Meeting the Challenges of Asian Language E-Discovery

By John Tredennick, Esq., and W. Peter Cladouhos, Esq.

As e-discovery reaches across borders into Asia, global companies face new and often unfamiliar challenges. Whatever the nature of the case, if it involves electronic information stored in China, Japan, Korea or elsewhere in Asia, be advised: You'll be managing case files differently than you would be if you were in the United States.

The challenges presented in managing electronic files in Asia stem from many causes—some geographical, some technical and some cultural.

In Asian countries, the laws governing data and privacy are quite different than in the U.S. For example, in China, collecting and exporting data involving “state secrets” can get you thrown in jail. In Japan, taking data out and hosting it in the U.S. may cause you to lose your client.

Language, too, presents multiple challenges. The so-called CJK languages (Chinese, Japanese and Korean) are the most difficult to process, search and review. Mangle the processing and you lose your data. Mess up the search and you may as well have lost your data. Either way, your review becomes costly and ineffective.

In this article, we have outlined some of the most common, and the most critical, challenges companies face when handling Asian data and keeping Asian e-discovery on track and on budget.

Consideration #1: Navigating Data Privacy Laws and Political Minefields

Different countries have different laws governing the collection and use of data. These laws may be national, regional or local. The safest practice is to retain local counsel for compliance advice.

In a recent U.S. matter involving a Korean pharmaceutical company, it was necessary to preserve potentially relevant data located in the home country. Korean counsel provided a formal legal opinion that employee consent was required before collecting network-based data from employees who had potentially relevant email residing on servers that required preservation. This consent had to be obtained and documented before the team could move forward with the collection.
But the problem goes beyond a simple reading of the law; you also have to consider the cultural and political climate. For example, in an antitrust litigation involving a Japanese multinational corporation, it was necessary to collect a custodian’s laptop located in China. Although the custodian had provided the proper consent, data shipments out of China sometimes “disappeared” or were confiscated. The safe move was to retain a forensic engineer with the appropriate travel documents to enter China on short notice and physically carry out the data.

**Consideration #2: Working with a Multinational Client**

Many Asian companies doing business in the U.S. are unfamiliar with our discovery rules and practices. In particular, their executives may be unfamiliar with the premise underlying our justice system—that a full exchange of documents among all parties is the best way to reach the truth about a dispute. This is not necessarily the case in Asia, where few if any documents are exchanged before trial. The parties simply bring their witnesses and state their case as best they can.

When working with Asian clients, those who are responsible for managing the legal matter must educate the client early on about the duties and processes surrounding e-discovery in this country in order to keep the project on track. When discussing these matters with your client, make sure you have a fluent native speaker on your team.

Disclosure is not cheap. The client must be warned that the costs of e-discovery are high in the best of circumstances, and multi-language discovery rarely offers the best of circumstances. Expect the process to be tedious, time-consuming and expensive.

In order to have the client’s support and manage expectations, corporate executives must understand the e-discovery process from start to finish, and that may be more difficult—and more expensive—to accomplish with Asian clients.

**Consideration #3: Choosing the Right Vendor**

We cannot emphasize enough the difficulties that Asian languages can present to vendors accustomed to managing English-language documents. To avoid risk, get experienced, qualified help. If the case involves data within Japan, then your e-discovery partner should have experience working in Japan. Ideally, the company should be local, either headquartered there or with a local office.
For collections, however, an Asian-experienced consultant can be based anywhere, provided he has experience working in your jurisdiction. At a minimum, the consultant should be conversant, if not completely fluent, in the language. Finding local vendors to assist is often an effective way to work with one primary consultant across jurisdictions.

Familiarity means more than language, law and customs. Your partner must also understand and be fluent in the technology. This starts with knowing about the operating systems and the various encoding pages you are likely to encounter. If Japanese and Chinese character coding terms such as Shift JIS and Big5 cause consternation for your vendor, you may need to look for another. The same for GUL files and Bequi email. Not all Asian clients use Microsoft Outlook or Lotus Notes; even if they do, they may use older versions that are not Unicode compliant. Collect data in the wrong way and you might see your data expressed as a series of ☐ ☐ ☐ ☐ ☐ and ??? —a disaster for your client.

**Consideration #4: Unique Challenges in Search and Review**

For search and review of Asian-language data, not all e-discovery platforms are up to the task. You need to select a vendor whose search engine incorporates a tokenizer—a tool with the ability to split a string of text into individual “tokens” based on language-specific rules—capable of handling the peculiarities of the particular language.

Search software operates by creating an index of every word within the data set. To create the index, the software identifies words by the spaces and punctuation that surround them. This process, called tokenization, is easy for Western languages, because words are separated by spaces and punctuation.

However, in the CJK languages, characters run together without clear breaks. To identify individual words and create an index, search engines must use special tokenizers to recognize the language and determine where words begin and end.

A further difficulty in searching CJK languages is their use of pictorial characters called logograms. The search engine must figure out how to group these pictorial characters into words. Even more challenging, the engine also must recognize how the characters are combined, because that affects their meaning.

For example, the traditional Chinese word for “Chinese” consists of three logograms that directly translate to “middle country people.” To say “China” rather than “Chinese,” you use two of these characters that represent "middle country."
A basic search engine would have no way of knowing whether to read these characters together or separately and therefore no way of discerning their meaning.

**Consideration #5: Using Machine Translation to Control Costs**

In Asian e-discovery matters involving large numbers of documents or short turnaround times (or worse, both), translation can present an enormous hurdle. Human translation is time-consuming, tedious and extremely expensive. In fact, in a multi-language matter, translation is often the single-largest expense, sometimes even more expensive than review.

Machine translation (MT) may present an alternative in many cases, depending on the nature, volume and complexity of the data.

The 2011 sale by South Korea-based Samsung Electronics of its hard disk drive operations to Seagate Technology for $1.4 billion is a good example of how MT can be used successfully under the right circumstances. Before the deal could close, Samsung had to await Second Request review and approval by the FTC, a fast-track process that gives the FTC the documents and information it needs to evaluate the proposed transaction.

In a Second Request scenario, each party must certify that it has “substantially complied” with the government’s requests, including a requirement that the documents be provided in English. With this substantial compliance standard in mind, the attorneys at Paul Hastings who were representing Samsung faced a quandary: the vast majority of the documents they had to produce to the FTC were in Korean, but they had to be produced in English.

With several hundred thousand Korean-language documents totaling well over a million pages, the cost to translate them all using human translators could easily have approached $15 to $20 million. The Paul Hastings team decided to use MT as an alternative to hand translation.

Although the machine-translated documents did not have the same level of quality as hand-translated materials, they provided a sufficient level of comprehension to allow the FTC to identify relevant materials for further review and, if necessary, to seek hand translation of particular documents. In the end, the machine translations were of sufficient quality to enable the Second Request process to conclude and the merger to go forward. In contrast to the estimated human-translation cost of $15 to $20 million, the final EMT cost was a small fraction of that amount.
That said, it is important to note that there are significant differences among MT offerings in the e-discovery marketplace, and very few are up to the specific challenges of CJK languages. Do not hire an MT vendor without performing a sample translation on data that’s comparable to what you’ll be working with on the case in question.

Conclusion

The challenges presented by an Asian-language e-discovery matter are broad and varied. They span every step of the process—from collection through trial—and they vary by country, case and parties.

In all cases, the key to success in an Asian e-discovery matter is to assemble the right team as early as possible. That includes lawyers, case managers, linguists and vendors with demonstrated experience in the host country and recognized expertise in addressing the language issues. With the right team in place, Asian e-discovery need not be any more burdensome or expensive than the American counterpart.

John Tredennick, Esq., is the founder and CEO of Catalyst Repository Systems, an international provider of multilingual document repositories and technology for electronic discovery and complex litigation. Formerly a nationally known trial lawyer, he was editor-in-chief of the best-selling book, “Winning with Computers: Trial Practice in the Twenty-First Century.” Recently, he was named to the Fastcase 50 as one of the legal profession’s “smartest, most courageous innovators.”

W. Peter Cladouhos, Esq., is firm-wide Practice Support Electronic Discovery Consultant for Paul Hastings LLP. He is a recognized electronic discovery expert and specializes in directing and managing international data privacy and risk management projects.

-END-