

Electronic Discovery Without Borders

Your Passport to Managing Multilanguage ESI



Forward

Electronic discovery continues to evolve. I hear time and again that figuring out how to deal with multilanguage electronically stored information (ESI) is a growing challenge for attorneys and their support staff. An increasing number of organizations manage data and employees around the world. With the click of a button, documents and emails can be sent across national borders. Even materials that stay entirely in a single country can be written in multiple languages – just think of “Spanglish” and “Franglais.” For the 2008 Socha-Gelbmann Electronic Discovery Survey, we spoke with 155 organizations; as a group, they conducted a broad range of eDiscovery activities in 83 countries on every continent except Antarctica. Increasingly complex discovery and government investigation requirements only make the challenges greater. When Epiq invited me to participate in the development of a how-to guide on managing multilanguage files, I saw an opportunity to add clarity to this complex topic. While international eDiscovery poses considerable challenges, organizations can streamline the process through a number of smart strategies, better ensuring the integrity of the ESI for a more successful review, analysis, production and ultimately presentation of the data. This guide provides tips and considerations for handling the discovery of multilanguage documents. I hope you will find it to be a useful tool for managing your litigation.

Sincerely,
George J. Socha, Jr., Esq.
Socha Consulting LLC

Introduction

Although litigators today benefit from advanced technology to research and organize their cases, the evolution of electronic data and the internationalization of business are adding new complications that require ever-improving solutions. Such complexities stem in large part from the ease and pace at which electronically stored information (“ESI”) is created, duplicated, disseminated, stored and hidden across the globe.

Investigations by the U.S. government under the Foreign Corrupt Practices Act have increased significantly in recent years. Business continues to expand across international borders, and with that so does the scope of government investigations, litigation and corporate due diligence reviews. As electronic discovery reaches a global scale, attorneys need to know how to handle information collected from foreign entities without incurring excessive risks or costs.

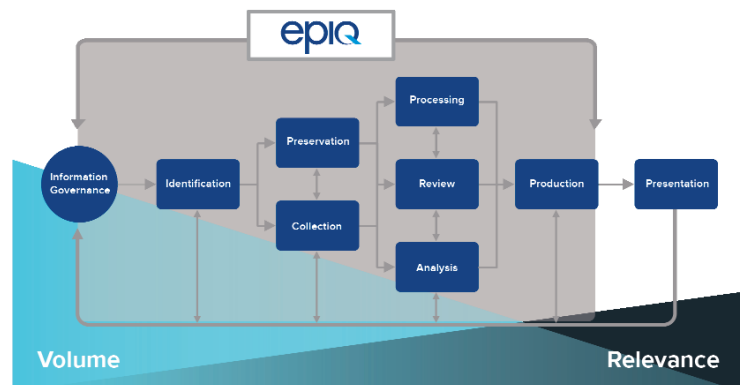
In this pamphlet, you will learn about managing multilanguage ESI throughout the Electronic Discovery Reference Model (EDRM) from identification through production. We will explore the various implications of, and provide recommendations for, handling legal matters that involve multilanguage ESI.

The topics you will read about include:

- The legal restrictions on the handling of ESI collected in a foreign country
- Best practices for the processing and review of multilanguage ESI
- Helpful technologies to assist in handling foreign ESI
- Tips for evaluating ESI translation services

EDRM

The Electronic Discovery Reference Model (EDRM) is a standardized model formally outlining the steps involved in the electronic discovery process. Developed to establish consistent procedures around the once-disconnected stages of electronic discovery, it helps maintain the integrity of ESI during discovery and provides guidelines for eDiscovery consumers and providers. For more information visit: www.edrm.net.



Identification

In this stage we will discuss:

- Determining if your ESI has multilanguage components
- The value of a litigation readiness plan

Determining if Your ESI Has Multilanguage Components

In any matter, the first step to ensuring a successful ESI collection effort is to learn about the locations of potentially relevant data. With businesses increasingly employing people overseas and the interconnectedness of information technology (IT) infrastructures, there may be no clear rhyme or reason to where multilanguage data is stored. Just because a server is located in a particular country, it does not automatically mean that all the documents contained on it will be written in that country's mother tongue or be from individuals in that country.

It is therefore important to conduct interviews with IT personnel and data custodians to assess:

1. where relevant data is located
2. whether it is in foreign jurisdictions
3. how much of the ESI includes foreign languages.

If you are unsure whether or not foreign languages are present, take a sample of key custodians and search the ESI with language identification processes and technology. Unless you know for sure that the ESI includes foreign languages, you should assume there might be and test accordingly.

Language searches use software that quickly scans and identifies the language of unstructured text within individual documents based on linguistic analysis of its word stems. By identifying the language in which a document is written from the outset, an appropriate review strategy can be deployed to route these documents to the correct location with the required expertise.

The Value of a Litigation Readiness Plan

A litigation readiness plan (LRP) is an important requirement to ensure the success of identification, preservation and collection efforts for any matter. However, it is particularly critical for corporations dealing with international eDiscovery. Faced with the Federal Rules of Civil Procedure (FRCP), state rules, common law requirements and foreign regulations, corporations will find that a well-developed LRP serves as an invaluable insurance policy against violations of this complex web of rules.

To develop your LRP, assemble a cross-functional litigation readiness team consisting of key senior level representatives from the following divisions:

- Legal
- Compliance
- IT
- Human Resources
- Risk Management
- Information Security
- Records Management
- Audit
- Operations

Of these groups, legal and IT are generally the key members of the team without whom any litigation readiness initiative is likely to fail. The more participation and coordination, the better conceived your LRP will be.

Preservation and Collection

In this stage we will discuss:

- Preservation and legal holds
- Data maps
- Determining who owns the ESI
- Data privacy laws and blocking statutes
- Harvesting ESI
- Transferring ESI

The overarching consideration when dealing with preservation and collection is that nearly any oversight can be corrected, except the failure to identify and preserve relevant ESI. Whether intentional or not, such an omission could lead to a spoliation claim with severe consequences in terms of penalties or damages.

Preservation and Legal Holds

To efficiently and adequately manage preservation, you should assemble your litigation readiness team as quickly as possible and include representatives from key international locations—if not every international location.

As with any litigation or regulatory compliance effort, your first step is to ensure preservation of relevant materials. When operating on an international scale, however, this process can be more complicated than issuing a standard directive to employees to stop deleting emails or drafted documents.

Outside of the United States, the concept of a legal hold is not always well understood and thus its efficacy can be limited. Therefore, it is imperative to develop a well conceived and targeted communication effort to educate international employees about what steps need to be taken.

Data Maps

At its most basic level, a data map tells you where ESI is located and how to take action on it. Developing a data map is an important component to ensuring well organized preservation and collection, and will help minimize the risk of spoliation.

The key components of an ESI data map include:

- Where ESI is located
- How it is stored
- When and how it is backed up
- Whether backups are rotated or destroyed and, if so, on what schedule

Remember, automatic deletion or rotation policies mean that if you do not take action you may lose ESI that may be subject to a regulatory or litigation hold. Opposing counsel, judges and juries often are not sympathetic to these mistakes. It is therefore crucial to complete your data map early so that the litigation hold can be implemented promptly.

Determining Who Owns the ESI

Rules defining ESI ownership vary dramatically. For instance, in the U.S., an employee's work files and email data generally belong to the employer. The employee has no real expectation of privacy. Outside of the U.S., the laws frequently provide for the opposite—an employee's work computer and email account are considered private, and subject to privacy protection. In certain jurisdictions, it may be a criminal offense to invade that privacy by accessing or transferring that data. Be sure to be fully versed in local laws.

Data Privacy Laws and Blocking Statutes

Some countries in Europe, Asia and South America have data privacy laws that disallow the transfer of personal data outside of their borders. Personal data includes basic information such as names, email addresses and phone numbers. These privacy laws may only provide that data specifically responsive to a request can be exported. In this instance, the data may need to be reviewed locally, rather than in the U.S., in order to minimize the risk of liability or criminal penalties in that country.

Blocking statutes are another consideration. Specifically designed to counter the reach of U.S. discovery, these laws prohibit nationals from disclosing information in response to U.S. discovery requests, even in the face of a subpoena. These statutes can put a respondent in the unenviable position of having to choose between contempt of a U.S. court, or civil and criminal penalties before local courts.

Companies should therefore retain local counsel to oversee official paperwork—including consent forms, commission(er) approvals and labor union approvals—to protect themselves from criminal penalties and civil liabilities.

Harvesting ESI

Choosing an experienced data collection team will be central to your success. "Experienced" doesn't simply mean having an IT background but also means possessing a deep, firsthand knowledge of collection requirements. When harvesting ESI abroad, the collection team must understand the applicable rules in each jurisdiction. If the proper paperwork and authorizations are not in place, collecting ESI subject to privacy protections could put you and your collection team in serious trouble. It is also essential to have a team leader who can credibly testify in court to explain and defend the collection process.

Additionally, if the tools used for the collection are not compatible with the targeted languages, the delays and added costs could be insurmountable. Any time Unicode is involved, it is critical that all of the tools used to collect ESI are able to handle the Unicode format; otherwise, spoliation can occur.

Transferring ESI

Once the ESI is collected, transporting it across international borders exposes the data to a host of potential issues. To mitigate risk, remember the acronym "RED":

- Redundancy
- Encryption
- Documentation

Transporting ESI always involves the risk of data loss. So when practical, a redundant set of data should be created to protect against accidental damage. We also recommend encrypting the data, as the media could be lost, stolen or simply viewed in transit. When dealing with Customs, be prepared to show necessary documentation to explain why encrypted data is in your possession and to explain the purpose of the transfer. You should also have a plan for how to accommodate a request from Customs to show the data that is being transported.

ESI Processing

In this stage we will discuss:

- The phases of ESI processing
- File encodings: Unicode and its importance
- Tips for reducing errors when processing multilanguage files

The Phases of ESI Processing

ESI processing, which is commonly referred to as “loading to the review database,” is the conversion of ESI retrieved from various information systems into a common format that can be filtered, organized, reviewed and produced by attorneys. For example, an Outlook® email may be converted into a record in a review tool that enables a lawyer to search the message, read it, see its attachments and view hidden information that may be stored within it.

ESI is often processed in four phases using specialized software:

Phase 1: Prepare ESI for processing

Phase 2: Process data

Phase 3: Validate processing and address any problems

Phase 4: Load ESI for review

Each of these phases may depend on different processes and technology. It is advisable to test and confirm that the requisite languages are correctly handled.

File Encodings: Unicode and its Importance

Generally, multilanguage ESI can be processed if the technology can support the Unicode set of characters. Also known as “doublebyte,” these characters are more complex and therefore require a second byte of memory to store.

Unicode provides a uniform way for software to encode and understand data irrespective of the language in which it is written. This is important because in addition to a proliferation of languages, there is also a proliferation of character sets. These include those used in Arabic, Chinese, Japanese, Cyrillic (used for Russian and some other eastern European languages) and Devanagari (used for Hindi, among others).

Characters in the English alphabet require a single byte and can be encoded using the more familiar ASCII format. Romance languages such as French and Spanish are also single-byte and can be handled in any system.

The implementation of Unicode support in software is not yet universal, so be sure to check. Beyond Unicode, the following factors will also impact the capacity to process multilanguage files:

- Individual languages contain many different types of encodings that may not be readily detectable. If these encodings are not detected and converted, they will not be readable—even in a Unicode system.
- Certain data collected from unfamiliar operating systems, as can be found in Asia, may not be legible on U.S.-based operating systems.
- Bi-directional languages, such as Hebrew and Arabic, require special accommodations to be displayed and handled properly in a review tool.

Multilanguage data collections can bring about a variety of issues not encountered in typical test environments, so working with a team of experienced professionals is imperative to avoiding common pitfalls and errors. Retroactively correcting problems can cause costly delays.

Tips for Reducing Errors When Processing Multilanguage Files

To avoid problems and reduce errors when processing multilanguage ESI, it's best to confirm the following up front:

- The tools used to collect, stage, copy, process, search, review and export the ESI are Unicode compliant.
- The service provider has specific experience in the jurisdictions where the data is to be collected.
- The service provider has specific experience handling the languages in the current matter. If you hear that a tool is “supposed to” work with a given language, seek evidence or organize a test.
- The platform used to search ESI has the necessary tokens installed to enable searching of specific languages such as Chinese, Japanese, Korean and Thai.
- The service provider should be able to index and search multiple languages in a single system; otherwise, the search and filtering process risks being too cumbersome.

Additionally, it can be helpful if the service provider can offer the following services:

- Identify the primary and secondary languages in the documents during collection.
- Convert image content to text in all necessary languages using Optical Character Recognition (OCR).
- Offer automated language translation.
- Offer human translation.

Document Review

In this stage we will discuss:

- Choosing the right document review software
- Maximizing the document review process
- Evaluating translation options
- Applying best practices

Choosing the Right Document Review Software

While the most important component of a successful eDiscovery review is strong project management, utilizing a quality document review tool is vital to ensuring that you can meet your objectives on time and under budget. In addition to foreign language capabilities, you should evaluate the following:

Technical Considerations

Although the Internet can quickly take you to a distant website in China, this does not mean that you can easily review ESI stored on a distant server. Even with fiber optic technology, a server halfway around the world requires information to travel across numerous Internet servers that relay communications. These long distances can produce excessive latency, or sluggish response. Reviewers in the U.S. may be able to efficiently access a server in Europe, but a team based in China might have difficulty being productive. For this reason, consider service providers with international data centers.

If the review tool is accessed via a web browser, or requires the installation of software on the reviewer's desktop to review native files or foreign language ESI, it could introduce security risks and technical challenges. Remote desktop technology (such as Citrix or Terminal Services) mitigates these problems and eliminates the need to employ software or security restrictions on the local desktop.

Privacy Issues

If the ESI was intentionally kept outside of the U.S. to comply with privacy laws, blocking legislation or other regulatory considerations, it is important to be able to certify that U.S. based users cannot access the data.

If ESI is to be brought into the U.S., working with a service provider that complies with current data protection rules (such as Privacy Shield) can add some assurances. EU-U.S. Privacy Shield Framework Principles issued by the U.S. Department of Commerce means that the service provider is capable of processing and storing ESI in compliance with European regulations. However, the benefit of certification is limited if the data is transferred from one Safe Harbor certified entity to another that is not

certified, which can happen as a result of a document production.

Review Tool Capabilities

Ideally your review tool should enable you to:

- Create a query in multiple languages that can be saved in the system for reuse.
- Create fields and folders in their various languages to accommodate foreign language reviewers.
- View email, Microsoft® Office and PDF documents in any language in native format.

Maximizing the Document Review Process

Even an experienced document review coordinator will find that a multilanguage data set can considerably complicate the planning for a review.

The following guidelines can help minimize the challenge:

- Use reviewers who are fluent in the languages and cultures found in the data set. Their affinity with the language will enable them to understand subtle forms of communication to ensure a more accurate review.
- To the extent possible, keep the team close together. Having all reviewers in close proximity will facilitate coordination and the natural exchange of information among reviewers.
- If the team must be dispersed among two or more countries, ensure that the project coordinators at each site are in regular communication, disseminating new information throughout the review.
- Taking advantage of an automated workflow tool will help organize the distribution of work among various reviewer teams focusing on different language sets, particularly if they are in separate locations.

Evaluating Translation Options

If you do not have attorneys to review documents in their original languages, you can choose to translate the data into a language your reviewers can understand. There are two options:

1. Automated Translation
2. Human Translation

Automated translation is the less costly option, but you generally get what you pay for. This translation software tends to be error-prone, particularly when interpreting the casual communications common to email.

Human translation is more accurate, but time consuming and more costly. Ultimately, its success depends on the skill of the translator, so if proceeding with this option, be sure to utilize capable people.

Important considerations when making your decision are the size of your data set, your budget, your timeframe and the capabilities of your review team.

In both cases, identifying and reducing the number of documents that require translation is the first step, which should proceed as follows:

- Reduce the size of the ESI set by filtering with keywords, preferably with agreement of opposing counsel.
- Detect the primary and secondary language of each document using specialized software, and use this designation to assign documents to the appropriate foreign language review teams. It is best to rely on an automated workflow tool for this process.
- Complete a first pass review using qualified attorneys who can make a high-level cut to identify relevant and non-relevant ESI. This could reduce the volume of the data by over 50%.
- During the first pass review, reviewers should tag relevant documents that require translation.
- For machine translation, process the tagged documents on a routine schedule, and load the translated text in a separate field, preserving the original text. This will provide the ability to search the files in either language.
- If necessary, follow the same process but with a human translation provider to increase quality.

Applying Best Practices

As with any project based on a technology process, test your processes and the technology up front. Request that a representative sample of the languages in your data set be fully tested and evaluated. There are more than 2,000 languages spoken by over 60% of the world population. So, testing and sampling are key components for providing realistic time estimates for completing the review.

Analysis Tools

In this stage we will discuss:

- Developing queries
- Concept searching
- Near-duplicate analysis
- Concept clusters/categorization
- Trained document categorization

Some review tools have advanced analytical features to help reviewers code large groups of files at once, accelerating review time, reducing costs and increasing consistency. When choosing a review tool, you will need to confirm if these analytical features also support foreign languages.

Developing Queries

Searching across foreign languages requires two key components. The obvious one is knowledge of the languages in play. The less obvious one is the development of the requisite tokens necessary to enable the queries to be entered into the system. This process is not difficult but requires planning.

Additionally, the search capability needs to enable complex searches that can handle multiple languages in a single query. Given that international projects will typically include several languages, this feature is more frequently a necessity than a convenience.

Concept Searching

Concept searching enables the retrieval of ESI by concept rather than by exact words. For example, a concept search for the term “device” may retrieve documents that refer only to a “tool” or a “gadget.” This technology is generally compatible across foreign languages but it needs to be specifically tested. Some users of this technology feel that it can result in overly broad results, even when used in English, so bear in mind that evaluating the results in several languages could require additional planning.

Near-Duplicate Analysis

Near-duplicate analysis is the detection of files that are highly similar but not exact duplicates. The benefit of this analysis is to enable the batch coding of groups of similar materials, reducing time and increasing consistency. This technology is generally compatible across foreign languages.

Concept Clusters/Categorization

Concept clusters are designed to show reviewers groups of materials that are related to each other by topic or concept. This technology enables reviewers to view the contents in their database from a high level, which can help them evaluate highly relevant or irrelevant portions of the data. Concept clusters are generally compatible across foreign languages, and are generally as effective with foreign languages as they are with English-only projects.

Trained Document Categorization

Document categorization is similar to concept clusters in that documents are organized by similarity; the approach, however, is entirely different. Categorization involves the creation of a document profile that is based on sample documents chosen by the user. Using this profile, the system can be trained to find more documents like it. This technology is generally compatible across foreign languages. However, regardless of the languages involved, it requires upfront involvement by the user to determine the optimal training documents and to fine-tune the document profiling.

About Epiq

Epiq (NASDAQ: EPIQ) is a leading global provider of integrated technology and services for the legal profession, including electronic discovery, bankruptcy, and class action and mass tort administration, federal regulatory actions and data breach responses. Our innovative solutions are designed to streamline the administration of litigation, investigations, financial transactions, regulatory compliance and other legal matters. Epiq's subject-matter experts bring clarity to complexity, create efficiency through expertise and deliver confidence to our clients around the world. For more information, visit us at www.epiqsystems.com.

