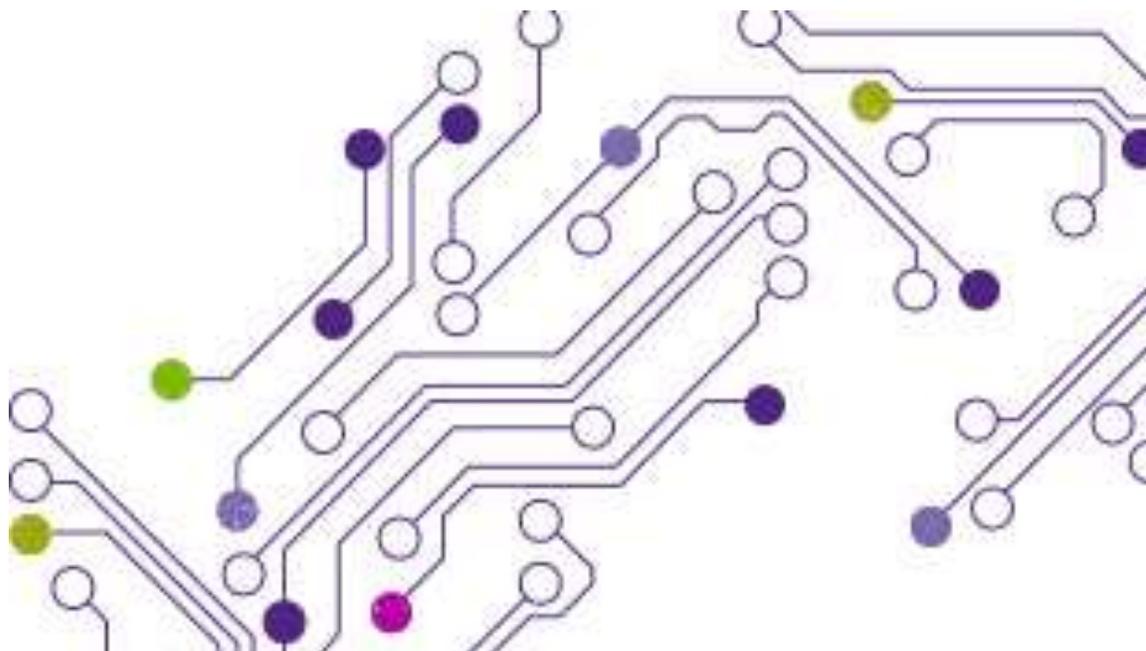




Comprehensive Defensible  
Litigation Solutions

# Enterprise Systems Discovery Management

A Granite Legal Systems White Paper



March 2016

Granite Legal Systems  
1201 Louisiana St., Suite 350  
Houston, TX 77002  
(713) 652-0881  
[www.granitelegal.com](http://www.granitelegal.com)

**Contents**

**INTRODUCTION** .....3

**OVERVIEW** .....3

**DISCOVERY PITFALLS** .....3

**EXEMPLAR PROJECTS** .....4

**IDENTIFY SYSTEMS** .....6

**IDENTIFY SYSTEM OWNERS** .....6

**IDENTIFY SYSTEM DOCUMENTATION** .....7

**SYSTEM ASSESSMENT** .....8

**SYSTEM HISTORY** .....9

**COLLECTION** ..... 10

**DATA ANALYSIS** ..... 10

**REVIEW AND PRODUCTION** ..... 11

**CONCLUSION** ..... 12

**ABOUT GRANITE LEGAL SYSTEMS** ..... 13

## **INTRODUCTION**

*The “smoking gun” may be in an e-mail message,  
however the smoldering embers may be buried in a legacy  
IBM DB2 database repository.*

E-mail, word processing documents, spreadsheets and other unstructured data are the typical focus in the discovery of electronically stored information (ESI). Highly publicized, landmark cases ensure that no one forgets to examine back up tapes and archives as well. Yet one often-overlooked source of ESI presents unique challenges to the litigation team: enterprise database systems.

Not only is the information stored in enterprise systems frequently relevant and discoverable, that information usually represents high value data that is key to the core litigation issues. The discovery process for enterprise database systems, however, involves significant challenges that require expert technical assistance to avoid errors, reduce costs and assure defensibility. This whitepaper outlines a disciplined, methodical and defensible approach for the efficient and accurate identification, collection, analysis, review and production of enterprise data.

## **OVERVIEW**

Organizations use enterprise systems to capture, store and transform data for core business functions such as finance, regulatory compliance, manufacturing, sales and human resource functions. Distinct from common electronic files such as Microsoft Office documents or e-mail repositories that individuals choose and determine how to organize in a personalized – unstructured - manner, data contained within an enterprise system will share a common organization regulated through the specific system interface. This standardized organizational data form – structured data – has different discovery planning, collection and processing requirements from the familiar loose electronic files and messaging data that the legal industry has successfully managed in the past.

The foundation of many enterprise systems are relational database management systems such as Oracle, SAS, SQL, IBM DB2, SAP and Lotus Notes/Domino. An end user performs a series of steps to enter or retrieve information. The output may be a screen display of information, a decision tree or outcome such as a document, report or export of raw data used to manage the business. Examples may include manufacturing history tracking, claim or complaint management systems and inventory tracking software. In general, an organizational need to manage large volume transactions or business process steps will likely result in an enterprise system implementation.

## **DISCOVERY PITFALLS**

Enterprise systems do not fit within the established discovery processes for emails, traditional loose files and other such “unstructured data”. The identification, collection, analysis, review and production of the ESI from enterprise systems is complex due to numerous factors, including but not limited to:

- **Capacity:** By definition, these systems store vast amounts of information. Enterprise system transactional data sets for a mid-sized company will commonly contain hundreds of millions of records or more stored as terabytes of data. (This is not “big data.”)
- **Diversity:** Each system is highly customized for the unique needs of each organization. Even systems with the same marketing name (i.e., SAP, PeopleSoft, JDE) are rarely structured the same way, requiring customer and system specific discovery solutions.
- **Complexity:** Responsive data identification requires an understanding of the system’s internal structure, relationships and connections to other systems that may span organizational business units. Incomplete data identification can hinder a complete organizational data view.
- **Functionality:** Corporations design systems to meet business needs, not with litigation in mind. As a result, they usually have no “easy button” for the transformation of data into a format suitable for review and production in litigation.
- **Sensitivity:** Clients have a legal or business obligation to protect private and sensitive information, including employee data, customer information, financial records, health records and credit card numbers contained within the corporate systems.
- **Usability:** Discovery team must be prepared to convert data received from enterprise systems into a format suitable for attorney review, analysis and production.

For these reasons, discovery teams should consider engaging experienced assistance for translating litigation requirements into technical specifications for identification, collection, analysis, review and production of ESI from enterprise systems.

### **EXEMPLAR PROJECTS**

Some examples of projects involving the identification, collection, review, analysis and production of data from enterprise systems include:

#### **Clinical Trial Data**

Dispute: Drug and Medical Device

Data source: SAS

Project Summary:

Collect raw SAS data files, identify and extract relevant records by product and subject, anonymize competitor manufacturer information, programmatically redact personally identifying information and produce data in delimited text format suitable for multiple platforms.

**Supplier Information**

Dispute: Drug and Medical Device/Product Liability

Data source: Oracle

Project Summary:

Develop script(s) with internal team to obtain data export, regenerate the dataset locally and perform iterative searches to reduce overall volume of potentially relevant data and create easily digestible reports for attorney review.

**Warranty Claims**

Dispute: Product Liability

Data source: SAP

Project Summary:

Obtain legacy archive data, recreate the database, rebuild the database schema, create data dictionary and perform filtering and culling. Develop reports for attorney review and raw data spreadsheets for expert analysis.

**Email Archive Analysis**

Dispute: Various

Data source: email

Project Summary:

Collect email archive database from third party vendor, recreate in SQL, perform date, name and search term filtering, reconstruct emails and associated attachments and provide to hosting vendor for attorney review.

**Medical Device Monitoring**

Dispute: Drug and Medical Device/Product Liability

Data source: IBM DB2

Project Summary:

Collect device-monitoring data from proprietary hospital system(s), perform programmatic redactions of personally identifying information, mask sensitive corporate information and produce the data in delimited text format sufficient for analysis in diverse platforms.

**Complaint Handling**

Dispute: Product Liability

Data source: Lotus Notes/Domino

Project Summary:

Collect complaint-handling databases from U.S. and foreign sources, extract both data and attachments, perform iterative name and product searches and create PDF reports suitable for TIFF production with programmatic redactions of HIPAA protected information.

### **IDENTIFY SYSTEMS**

The first task is to identify the potentially relevant systems vital in supporting litigation. There are multiple questionnaires and checklists available to interview business staff regarding the location of ESI. For enterprise systems, this information can frequently be obtained from resources across many functional areas in the organization such as IT, engineering, quality assurance, accounting, compliance, customer service, etc.

In absence of written answers to interrogatories designed to identify potential sources of relevant ESI, eDiscovery personnel might start with a review of a corporation's organization chart to identify key management personnel. This personnel can provide guidance and direction to discovery teams tasked with gathering information about key systems of interest in the discovery process.

Organizations that are frequent targets of litigation or regulatory investigations will likely have some version of a systems inventory or other documentation depicting major repositories of documents or data. The discovery team should request the existing documentation and supplement with current updates or proposed system changes along with any periodic audits or inventories performed by IT personnel or individual department managers.

Understanding the regulatory environments and constraints an organization faces provides valuable guideposts to identifying relevant systems. For instance, medical device manufacturers are broadly bound by the Code of Federal Regulations Title 21, Part 820<sup>1</sup>, which addresses the quality management systems required to meet regulatory requirements established by the U.S. Food and Drug Administration. While this knowledge may not identify specific systems within such an organization that are subject to discovery requests, it should inform a qualified investigating party to focus their inquiries on systems that address required regulatory reporting.

### **IDENTIFY SYSTEM OWNERS**

Once the discovery team identifies potentially relevant systems, the next step is to identify the business and technical system owners who can provide system background information and documentation. There may be multiple system and/or business owners for a particular system, depending on the scale, scope, system architecture, geographic considerations and functional components of a system.

Technical Owners can provide information about the hardware, operating system, database and applications. The technical team will typically have specific information on underlying system technologies, disaster recovery planning, data retention and archival, security implementations and data

---

<sup>1</sup> <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfCFR/CFRSearch.cfm?CFRPart=820>

structures documentation. These resources will also have access to tools or technical resources to assist with data extraction and reporting requests.

Business Owners will have a deep understanding about how the business information flows into and out of the system, how it is used and the general business purpose. Additionally, they will likely be fluent in the system data nuances – issues such as when and how is information entered, how particular entries are worded and specific changes in the business processes and requirements over time.

The roles played and knowledge residing with these individuals may vary from organization to organization and from system to system. Larger systems may have dozens of business and technical owners associated with it.

System and business owners typically have limited time to respond to inquiries and requests for assistance from investigating parties. Respect that these individuals have real jobs beyond responding to discovery requests and try to minimize the interruptions to them and their staff. Designating key contact points and communication protocols can go a long way towards minimizing confusion, redundant requests and unnecessary business interruptions throughout a lengthy discovery process.

Technical owners may not be employed by the organization facing the discovery requests. Many organizations outsource the development, maintenance, support and overall management of their larger, more complex enterprise systems. Such contract support resources may not always be available or consistent throughout the litigation lifecycle and/or may need to receive permissions (and budget) from the client organization before completing any requested work.

### **IDENTIFY SYSTEM DOCUMENTATION**

Identification and collection of system documentation is key to the investigation process. Not only will this documentation facilitate data extraction, analysis, review and production, but it will also establish a due diligence audit trail for the discovery process. This information will help address questions or challenges posed by opposing counsel during meet-and-confer sessions or following production of data extracted from a specific system.

Below are documentation examples the discovery team should request during the investigation process:

- System Diagrams (on premise and externally hosted)
- Database Entity Relationship Diagrams
- Data archive and retention policies and practices documentation
- Business continuity plans
- System Manuals (user and administrative)
- Training manuals
- Vendor contract materials (3<sup>rd</sup> party support, development, maintenance, hosting, etc.)

While the documents collected during this project phase may seem overly technical and marginally relevant to the core litigation issues, the discovery team will find these materials highly valuable to support data retention and preservation requirement compliance, collection planning and discovery burden and proportionality arguments. In addition, these efforts are valuable across multiple litigations. Indeed, the discovery team's comprehensive efforts should form the basis for a general corporate "Discovery Encyclopedia" for legal teams to reference during litigation planning phases.

### **SYSTEM ASSESSMENT**

The discovery team should perform a detailed assessment when data residing in an enterprise system is identified as potentially relevant to a discovery effort. The purpose of this assessment is to gather information required to design and implement a process for extracting data from the identified system for subsequent collection, review and production processes, such as:

- Data and document types managed by system
- Locations where system resides (servers, cloud, third party locations, etc.)
- Details of system users (population size, user types, etc.)
- Key IT support resources
- System upgrade schedule
- Impact of critical regulatory/compliance schedules on assessment/discovery (e.g. year-end financial reporting, quarterly SEC filings, etc.)
- Data size (data storage, tables, rows, users, licenses, servers)
- Archival, retention, disaster recovery processes, procedures, schedules
- Report generation functionality, data warehouse inputs/outputs
- Quality Assurance/Development environment description
- Data export, extraction, data transfer capabilities

Data size and type (data, propriety file formats, encrypted formats, document revision sets) by the system may dictate the approach required to extract relevant data. Large data extracts from multiple system tables may require custom programming available only through third party vendors. On the other hand, ad hoc document retrieval may be accomplished from the front-end user interface. The legal team should balance the assessment detail with the common limitations associated with resource availability, time and budget constraints and the likely data relevance to the discovery effort. Minimally, discovery teams should gather enough information to satisfy the specificity requirements for burden and proportionality arguments.

Physical system location may create data privacy or repatriation issues. For instance, data in a system with servers in both the United States and Europe may require additional resources to facilitate specialized treatment of potentially sensitive data sources.

## SYSTEM HISTORY

Time is a critical component to system investigations. Discovery teams must consider legacy – predecessor – systems and successor systems when identifying potentially relevant data sets. Discovery teams should request and review any documentation regarding predecessor/successor systems and legacy system archives to gain a better understanding of how technical teams implemented data migration design and execution and to prepare for possible subsequent inquiries such as:

- Availability of legacy system archives for subsequent analysis and extraction
- Migration audit documentation availability to verify project completeness and accuracy
- Business or technical decisions made during migration planning or execution affecting content or functionality migrated from legacy to successor system

Common issues legal teams encounter in migration assessments include:

- Multiple field level data values consolidated into singular field values
- Data type conversion resulting in corrupted or altered data representations
- Incomplete data migrations where selected data is not moved to the new system

In a best-case scenario, migration project documentation would answer most, if not all, technical questions regarding inconsistencies between old and new data systems. The worst-case scenario would require the consulting team to match like data elements and seek client clarification regarding data set inconsistencies.

New systems are frequently implemented on a date cut-off basis. For instance, a legacy customer complaint management system may be managed in parallel with its replacement counterpart until the disposition of a complaint managed in the legacy system is finalized. In these instances, it may be necessary to prepare for the extraction of data from two complaint management systems.

Conversely, a business decision may have been made to perform minimal data mapping of legacy data to its new repository. Legacy data may reside in the new system but may require special processes for identification and extraction as it may not correspond to the same data structure as material created “day-forward” in the new system.

While all data from a legacy system may have been successfully migrated to a new system, the “look and feel” or contextual significance of those data elements may appear different once migrated to a new environment. The data differences may result from technical specifications and/or limitations for the successor system or from changing business requirements that alter the required data to be stored, processed and reported within the new system.

## **COLLECTION**

Once the technology assessment is complete, the most efficient and accurate method of acquiring the data must be determined. The discovery team and corporate resources should work together to ensure data collection occurs by the most cost effective method that generates responsive deliveries to discovery requests that are consistent with the legal teams' litigation strategy and the document production schedule.

Organizations often store similar or overlapping data sets within multiple systems to facilitate data entry and reporting functions. Using the company's systems inventory will allow the legal team to identify the specific systems and contained data subsets for a collection responsive to project specific requirements. The systems inventory also provides the discovery team the opportunity to identify the easiest access point to responsive data for both initial and any anticipated supplemental data collections.

The challenge for the legal team is to balance the data collection between too narrow – generating repeat requests and over burdening the client resources – and too broad requests that could potentially generate mammoth and expensive data sets for filtering, review and production. The discovery team should employ legal team strategy decisions regarding responsive data scope and data elements to identify the “Goldilocks” collection approach – the “Just Right” data collection criteria.

Structured data sets in many ways are easier to filter during collection than traditional unstructured data sources. The organized data structures and controlled data entry processes provide – generally – consistent data values upon which the discovery team may construct data requests that closely match the litigation requirements.

Collecting data sets should be based upon the simplest criteria that fits the technical and strategic project requirements. Litigation is a living, breathing entity and legal teams cannot predict future data requests. Having broad, generally responsive data sets available to the legal team will enable quicker responses to updated discovery requirements. Working with previously collected data sets in place is more expedient than having to return to the client to request additional narrow data set identification and collection.

When planning enterprise system collection, leave the chainsaw behind (we are not taking the whole tree), use a hatchet and save the scalpel for the legal team review. This approach will provide greater flexibility to the legal team over the complete litigation lifecycle.

## **ANALYSIS**

Identification of potentially relevant data within the system is the next area of examination. Collected datasets undergo thorough analysis to determine field and data structure. The discovery team should verify that the system data dictionary describing the content, format and structure of the system gathered during the investigation phase matches the collected data set. If not, the team should update existing data dictionaries as necessary. If legacy or archived datasets do not have system information or

documentation available, the project consulting team will reconstruct and analyze the legacy data sets to determine usability, validity and relevancy of data.

Datasets containing privileged, confidential or otherwise sensitive data values must be uniformly redacted, masked or anonymized to meet data privacy considerations and/or to avoid inadvertent disclosure of sensitive or protected data. The legal team should utilize proven programmatic methodologies for consistent redaction, masking and anonymization of data to reduce unnecessary manual review time and costs. The discovery team coordinates with the legal team resources to determine suitability for applying these methods in accordance with project wide processes and standards. The discovery team will provide data descriptions and representative data samples, along with production format and programmatic redaction protocols, in accordance with established redaction protocols.

Final filtering criteria is developed and refined based upon the totality of the available data set (as opposed to data samples or system documentation). The discovery team will document filter criteria at each iteration resulting in a protocol that is repeatable amongst disparate datasets and future supplemental efforts. This method also ensures consistency with any agreed upon criteria developed among the parties to the litigation.

Again, structured data sets provide the ability to implement objective criteria for selecting responsive records. Using standardized keywords against full text fields is discouraged absent specific need. Standardized classification and controlled data entry, limiting and standardizing field values across records, provides a much more accurate and cost effective data filtering tool than common full-text searching options.

## **REVIEW AND PRODUCTION**

When collection is complete, data received from enterprise systems must be transformed into a relevant review set suitable for attorney review, analysis and production. Analysis, filtering and culling occurs in accordance with the legal strategy. Formats for review and production are developed and samples are provided for approval based upon the data selected for inclusion in the production, type and volume of data and programmatic redaction decisions. Review can be conducted in customized interfaces to closely mimic the look and feel of the underlying repository or report view from which the data was extracted.

Flexibility is a key component for successfully producing enterprise system data. Some datasets require multiple formats: PDF files suitable for TIFF production, native data production in spreadsheet, database and/or text format. Attorneys may need PDFs or TIFFs, while experts may require the data in native form.

One useful technique for legal teams is to assign production Bates numbers to each data row produced in spreadsheet or database form. The production number provides a familiar mechanism for legal teams

to reference specific produced data subsets, removing confusion should a data portion be selected for use or when inadvertent sorting occurs in a spreadsheet. Additionally, each data row may have a Hash value generated for the produced data to protect data integrity against inadvertent data modification.

### **CONCLUSION**

At a time when the legal industry has finally mastered the art of managing the discovery processes for loose electronic files and email, along comes the next challenge. Technology, regulatory and political influences have driven the corporate world to adopt enterprise database systems at an increasing pace. The rise of the computing “cloud” and Software as a Service have reduced the cost entry points for both initial implementations and migrations to new generation software solutions. Regulatory compliance requirements (and the costs of non-compliance) have driven implementation of sophisticated solutions. Key, high value, litigation-relevant data sets are contained within these corporate systems now.

Enterprise database systems are technically complex, using data structures, data types and technologies that are unfamiliar to most discovery teams. Often highly optimized and technically complex systems, discovery efforts are usually unexpectedly costly and time consuming. Understanding these systems in a detailed manner sufficient to support timely, cost effective discovery efforts is a new challenge. Discovery teams implementing the standardized approach, such as discussed within this document, will have the best opportunity to both further general corporate litigation management goals and specific litigation strategies.

## **ABOUT GRANITE LEGAL SYSTEMS**

Granite Legal Systems, Inc. is a litigation consulting firm with services spanning the Electronic Discovery Reference Model. We specialize in solving the complex technical challenges inherent in large-scale discovery projects. We provide solutions to clients facing high stakes litigation, arbitration, compliance investigations and regulatory scrutiny.

Granite Legal Systems, Inc. offers a uniquely distinct level of expertise in the area of enterprise systems discovery management. Granite has successfully handled the discovery process for several hundred-million-row data sets and terabyte sized collections from highly specialized and proprietary enterprise systems

Our team is comprised of veteran business and technical analysts and programmers and experts with significant experience in the e-discovery field.

Granite Legal Systems, Inc.  
1201 Louisiana St., Suite 350  
Houston, TX 77002  
(713) 652-0881  
sales@granitelegal.com  
www.granitelegal.com