear and convincing evidence that S4 embodied all of the [patent] claim limitations."

Source code itself is not in a database form, but it is usually stored and maintained in a versioning system that acts as a database. While current databases may no longer contain relevant information about the prior art, most com-

panies routinely back up all electronic data on a regular basis. It is not uncommon for a party to simply state (almost reflexively) that the data no longer exists. But with detailed and particularised requests for information, effective crossexamination of key witnesses, collaboration between counsel and an e-discovery expert, and effective communication with the other party's IT person most knowledgeable, it may be possible to locate archived databases and recover case-altering evidence. Unfortunately for Microsoft, such discovery, if attempted, was not sufficient to locate the missing information.

phering, and recompiling the electronic data into a readily understandable format can also be time consuming. Further, the e-discovery expert must be educated sufficiently about the

case to understand what information is being sought and why. Accordingly, coordination with an e-discovery expert early in the case will help to ensure that critical discovery is not overlooked and obtained in a timely fashion.

### A more focused fight

Database discovery can be a powerful tool for quickly assessing the facts of a case and can help minimise or focus more traditional e-discovery of unstructured data. While discovery directed to business databases does not always receive as much attention as traditional e-discovery of unstructured data, databases can be equally valuable, and are sometimes of even greater value. An early focus on obtain-

ing database discovery, including discovery of the database schema needed for reasonable interpretation of the data contained therein, can yield large dividends.



**Brent** Babcock



Jeff

© Brent Babcock and Jeff Johnson 2011. Babcock is a partner with Knobbe Martens Olson & Bear in Irvine, California and Johnson a principal consultant at Advanced Discovery

Busting US trade mark litigation myths, June 2011

E-discovery rules released in Australia, March 2009

Plan early to avoid e-discovery risks, September 2007