

BLAZING THE TRAIL: COURAGE | CHALLENGE | CHANGE

National Employment Lawyers Association

2014 Annual Convention

Boston, MA

June 25-28, 2014

REQUESTS FOR PRODUCTION OF DATABASES: DOCUMENTS v. DATA

Christine E. Webber

Cohen Milstein Sellers & Toll PLLC

1100 New York Ave., NW suite 500

Washington, DC 20005

(202) 408-4600

cwebber@cohenmilstein.com

Jeff Kerr

Mays & Kerr LLC

235 Peachtree Street NE

North Tower, Suite 202

Atlanta, GA 30303

404-446-1010

jeff@maysandkerr.com

I. DATABASE CONCEPTS

Employer databases contain rich stores of information; understanding how a particular database works and what it contains is the key to unlocking its value as a source of discoverable information.

Before considering common databases in employment litigation, we will provide some down-to-earth explanations of database concepts by examining a something all lawyers are familiar with: billing one's work. You may not have thought of your time records as constituting a database, but that is exactly what they are. Many of us track our time using practice management software, but database concepts apply equally even if you track your time on paper only. In database parlance, the template or pattern for a time entry is an "**object**," and it is generally comprised of several "**fields**": the name of the biller, the date the work is performed, the name of the matter, the amount of time, the rate, and a description of the work performed.

Each individual time entry that you create is called a “**record**,” which is a particular instance of the time entry object. The database itself is the collection of all existing time entry records for all cases, for all billers, as well as records for other objects (attorneys, cases, contacts, etc.). When you create an invoice, you are selecting a subset of time entry records from the database for easy presentation to the client or the court. The invoice itself is an example of a “**report**,” which refers to a snapshot of the database that is used for a particular purpose—here, billing a client. The criteria used to select the pertinent records for a report is called a “**query**.” By understanding the terms **object**, **record**, **field**, **report**, and **query**, you will be well on your way to knowing how to conduct database discovery.

II. COMMON DATABASES RELEVANT TO EMPLOYMENT LITIGATION

These are just some examples of commonly available, and commonly sought, databases. Obviously, this is case-specific information, and you should not let these suggestions limit your imagination.

A. Human Resources Databases

These will typically include, in addition to names, addresses, and other contact information, demographic information about each employee (i.e. date of birth, gender, race), perhaps some educational background, and typically their job history with the employer. Job history will generally show not only the dates during which the employee held different titles, but whether a move was lateral or a promotion. Such databases also generally include salary history, and may include performance evaluation scores. These databases may also be integrated with payroll and timekeeping databases (discussed below), and the information contained on the HR side could be virtually limitless, with entire performance files no longer “living” in file folders but instead maintained in an electronic HR database. For example, modern HR databases often allow supervisors and HR personnel to auto-generate performance reviews from within the

database and then maintain retain the completed review in digital form.

This data is essential for employment discrimination class actions, and can also be extremely useful in individual actions.

B. Job Posting/Bidding data

Job posting and bidding data are frequently stored in databases referred to as Applicant Tracking Systems (“ATS”). Such databases include job vacancies that have been posted (including the dates during which they were posted) and identify which employees applied to be considered for each vacancy. Some ATS services use algorithms to automate how resumes are processed and prioritized. As a result, the primary evidence of a discriminatory hiring scheme might be the algorithms themselves, which could, for example, score applicants higher or lower based on impermissible criteria like age or gender. Discovering the algorithm would then be the equivalent of noticing the letter “O” was only filled in in the interview notes for African-American candidates. This information, when available and complete, is an important addition to the basic HR databases when litigating promotion discrimination cases. Most electronic databases allow for lightning-speed analysis of large sets of data, which can be invaluable for gathering promotion or hiring statistics and comparing outcomes based on membership in protected groups.

C. Payroll Data

Payroll databases generally have in electronic form all of the information ultimately included on an employee’s pay stub, including hours worked, pay rates, etc. Such data provides essential information in easily manipulable form for calculating damages in wage and hour cases. It may also provide more precise information for pay discrimination cases, and essential information for claims related to access to overtime assignments. Many employers utilized third party vendors, such as ADP and Paychex, for payroll, and these databases contain rich sets of

information useful for employment claims.

D. Timekeeping Data (and audit trails re same)

Timekeeping data, which may be linked to time clock showing when an employee punched in and out, or to manually entered start and end times, provides far more detail than payroll data in showing which hours employees are credited with working out of each day, rather than just the total number of hours for the week. This information can be invaluable in wage and hour cases. It may show, for example, employees punching in (or logging in) at a time that is earlier than the time for which compensation begins or automatic deduction of lunch hours the employee never took. Audit trails in the same data may show where employee start and end times were edited, and who edited them. These databases can be a motherlode in wage and hour litigation and have proven to be “smoking gun” evidence in several of our own cases. In one of our wage and hour collective action cases, we requested complete time and pay database records for our clients during the statutory period. The records, which we requested in Excel format, contained the following fields: Employee ID, Last Name, First Name, Department, TimeZone, Module, Date of Change, Action, Reason, Changed By (User ID), Attribute, Old Value, New Value. We found numerous records where the Action field was equal to “Modified,” the Reason field was equal to “Timecard Edit,” and the Change By (User ID) was a manager. Obtaining this evidence changed our case dramatically

E. Access logs and other potentially probative databases

Almost any database maintained by an employer can be relevant to an employment claim. Here, we will consider some less obvious databases that you may want to investigate in litigating your claims. Say that you are litigating an individual case in which showing that the employer’s stated reason for termination or demotion was allegedly poor work performance, and your goal is to prove not only pretext, but also that the employer’s real intent was to retaliate. If your client

performed work in a database—for example, doing data entry—then the records of your client’s activities would be highly probative, and perhaps the database will have error-counting features that could be used to prove that your client made no more errors than her colleagues.

Alternatively, in a wage and hour case where hours worked are disputed, you may discover that the employer maintained a database showing when its alarm was turned off in the morning and activated in the evening. If your client was the person responsible for this task, then obtaining the database records from the employer or from the alarm company would be advisable. Similarly, access logs may be able to show which employees were logged into the computer system at which times (which may help to establish when they were working).

III. PRACTICAL TIPS FOR DISCOVERY

A. Locating databases

Discussion about available databases should be a high priority at the initial Rule 26(f) conference. If you sense that your opposing counsel does not intend to gather information on databases and other sources of ESI, send a letter stating that you expect to have a detailed conversation regarding defendants’ databases, including databases for HR, hiring, timekeeping, payroll, and whatever else could be relevant.¹ Unlike individual documents and email messages for which a search strategy is needed (outside the scope of this paper), employers should be aware of the databases that they are using, although they may need to be nudged with pointed questions about what systems they use to track whatever may be of interest to you (time, pay,

¹ In the Rule 26(f) conference, counsel must “discuss any issues about preserving discoverable information; and develop a proposed discovery plan.” Rule 26(f)(2). “When the parties do anticipate disclosure or discovery of electronically stored information, discussion at the outset may avoid later difficulties or ease their resolution.” Rule 26, advisory committee notes to 2006 amendment, subdivision(f). Among other topics, the discovery plan must state the parties’ proposals on “any issues about disclosure or discovery of electronically stored information, including the form or forms in which it should be produced[.]” Fed. R. Civ. P. 26(f)(3)(C).

employee information, vacancies, applications, etc.). Databases are generally *not* the type of documents that can be discovered with key word searches, concept searches or predictive coding. This is so because a database is not itself a document, but a method of organizing information and documents. Approach databases as a separate category from other forms of e-discovery.

B. Entire database vs. excerpts

One of the decisions you will have to make is whether to seek discovery of the entire database, or some excerpt from it. Generally, unless the database is very small (e.g., payroll for a company with 5 employees), we advise against requesting the entire database because the request may be deemed overbroad and the produced database may be less useful than a more targeted production. There are three major ways in which requests for production from databases are commonly limited, as well as an infinite number of variations. The three major areas of limitation to consider are: limiting the fields of data to be produced, limiting the records produced by individuals/groups of individuals, and limiting the records by temporal criteria.

1. Selecting Fields

Often HR databases will contain information that you do not need, for example, information about emergency contacts or the number of dependents covered by the employer's insurance. Or there might be demographic information (race, age, gender) that would be relevant in a discrimination case, that you would not request in an FLSA case. You can approach this from two directions simultaneously. First, identify all the categories of information that you hope that one or more databases includes and that you would like to have. For example: job history (positions held, date and reason for change in position); job performance; education/training/certifications; demographics (race/gender/age); name and contact information; applications for positions; compensation history. Second, ask for production of a data dictionary or equivalent information in which defendant identifies the specific fields included in its

database and what sort of information is stored in each field. The data dictionary will serve many purposes once you receive the data as well, but getting information on what fields of information the employer tracks is valuable, and may give you ideas about data you can request that you would not have had on your original wish list.

The data dictionary, in addition to other information you gather about the database itself, will help you to target your requests for production of key fields. Most databases contain fields whose value is determined by an algorithm or formula, and you will likely want both the *output value* of the field in the records you request *and* the formula by which the value is determined. For example, an employer may have a simple payroll database that it maintains in Excel format, or it may have a complicated proprietary database running on Salesforce.com. In either case, the payroll database will likely have a “gross pay” field whose value is determined by a formula, for example, straight time hours worked * rate of pay + overtime hours worked * rate of pay * 1.5. In an overtime case, you will want both the amount of gross pay in certain weeks, and the formula. In fact, the formula may be your key to establishing liability in a case based on systematic miscalculation of overtime premiums.

2. Limiting Individuals Included

For example, you might limit your request to data on individuals who held specified positions, or who worked in a particular geographic area or organizational unit. *See, e.g., Johnson v. Kraft Foods N. Am., Inc.*, 238 F.R.D. 648, 657 (D. Kan. 2006) (imposing such limitations when ordering discovery). Otherwise, the employer’s production of records might bury the needle for which you are searching in a haystack of data. Gathering *useful* database discovery often will depend on your preliminary discovery of database objects, fields, records, tools for performing queries, and reporting capabilities. Rather than receiving a data-dump, you can target your production in a manner that packages the data tailored to the needs of your case.

3. Limiting Time Period Covered

While it is common for discovery to be limited to the time period encompassed by the claim, with perhaps some time immediately preceding the claim, beware of how time limits are implemented. For example, while you may be content to have personnel data regarding employees who worked for an employer in a particular position from 2010 to the present, for any such persons, you will likely need information about those individuals' experience with the employer prior to 2010. In a promotion case, for example, there may be a big difference between an employee hired on December 2, 2009 and one hired on March 3, 2005 when it comes to competing for a promotion in May 2010. Thus, be careful in how date limitations are framed.

Another consideration here is the multiplicity of time and date fields in most electronic databases. Reverting back to our example at the beginning of the paper on lawyer time entries, we noted that a time entry record always has a "date" field. But in electronic databases, there will usually be several other date fields that contain useful information: for example, the created date and the last modified date. These fields may be thought of as database metadata, but they can be highly relevant evidence in their own right. For example, imagine the relevance of created and last modified dates in litigation over attorney billing. If the created date for every record is equal to the day the bill is prepared, the fact-finder might wonder if the time entries were simply "made up" at the last minute. Alternatively, last modified dates before sending a bill might indicate an attorney exercising reasonable billing judgment by cutting time for redundant tasks. In both cases, the metadata fields are probative evidence, and should not be ignored.

C. Discovery of information needed to understand the database

Crucial to your ability to work with the database is obtaining discovery of how the database is organized, what types of data are stored in which fields, what codes are used to record the data, whether a field is set up specifically to record dates in a particular format, etc,

etc. A good description of databases and the information you will need to obtain about them is included in *N.A.A.C.P. v. Acusport Corp.*, 210 F.R.D. 268, 278-282 (E.D.N.Y. 2002). In particular, there is often a “data dictionary” which summarizes all of this information, and if available, is a key document to request. *N.A.A.C.P.*, 210 F.R.D. at 280-82.

You are entitled to this discovery:

1. *Fautek v. Montgomery Ward & Co., Inc.*, 96 F.R.D. 141, 144-45 (N.D. Ill. 1982) (granting sanctions for, among other conduct, failing to provide complete information responsive to plaintiffs’ requests for codes necessary to understand the HR database)
2. *In re Seroquel Products Liab. Litig.*, 244 F.R.D. 650, 660-61 (M.D. Fla. 2007) (finding sanctionable party’s failure to produce IT employees for informal discovery of how to understand various databases, as agreed to in the parties case management order)
3. *Zamora v. D'Arrigo Bros. Co. of Cal.*, C04-00047JW(HRL), 2007 WL 806518, at *3 (N.D. Cal. Mar. 15, 2007) (granting motion to compel where 30(b)(6) deponent did not have complete information about the database codes that were the subject of the deposition)

The question then arises how best to obtain information on database objects, fields, processes, and architecture. Certainly, requests for production are valuable for obtaining existing documents setting forth or explaining the structure of the database in general and the definitions of fields, objects, and processes. A 30(b)(6) deposition can also be extremely useful to obtain information on employer databases, but both of these formal discovery devices have limitations in the context of database discovery. In our experience, the best “tool” for obtaining the information you need to frame your requests for production of information from databases is informal, cooperative ongoing conversations with opposing counsel, IT personnel, and, as needed, experts. If your opposing counsel does not want to cooperate with you, then we recommend sending a letter with case cites and other authority on the need for cooperation.² If

² The Third Sedona Principle states: “Parties should confer early in discovery regarding

this fails, then a conference with the judge or a motion to compel may be in order.

D. Form of Production

A database is the quintessential example of a document that should be requested and received in some native form – if not its original form (which might be proprietary programming that would be awkward to deal with), as .csv³ or other file format that can be imported into database programs so that you can analyze the data. Still, the meaning of “native” production in the database context may be hard to pin down in some cases. For example, the opposing party may maintain a complicated proprietary database that stores information for day-to-day use in file types that you will never be able to use without literally sitting at the employer’s computer terminal after receiving hours (or days) of training in managing the database’s user interface. The key in such cases is to learn as much as you can about the database so that you can put it to work for your case. Prioritize understanding the objects, the relationships between types of objects, and the key fields, and then you can request data in a form that will be both useful and useable. As a rule of thumb, .csv or Excel production is preferable to most alternatives when the native database files are not usable. Assuming you have Excel on your computer, you will be able to open and review .csv files in Excel, and then save them in Excel format. But also note that some data maintained in databases does not translate well into .csv or Excel. For example, an HR

the . . . production of electronically stored information when these matters are at issue in the litigation and seek to agree on the scope of each party’s rights and responsibilities.” *See also* Fed. R. Civ. P. 26(f) advisory committee note (2006) (“Rule 26(f) is amended to direct the parties to discuss discovery of electronically stored information during their discovery-planning conference.”); *In re Seroquel Products Liab. Litig.*, 244 F.R.D. 650, 658 (M.D. Fla. 2007) (finding that a party’s “refusal to allow contact between individuals with appropriate technical backgrounds as part of the effort to resolve technical issues is an inexplicable departure from the requirements of Rule 26, the Sedona Principles and this Court’s expressed expectations.”); *id.* at 664 (finding that Rules 26 and 34 require “dialogue to discuss the search terms”).

³ .csv files are commonly seen when dealing with databases because they are the standard form in which data is *exported* from a database.

database often will link PDF, Powerpoint, Word, or other useable native documents to the records within the database. Ensure that these files are produced in their native format.

On the other hand, beware opposing counsel's attempts to produce database records or whole databases as mere images of the data – whether in .tiff or .pdf or hard copy. The fact that a .tiff or .pdf may be “electronic” does not make it usable – it is a static visual presentation, depriving you of the opportunity to analyze the data directly. Courts have recognized this for years. In addition to Fed. R. Civ. P. 34, which permits the requesting party to designate the form of production, there are numerous court rulings requiring parties to produce electronic databases, even where the same information has already been made available in paper form.

1. *Adams v. Dan River Mills, Inc.*, 54 F.R.D. 220, 222 (W.D. Va. 1972) (“Because of the accuracy and inexpensiveness of producing the requested documents in the case at bar, this court sees no reason why the defendant should not be required to produce the computer cards or tapes and the W-2 print-outs to the plaintiffs.”) Thus, as early as 1972, when producing “electronic” versions of databases might mean producing punch cards encoding the data, courts had figured out that was more useful than mere paper printouts.
2. *Ayers v. SGS Control Servs.*, 03 CIV. 9078 RMBRLE, 2006 WL 1519609, at *2 (S.D.N.Y. Apr. 3, 2006) (requiring defendant to produce payroll data in electronic form even though already produced in paper form, finding otherwise “plaintiffs will have to recreate the pay and time data in an electronic, manipulable form. Although defendants already have the information in electronic form-because their payroll system is electronic-they insist on having plaintiffs go through the burdensome, time consuming, and expensive process of recreating the data in electronic form. “)
3. *Margel v. E.G.L. Gem Lab Ltd.*, 04 CIV. 1514 PAC HBP, 2008 WL 2224288, *5 (S.D.N.Y. May 29, 2008) (compelling production of electronic data, even though report had been printed and produced containing the information). The court relied upon Fed. R. Civ. P. 34(a)(1)(A) noting that the electronic version of the data was a non-identical copy. The court also noted that the report that the defendant had offered in lieu of the underlying data could not even be admitted into evidence unless the underlying data were produced to plaintiffs, citing Fed.R.Evid. 1006 (governing compilations or summaries of records).

IV. CASE AUTHORITY ON COMMON DISPUTES

A. Rule 34 Basics

Databases are subject to the same rules of discovery as any other document. In addition to *Margel, supra*, some good cases laying out the basics are:

Johnson v. Kraft Foods N. Am., Inc., 238 F.R.D. 648, 657 (D. Kan. 2006) (granting motion to compel “the contents of electronic databases that contain personnel-related data (e.g., salary, race, gender, promotions, etc.) concerning the Kraft Sales Organization in the Kansas City Region” after resolving disputes over the geographic, organizational and temporal scope, and rejecting claims that requests for items such as “electronic databases” and “data dictionaries” were vague and ambiguous)

Chen-Oster v. Goldman, Sachs & Co., 285 F.R.D. 294, 301-02 (S.D.N.Y. 2012) is particularly helpful at parsing the standard of “accessibility” (as courts, and now the rules, distinguish between requests for accessible and inaccessible data) and finding that the concept of accessibility refers only to technical impediments to accessing the data, not to the cost or related burdens, although it separately considered the costs in analyzing proportionality issues (discussed below at C).

B. Responding to claims that you are asking defendant to create a document that doesn't exist

If you request a report be run from a database, or a subset of the data extracted, you might run into the objection that you are asking the defendant to create a new document, one which doesn't exist, and that Rule 34 does not require them to create new documents. Rule 34 applies to all documents, including electronically stored information which may not remotely resemble a physical document. Indeed, Rule 34 specifically authorizes requests for “documents or electronically stored information . . . stored in any medium from which information can be

obtained either directly or, if necessary, after translation by the responding party into a reasonably usable form.” Fed. R. Civ. P. 34(a)(1)(A). Seeking a report from a database is essentially a request that the data be “translat[ed] by the responding party into a reasonably usable form.” *Id.*

This is not a novel construction of Rule 34. As far back as the 1970 Advisory Committee Notes, it was established that “when the data can as a practical matter be made usable by the discovering party only through respondent's devices, respondent may be required to use his devices to translate the data into usable form. In many instances, this means that respondent will have to supply a print-out of computer data.” Fed. R. Civ. P. 34, Adv. Comm. Notes (1970).

Courts have applied the clear direction of Rule 34 to similar circumstances, and found it appropriate to require the responding party to produce responsive documents by generating requested computer reports. For example, in *Evans v. Tilton*, 2010 U.S. Dist. LEXIS 36953 (E.D. Cal. Mar. 18, 2010), plaintiffs requested that defendant search its database system and report on which medical personnel had approved a specifically identified request. Defendant took the position that no such report currently existed, and thus it was not required to produce the requested information. The court rejected that argument, holding that the generation of such a report simply constituted production of electronically stored information in a useful, requested form, as provided for under Rule 34, and required defendant to produce such information. *Id.* at *12-15; *see also Flying J Inc. v. TA Operating Corp.*, 2007 U.S. Dist. LEXIS 55567, *13-14 (D. Utah July 30, 2007) (granting motion to compel defendant to “extract, download, copy, or print” electronically stored information, as there was no evidence the data was inaccessible); *Jinks-Umstead v. England*, 227 F.R.D. 143, 148 (D.D.C. 2005) (requiring defendant, who had already produced all existing hard copies of reports that were generated from its database, to formulate

queries to extract relevant data from its database, and produce to plaintiff).

A similar conclusion has been reached regarding requiring defendant to respond to an interrogatory by analyzing and summarizing data from its database, rather than respond to a document request. In *Powerhouse Marks, L.L.C. v. Chi Hsin Impex, Inc.*, CIV.A.04CV73923DT, 2006 WL 83477 (E.D. Mich. Jan. 12, 2006), the court held that:

The Court is convinced that given the nature of the raw data and the fact that it is much more easily used in conjunction with a financial database, Defendant's burden in deriving the information sought in Plaintiffs' interrogatories is significantly less than Plaintiffs'. The Court is also convinced that Defendant is better positioned to accurately interpret and explain how the documents produced are responsive to Plaintiffs' Interrogatories 1 and 2. Certainly Defendant has the capability to compute and provide in summary fashion annual sales figures and expenditures for specific products. Under the Federal Rules, a party may be ordered to produce such information even when the electronic information does not exist in the format requested. See generally Fed.R.Civ.P. 34, Advisory Committee's notes specifying that a "respondent may be required to use his devices to translate the data into usable form." Furthermore:

Although there may be some differences between requiring the production of existing tapes and requiring a party to so program the computer as to produce data in computer-readable as opposed to printout form, we find it to be a distinction without a difference, at least in the circumstances of this case.

In re Air Crash Disaster at Detroit Metropolitan Airport on August 16, 1987, 130 F.R.D. 634, 636 (E.D.Mich.1989)(citing *National Union Electric Corp. v. Matsushita Electric Industrial Co., Ltd.*, 494 F.Supp. 1257, 1262-63 (E.D.Pa.1980)). This Court similarly concludes that Defendant must produce to Plaintiffs a more usable form of data that is responsive to Plaintiffs' Interrogatories 1 and 2.

C. Burdensomeness and Proportionality

The court in *Chen-Oster* addressed the proportionality argument with a very sophisticated understanding of databases. *Chen-Oster v. Goldman, Sachs & Co.*, 285 F.R.D. 294 (S.D.N.Y. 2012) In that putative class case, the defendant claimed that extracting the requested information from its HR database would require 90 to 150 hours of staff time, plus another 40 to 80 for

quality control checks.⁴ Defendant argued this was too great a burden to ask them to bear in proportion to the needs of the case.

The court first considered two different approaches for reducing the cost to defendant: (a) extracting a sample from the database, which both parties opposed for different reasons, and which the court conceded would not reduce the time required to program the query to extract the data, though it might decrease the time spent on quality control reviews (*id.* at 304); (b) a data dump – simply copying the database, instead of running a query to extract only certain information, which plaintiffs proposed but the court rejected because whatever savings in time was made on the front end would be lost in the end requiring defendant to provide all of the information needed to work with the data (*id.* at 305).

Having concluded there was not a significantly cheaper, equally effective alternative, the court turned to weighing the need for the data and found the burden imposed on defendant was proportionate to the needs of the case. Specifically, the court held “[t]here is little doubt that the needs of this case justify the discovery sought by the plaintiffs. The information in the databases is central to the plaintiffs' claims of gender discrimination in compensation, promotion, and evaluation. The amount in controversy, while not specifically quantified, is surely substantial.” The court also found the defendant had ample resources to provide the discovery. *Id.* at 305. Moreover, the court recognized that “the importance of this litigation is not measured in dollars alone; the plaintiffs seek to vindicate the civil rights of the class members, and thus further an important public interest.” *Id.* at 306.

Finally, the court understood that Goldman Sachs had somewhat inflated the burden by building in so much quality control, noting that their estimate “which is rather conclusory,

⁴ Other databases were also discussed, each with varying numbers of hours estimated.

appears to be based on a goal of providing a pristine set of data.” *Id.* The court pointed to the Sedona Conference to support its conclusion that “the standard for the production of ESI is not perfection. Rather, ‘[a] responding party must use reasonable measures to validate ESI collected from database systems to ensure completeness and accuracy of the data acquisition.’ The Sedona Conference, *The Sedona Conference Database Principles: Addressing the Preservation and Production of Databases and Database Information in Civil Litigation*, March 2011 Public Comment Version, at 32 (emphasis added).” The court held that defendant could sample the data extracted to identify any systematic errors rather than conduct the comprehensive quality review defendant proposed. *Id.*

As *Chen-Oster* demonstrates, it is important to press for detail underlying defendant’s estimates, so that the “padding” in the estimate can be identified. It is also useful to find out how often defendant engages in similar searches and extraction of data for its own internal purposes, to show how routine such work is. Finally, on the benefit side, it is important to ensure that the court considers the non-economic value of civil rights litigation, and does not merely consider the lost wages or other damages sought.

The other end of the scale, an example of a case finding a lack of proportionality with a relatively cursory discussion is *E.E.O.C. v. Supervalu, Inc.*, 09 CV 5637, 2010 WL 5071196, at *8 (N.D. Ill. Dec. 7, 2010). In that case, the EEOC sought data showing employee histories that they hoped to use to identify instances when there were vacancies that could have been filled, a request defendant estimated would take a week of staff time to comply with. However, the court found that “as plaintiff appears to acknowledge by virtue of its ‘piecing together’ argument, the information it seeks would not definitively prove the existence or number of open or filled positions at defendants’ retail stores at any particular time. Instead, that information would

require significant analysis as well as the inference that each time one type of employee position ended—whether by termination, resignation, or otherwise—a corresponding position became open and ready for hire.” Given the lack of evidence supporting that inference, and contrary evidence from defendant, the court refused to order defendant to produce the requested data. The court also noted that discovery closed in one week, and found that the EEOC's explanation for why it was seeking the database had shifted over the course of the litigation, making it less inclined to grant the request.

D. Direct Access to Defendant's Computer System

In a case defendants are fond of citing – and mis-citing – the Eleventh Circuit held that permitting an opposing party direct access to the producing party's computer system should generally not be permitted. *In re Ford Motor Co.*, 345 F.3d 1315, 1316-17 (11th Cir. 2003). Of course, this does not mean, as defendants sometimes suggest, that the requesting party is not entitled to discovery of the computer system, just that they should not ordinarily be permitted to roam at will through the entirety of the opposing party's data. *Id.* The court found that “Rule 34(a) does not grant unrestricted, direct access to a respondent's database compilations. Instead, Rule 34(a) allows a requesting party to inspect and to copy the product—whether it be a document, disk, or other device—resulting from the respondent's translation of the data into a reasonably usable form.” *Id.*

The Eleventh Circuit did acknowledge that in cases of “improper conduct on the part of the responding party” the requesting party might be entitled to check the data produced against the original files. *Id.* Other courts have found misconduct does permit such access. *U & I Corp. v. Advanced Med. Design, Inc.*, 251 F.R.D. 667, 674 (M.D. Fla. 2008) (“to gain direct access to the respondent's databases, the court must make a factual finding of some non-compliance with discovery rules and protect respondent with respect to preservation of his records, confidentiality

of non-discoverable matters and costs.” The court found a failure to produce responsive emails and other documents, despite multiple motions to compel, and granted the requesting party the opportunity to inspect various computers of the producing party under certain conditions).

Moving further afield, at least one court has permitted forensic imaging of a computer system in order to establish the merits of a claim that defendant has misappropriated client data and shared it with third parties. The court explained that “there is simply no other way in which to seek this information, and it should establish once and for all what Revonet did with the Covad data that was in the Federated Database. Indeed, the investigation may have a determinative impact on whether this case will survive” and thus that the benefit of the discovery outweighed the burdens. *Covad Commc'ns Co. v. Revonet, Inc.*, 258 F.R.D. 5, 12 (D.D.C. 2009).

In another useful decision from the same court, a requesting party was held to be entitled to discovery of the specific query used by the producing party to extract data from their database. *Barnes v. D.C.*, 289 F.R.D. 1, 21-22 (D.D.C. 2012) (explaining that the query was “electronically stored information” that counts as a “writing” under Federal Rule 34(a)(1)(A), and was a “document needed to interpret the data in the database,” one of plaintiffs’ requests.

Access to the responding party’s system may also be provided when the responding party objects to the burdensomeness of preparing the reports sought by the requesting party. The court in *L.A. County Emps. Ret. Ass'n (LACERA) v. Towers, Perrin, Forster & Crosby, Inc.*, 2003 U.S. Dist. LEXIS 10233, at *4-6 (C.D. Cal. Jan. 28, 2003), ordered the responding party to provide the moving party access to its software to run certain analyses and produce output.

E. Metadata

Metadata is particularly important in database and spreadsheet documents.

At one end of the spectrum is a word processing application where the metadata is usually not critical to understanding the substance of the document. The information can be conveyed without the need for the metadata. At the other end

of the spectrum is a database application where the database is a completely undifferentiated mass of tables of data. The metadata is the key to showing the relationships between the data; without such metadata, the tables of data would have little meaning. A spreadsheet application lies somewhere in the middle. While metadata is not as crucial to understanding a spreadsheet as it is to a database application, a spreadsheet's metadata may be necessary to understand the spreadsheet because the cells containing formulas, which arguably are metadata themselves, often display a value rather than the formula itself. To understand the spreadsheet, the user must be able to ascertain the formula within the cell.

Williams v. Sprint/United Mgmt. Co., 230 F.R.D. 640, 647 (D. Kan. 2005). The court then discussed authorities governing the production of metadata, finding little relevant case authority and turning to the Sedona Principles.

At the time the *Williams* case was decided, the Sedona Principles did not have strong support for production of metadata. The court quoted Principle 12: “[u]nless it is material to resolving the dispute, there is no obligation to preserve and produce metadata absent agreement of the parties or order of the court,” but also one of the comments, “Of course, if the producing party knows or should reasonably know that particular metadata is relevant to the dispute, it should be produced.” *Williams*, 230 F.R.D. at 650, 651. Ultimately, despite the iffy guidance from the Sedona Principles, the court came to a helpful conclusion supporting production of metadata:

the Court holds that when a party is ordered to produce electronic documents as they are maintained in the ordinary course of business,⁶⁸ the producing party should produce the electronic documents with their metadata intact, unless that party timely objects to production of metadata, the parties agree that the metadata should not be produced, or the producing party requests a protective order.

Williams, 230 F.R.D. at 652.

Since then, and following amendments to the Federal Rules, the Sedona Principles have also evolved to support the same conclusion. The current version of Principle 12 states:

Absent party agreement or court order specifying the form or forms of production, production should be made in the form or forms in which the information is ordinarily maintained or in a reasonably usable form, taking into account the need

to produce reasonably accessible metadata that will enable the receiving party to have the same ability to access, search, and display the information as the producing party where appropriate or necessary in light of the nature of the information and the needs of the case.

Thus, there should be no dispute that the metadata needed to make use of databases and spreadsheets should be produced when databases and spreadsheets are requested in discovery.

F. Disputes about which fields Defendant will produce

In addition to potential differences about which fields in the database contain relevant information (see above at ___), which tend to be very case-specific disputes, there are some recurring issues, not specifically limited to databases, but highly relevant to them. Those are disputes regarding inclusion of employee names, social security number, addresses or other contact information.

Name: Attaching names to job histories is vitally important. If documents produced by defendant include, for example, discussions between managers about whether to promote Susan Smith or Robert Jones, accessing objective data about those individuals and comparing it to the managers' discussion could be very illuminating. Further, Plaintiffs seek anecdotal evidence from class members. If a class member avers that she was paid less than Georgie Porgie, despite superior qualifications, being able to search the personnel database for Mr. Porgie's records permits Plaintiffs to submit, under seal, corroboration of key facts, without sharing such information with the class member. Finally, knowing the names of potential class members means knowing the names of potential witnesses whom counsel may seek to interview to provide anecdotal information.

Social Security Number: SSN is the best means of obtaining updated addresses for individuals whom Plaintiffs may wish to contact as potential witnesses. Thus, SSN may be as important as names and contact information in permitting collection of anecdotal evidence.

Addresses: Plaintiffs often want to contact putative class members to obtain information about their personal experiences.

There is ample authority supporting disclosure of this type of information in employment class actions. “The disclosure of names, addresses, and telephone numbers is a common practice in the class action context.” *Artis v. Deere & Co.*, 276 F.R.D. 348, 352 (N.D. Cal. 2011) (collecting cases), motion for relief from judgment denied (Aug. 8, 2011).⁵ Plaintiffs are entitled to contact information to show that they meet the requirements of Rule 23. *Id.* As in *Artis*,

Here, the putative class members may possess relevant discoverable information concerning issues dealing with Plaintiff’s gender discrimination claims, as well as other class certification issues. Further, the privacy interests at stake in the names, addresses, and phone numbers must be distinguished from those more intimate privacy interests such as compelled disclosure of medical records and personal histories. *Id.* While the putative class members have a legally protected interest in the privacy of their contact information and a reasonable expectation of privacy the information sought by Plaintiff is not particularly sensitive.

Id. at 353. Plaintiffs need for the discovery outweighs Defendant’s privacy objection.

A protective order can protect the legitimate privacy interests of employees whose information is thus disclosed.

⁵ See also *Khalilpour v. CELLCO Partnership*, 2010 WL 1267749, at *3 (N.D. Cal. Apr. 1, 2010) (“the disclosure of names, addresses, and telephone numbers is common practice in the class action context because it does not involve revelation of personal secrets, intimate activities, or similar private information, which have been found to be serious invasions of privacy”); *Soto v. City of Concord*, 162 F.R.D. 603, 617 (N.D. Cal. 1995) (compelling disclosure of personnel files in case alleging police misconduct); *Hill v. Eddie Bauer*, 242 F.R.D. 556, 563 (C.D. Cal. 2007) (ordering production payroll records including names, addresses, social security numbers in employment class action, so long as personal identifiers were redacted when documents were filed with the court); *Currie–White v. Blockbuster, Inc.*, 2010 WL 1526314, at *2 (N.D. Cal. Apr. 15, 2010); *Babbitt v. Albertson’s Inc.*, 1992 WL 605652, at *6 (N.D. Cal. Nov. 30, 1992) (at pre-certification stage of Title VII class action, defendant employer ordered to disclose names, addresses, telephone numbers and social security numbers of current and past employees); *Putnam v. Eli Lilly & Co.*, 508 F.Supp.2d 812, 814 (C.D. Cal. 2007) (ordering production of the names, addresses, and telephone numbers of putative class members, subject to a protective order, including those who worked in a sales division other than the plaintiff’s own).

G. Authority Supporting Discoverability of Data Regarding Other Employees

While not directly addressing electronic database issues, many courts have addressed non-ESI related questions of whether a plaintiff may obtain discovery about how other employees have been treated – establishing the basic relevance of such information is a prerequisite for being able to seek it in the form of an electronic database, so here are some helpful citations. The first several cases specifically address this issue in the context of individual claims, and find that data about broad groups of employees should be discoverable, not merely data about the individual plaintiff and his or her facility. The last several cases provide support for discovery of such data prior to a ruling on class certification.

1. *Diaz v. American Telephone & Telegraph*, 752 F.2d 1356, 1363-64 (9th Cir. 1985) (individual employee entitled to discovery for entire region, not just his facility) The Ninth Circuit specifically considered and rejected defendants' assertion that discovery should be limited to the districts where plaintiffs are employed because decisions about promotions from supervisor to manager are made at the district level, stating:

It is irrelevant whether the same Operations Manager made all of the hiring and promotion decisions about which Diaz seeks information. One way of reaching conclusions about an employer's motives is by ascertaining whether the employer's explicit or implicit policies encourage or permit discriminatory employment decisions by its supervisory personnel. The employer is responsible for such decisions, *see Miller v. Bank of America*, 600 F.2d 211, 213 (9th Cir. 1979), because its policies control the manner in which its supervisory employees make them. We have previously recognized that the promotion patterns of the employer, as a whole, are relevant to such an analysis of motive. *Diaz*, 752 F.2d at 1363-64.

2. *Guruwaya v. Montgomery Ward, Inc.*, 119 F.R.D. 36 (N.D. Cal. 1988) (individual employee entitled to discovery for entire region, not just his facility)
3. *Burns v. Thiokol Chem. Corp.*, 483 F.2d 300, 305-06 (5th Cir. 1973) (holding that overall statistical evidence was discoverable even in an individual case because data reflecting overall pattern of conduct is probative of whether employer has discriminated against particular individuals)

4. *Rich v. Martin Marietta Corp.*, 522 F.2d 333, 343 (10th Cir. 1975)
5. *Hollander v. American Cyanamid Co.*, 895 F.2d 80, 84-85 (2d Cir. 1990) (“Evidence relating to company-wide practices may reveal patterns of discrimination against a group of employees,” that was relevant to this individual case).
6. *Flanagan v. Travelers Ins. Co.*, 111 F.R.D. 42 (W.D.N.Y. 1986). The court allowed discovery in this individual case for the entire region, rather than limiting discovery to the office where the plaintiff worked. The court distinguished cases which limited discovery, stressing that in this case the personnel decisions were not made at a local level, and noting that the defendant failed to show that production of the documents would cause an undue burden.
7. *Henderson v. National R.R. Passenger Corp.*, 113 F.R.D. 502, 507 (N.D. Ill. 1986) (“The plaintiff needs access to statistical data in order to prove that racial discrimination exists.”) In this case, the court allowed nationwide discovery of all information and documents relating to the investigation of alleged employee violations of certain rules of conduct. The court reasoned that the plaintiff needed access to statistical data in order to prove that racial discrimination existed. Thus, the court allowed nationwide discovery.
8. *Holley v. Pansophic Sys. Inc.*, No. 90 C 7505, 1993 U.S. Dist. LEXIS 13910, *16-17 (N.D. Ill. Sept. 30, 1993) In this multi-plaintiff case, the court held that allowing discovery of similarly-situated employees “sets a minimum level of discovery rather than an optimal formula.” and permitted discovery of positions not held by plaintiffs, as well as nationwide discovery of all persons with supervisory or managerial positions.
9. *Lyoch v. Anheuser-Busch Cos.*, 164 F.R.D. 62, 66 (E.D. Mo. 1995) (company-wide statistical information is discoverable to demonstrate pattern and practice of discrimination, even in an individual disparate treatment case).
10. *Duke v. University of Tex. at El Paso*, 729 F.2d 994, 996-97 (5th Cir. 1984). The Fifth Circuit held that it was an abuse of discretion to deny plaintiff discovery of employment patterns for entire the university, despite fact that the decision about the plaintiff was made at the department level. The information was ruled relevant to plaintiff’s motion for class certification as well as to her individual claim.
11. *Hubbard v. Rubbermaid, Inc.*, 78 F.R.D. 631, 639 (D. Md. 1978) (granting motion to compel discovery from divisions other than the one plaintiff worked in, and noting “the weight of authority is that discovery in a Title

VII action is not limited to plaintiff's department or division of employment, but may be company-wide in scope," even though the class certified was limited to the sales operations in which named plaintiff was employed).

12. *Owens v. Bethlehem Mines Corp*, 108 F.R.D. 207, 213-14 (S.D.W.Va. 1985) (court certified class and granted motion to compel, also noting that company-wide statistical evidence would have been discoverable even in an individual suit)
13. *Riggs v. United Parcel Service*, 24 F.E.P. Cases 93, 94 (E.D. Mo. 1980) The plaintiff, who was seeking to represent a class of women who may have faced discrimination in the Missouri District, was entitled to discovery of information for the entire District. The court ruled that the scope of discovery in a putative class action was defined by the scope of the potential class for which certification is being sought.
14. *Canty v. Philip Morris*, 18 F.E.P. Cases 86 (E.D. Pa. 1978) (certifying class for two regions, directing disclosure of nationwide statistical information, and noting plaintiff could seek to broaden the class certified based on the nationwide statistics — allowed nationwide discovery of the defendant's statistical information regarding certain personnel decisions. The court allowed nationwide discovery despite limiting the plaintiff's class to two regions whose personnel activities were centered in Philadelphia.);
15. *Karan v. Nabisco, Inc.*, 78 F.R.D. 388 (W.D. Pa. 1978). The court, in granting class certification, allowed nationwide discovery. The court, in addressing concerns regarding the plaintiffs' ability to represent such a large class, stated that “[w]hen the Court has required an early class determination on an incomplete record and such questions arise, the preferable course is to permit discovery to continue as to the national class rather than prematurely to exclude the class.”
16. *Barhart v. Safeway Stores, Inc.*, No. S-92-0803WBS JFM, 60 Fair Empl. Prac. Cas. (BNA) 751, 25 Fed. R. Serv. 3d 35 (E.D. Cal Dec. 14, 1992).The named plaintiffs in a putative (pre-certification) class action for employment discrimination were full-time, female food clerks. They moved to compel discovery of documents related to non-food department employees and part-time employees of defendant Safeway's Northern California retail stores. The magistrate judge presiding over discovery granted the motion because the plaintiffs might thereby obtain information relevant to the court's determination of whether to certify a class containing non-food and/or part-time employees.