

# Discovery of Databases in Litigation

Deborah H. Juhnke, Vice-President  
Computer Forensics Inc.™

As the production of electronic data in litigation has increased in volume so has its scope. Early forays into electronic discovery focused primarily on email messages and, to a lesser extent, electronic versions of word processing documents. Attorneys, however, have become increasingly aware of new data types that may lead to useful evidence, and they are regularly propounding broader electronic discovery requests.

Unfortunately, the technical understanding necessary to make informed requests has in many cases lagged behind the desire to do so. As a consequence, we are finding that requests for electronic discovery are often overly broad and unfocused, leading to confusion at best and to disputes at worst. Although the issues are many—ranging from how best to capture electronic data to who should bear the cost—the focus of this article is on databases specifically: what they are, how they should be requested, and how they should be produced.

## What is a Database?

A database is a collection of data arranged for easy computer retrieval, or, *a collection of non-redundant data that can be shared by different application systems*. The importance of this extended definition becomes clear as we examine real world examples in the context of real world litigation. It is tempting to rely merely on the first half of this definition since it appears to provide the most simple definition for discovery purposes. In fact, however, reliance on this simplistic definition creates havoc when applied to the complex databases found in corporate America.

## Defining a Database for Electronic Discovery Purposes

One way to look at databases is to define them according to their complexity and format. For example, we might look at these four types of databases, all of which fit the first definition, but not necessarily the second:

### A. Table or Spreadsheet

A table in a word processing program or a spreadsheet falls under the definition of a database. It consists of rows and columns of information (records) that may be used in sum or in part and is designed for easy computer retrieval.

## **B. Flat File Database**

Most attorneys and courts are familiar with this type of database. Two of the most commonly used litigation support databases are based on a flat file format, DB Textworks,<sup>®</sup> and Concordance<sup>®</sup>. These types of databases have a single layer of data in which every record is equal in form to every other record. Each will generally have the same number of fields, data types, and format. Boolean searches (and, or, not, etc.) yield a number of “hits” or records, and reporting is either built into the program or is easily defined by the end user.

## **C. Relational Database**

The next level of complexity is represented by relational databases. Multiple tables form a single database, and each table may or may not be “related” to another table. There is always, however, at least one common key to link every table to at least one other table.

A key distinction of the relational database is that it is built using a database development tool, such as Microsoft Access<sup>®</sup>, FoxPro<sup>®</sup>, or Paradox<sup>®</sup>. Depending on the tool used to build the database, its contents may reside in one or more files. Unlike flat file databases, relational databases must be designed and built from scratch, usually by more sophisticated programmers.

Additionally, and critical to understanding how they are used in electronic discovery, queries and reports must also be customized to each specific need. They do not exist unless they are programmed. This means that a new request for specific output from a relational database may require the development of forms for queries and reports.

## **D. Enterprise or Corporate Database**

The highest level of database complexity rests with what we will call enterprise databases, examples of which are built in Microsoft SQL Server<sup>®</sup>, Informix<sup>®</sup>, Oracle<sup>®</sup>, and similar programs. These are the databases that power industry and are used for managing extremely large volumes of data. They are most often integrated into business processes such as supply chain management, e-commerce, and transactional functions. Enterprise database platforms are also called upon as the underlying platform for customized or pre-packaged applications from third-party vendors.

Unlike their relational or flat-file counterparts, these types of databases are more amorphous, and harder to define or codify. This is because this type of database implies separation of physical storage from the use of data by an application program. The data, in other words, is separate from the program that uses it.

## **Playing by the Rules**

Unfortunately, Federal and State rules of court do not address, nor even contemplate, the many shadings and rapid evolution of meaning inherent in computer science. In fact, in 1999 the Advisory Committee on Civil Rules (Federal Rules of Civil Procedure) considered whether to modify the rules to deal specifically with the increasing use of electronic evidence. Their decision at that time was to rely on the rules as written, and to allow judges to address specific electronic discovery questions on a case-by-case basis. The decision was based in part on the theory that civil rules can't answer every question. They are only designed to give judges the tools they need to rule fairly.

From a practical standpoint, however, most judges are not well versed in the technologies that are the subjects of their rulings. Whether the issue is backup tapes, databases, or hard drives, it is only human nature to fall back on personal experience when no objective understanding exists. This reality was borne out recently when a judge, by her remarks, made it clear that her decision on the production of backup tapes was driven by her personal experience in a network and software upgrade within the court's own IT department. Her implied conclusion was that because the court's upgrade did not go well, nothing could possibly be as easy as counsel predicted.

There are movements afoot to modify the rules, but they aren't likely to occur anytime soon. In the meantime, the best approach appears to be systematic education of the Judiciary and the Bar.

## **Structure, Documents, and Databases**

The historical focus in litigation discovery has been on the *document*, most often described as "writings of every kind and description that are fixed in any form of physical media." The problem is that the common legal definition of a document is conceptually misleading in the context of electronic discovery issues. This is particularly true for collection and review of voice, video, and databases. When addressing these types of data, the average person's concept of a document—something that may be printed, read, and held in one's hand—begins to blur.

Although expanding the legal definition of a document to include electronic data does create the obligation to produce such data in discovery, it offers no guidance on how that production should be carried out. Consequently, there is significant variation in the

methods used to produce electronic data. If we assume that the *intent* of production is to provide meaningful information, we begin to see the ways in which this intent may be intentionally or inadvertently circumvented. With paper documents, or even word-processing files, meaning is fairly clear. There is a beginning and an end, and a logical structure. True documents tend to be self-contained. At worst, they refer to other documents in support of their content.

On the other hand, fitting databases into the conceptual framework of a document is particularly troublesome. Consider the following characteristics:

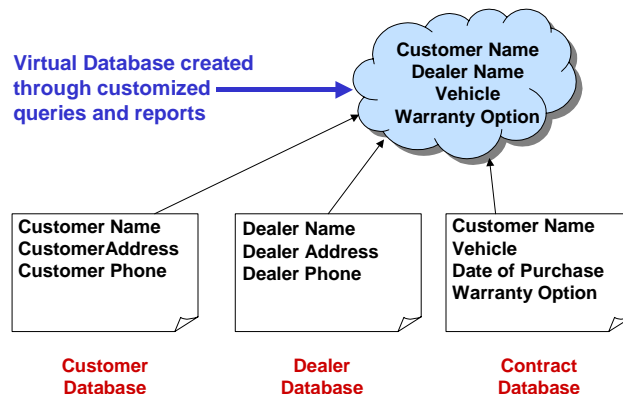
Document	Database
Self-contained	May or may not be self-contained
Meaningful on its face	Usually meaningful only in reports
Physically manageable	Almost always physically unmanageable
Easily printed	Not easily printed, sometimes impossible to print

Let's look at a real world example of a database comprised of millions of records—records that hold little meaning by themselves, except in the context of queries and reports.

### Real World Example

Following is a diagram of how a database we'll call "Customer Service" might be viewed. Note that the data contained in the Customer Service database (*customer name, dealer name, vehicle, warranty option*) exists only in other databases, and is pulled together into a virtual database through queries and reporting.

## Customer Service Database



Consequently, a request to produce or print the "Customer Service Database" would be difficult, if not impossible, because it does not really exist in one place in time, or in one location. It is only generated when a certain query is executed, and its content will change with time as the underlying tables or source databases change. Printing any one table or source database would also not give us the desired result.

## Searching – Databases and the Internet

Nearly everyone is familiar these days with searching the World Wide Web, more commonly referred to as the Internet. This is made possible by three things:

1. Powerful full-text **search engines** such as AltaVista, Google, Yahoo, and (a personal favorite) Dogpile.
2. **Crawlers** that constantly surf the net, capturing anywhere from a few key words to the entire text of a document and creating an index for the search engine.
3. **Metatags**, defined by the owner of a web site, that act like keywords (think Westlaw<sup>®</sup> keys). These are what cause some sites to pop up higher on your hit list than others.

Database searches and Internet searches, therefore, have vastly different characteristics. Compare:

Internet Search	Internet Example	Database Search	Database Example
Free form, full text	1992 World Series	Structured, fielded	"Date" > 12/31/1991 and "Date" < 01/01/1993 and "Event" = "World Series"
Index created automatically, constantly updated	Crawlers & metatags	Indexed by database administrator; updates periodic	Not all fields are index fields (so not all can be searched)
Searches across all or part of WWW, depending on what has been automatically indexed by the search engine being used	Results highly variable	Searches only pre-defined set of data. May be only one or selected tables	Results generally uniform
Search is easily entered by the user, with little or no restriction	Free form, full text, may even support "natural language" searching	User is restricted to queries of selected fields; may be different forms for different queries	Query forms are generally pre-defined by database administrator
Results may simply be printed or saved as a file	Internet search results are displayed as .html files	Results may only be printed through pre-defined, structured report	Reports are generally pre-defined by database administrator

As you can see, when searching the World Wide Web, a user may type almost anything in any order and get some type of result. The user's facility with the search engine will determine the success and accuracy of his search results. Searching databases, however, is subject to a variety of restrictions and relies heavily on the database administrator's ability to develop the queries and reports necessary to make the database useful. The benefit of all this structure is more accurate results.

## Searching Intranets

An intranet is a private network, generally within a company or organization, that uses web browsers to access data that is accessible only by the organization's employees or other authorized users. An intranet is not part of the World Wide Web. What does this mean for searching?

1. A corporate intranet will likely *never* offer full access to all types of data within a corporation. Intranets offer access to selected data sets. Therefore, an intranet does not offer "global" search capability.
2. A corporate intranet will generally rely on a full text search engine, not a structured search, so when data is available through the browser, results will be variable and unstructured.
3. Search results from an intranet search will only point to locations or files. These files must then be individually accessed for review and printing, since they are likely to be spread among many servers or locations.
4. Search results will also be variable because as with web content, an intranet environment is dynamic and may change at least daily if not more often.

Note: For reference, an *extranet* is merely an intranet that has been opened, in whole or in part, to users outside the corporation. An extranet also is typically not part of the World Wide Web and its only distinction from a search perspective is that it allows limited searching from the outside by a limited number of users.

## Printing and Production of Databases

It is rarely a simple matter to produce a database, particularly a *relational* or *enterprise* database. Issues include type of database, size, functionality, reporting formats, and so on. Nevertheless, because databases fall under the legal definition of a document they must somehow be produced. How to resolve this difficulty?

As with all discovery, the data sought must first be "relevant to the claim or defense of any party."<sup>1</sup> Though always important, adherence to this rule is *critical* to effective discovery of databases. What may on its face appear to be a reasonable request may, in

---

<sup>1</sup> Rule 26(b)(1), Fed.R.Civ.P.

fact, be a fishing expedition of the greatest magnitude or may be impossible to fulfill. We begin this discussion, then, with a guiding principle:

*Only ask for what you need, and only produce what is requested.*

Easier said than done? Yes, but here are some ideas on how to meet that goal:

**For the requesting party:**

- Seek to determine the nature of the databases you are requesting. Are they flat file, relational, or enterprise?
- What is their size? Remember, most corporate databases cannot fit on a single backup tape, much less a CD.
- What is their content? In many cases, understanding the basic structure of a database—its fields, query forms, and reports—can go a long way in helping you focus your requests.
- Understand that it will not be possible to “import” all databases into your own system for review. Some smaller databases and spreadsheets may be output to a delimited format suitable for import into MS-Excel or MS-Access. Enterprise databases, however, do not typically lend themselves to this treatment. Instead, request reports or other formatted output for review.
- In rare instances, a focused and guided on-site inspection of data structure may be appropriate if the producing party cannot provide written information regarding the database structure.

**For the producing party:**

- Ask that the production request be specific. If possible, provide information regarding field definitions, query forms, and reports to help refine the request.
- Production must be in a form usable to the requesting party, but will vary from database to database. In some cases a delimited file may be appropriate, while in others a report will be best.
- In those cases where no single report exists that will give a full view of the data, advise the requesting party whether it is possible to create such a report and what the cost to develop a custom report would be.
- Only allow on-site access to databases for purposes of defining queries and if no other alternative information regarding structure is available.

We have not considered the concept of “printing” databases, with the exception of reports, for good reason: either most databases cannot be printed in any meaningful way or they are so large that printing is a physical impossibility. Databases have this in common with large or complex spreadsheets. The data contained therein is only

understandable by its relationship to all other data within the file, whether by formula, calculation, or link. The only reasonable means of translating this to paper is through reports.

### **Some Closing Thoughts**

1. Databases, even if they are accessible through a corporate intranet, may rarely be searched in the same manner as an Internet search.
2. All databases are not created equal, in size, format, or function.
3. There is no “one best way” or standard format by which to produce databases.
4. The size and complexity of most relational or enterprise databases demand focused queries.

The use of electronic discovery in litigation will continue to evolve and, through trial, error, and education, standards for production may develop. However, for right now, the discovery of databases in litigation will remain more of an art than a science.