



Control Discovery offers a wide range of e-discovery services designed to help law firms and corporate law departments control the cost, quality, and accuracy of e-discovery.

## Quality Questions for the Electronic Discovery RFP

By Jim Barrick, President, Control Discovery, Inc.

### Introduction

The key quality issues facing the electronic discovery industry today relate to conflicts between electronic discovery vendor capabilities and the content of client data. As the volume of electronic discovery data has grown, so too have the number and severity of quality issues.

However, there are still many basic quality challenges that occur solely within electronic discovery vendor's operations. A carefully designed RFP process can help clients identify many of these issues before they make their final vendor selection.

Here we provide "early indicator" questions relating to the most problematic areas of electronic discovery processing. We present a series of suggested RFP questions and explain the relevance of each question.

This document is not meant to be exhaustive, and your client may be best served by a more extensive or more tailored list of questions. We offer this document as one way to try to work toward error reduction in your electronic discovery efforts.

### Overview

Our experience suggests that around 75% of all discovery processing errors are due to issues contained in the common email and office file of client data. We have found that two types of client data issues lead to discovery processing errors:

- Those related to Enterprise IT Practices
- Those related to Custodian Software Usage Practices

The leading contributor to discovery quality errors is what we call "Enterprise IT Practices." We define the term Enterprise IT practices as "the IT practices that define the operation of software applications for all employees in the normal course of business." Slightly over one half (54%) of all discovery quality problems have a root cause in the Client's routine IT management practices.

Another 19% of discovery quality issues have a root cause in the personal technology usage patterns of individual custodians. Custodian data practices are defined as "the personal habits and practices in use of software applications in the normal course of business."

Vendor issues accounted for the final 25%, of all discovery processing error incidents. Discovery quality issues attributable to a failure to "properly test processing software or operational documentation" and those attributable to "missing functionality" ranked about equally within the sample.

### Summary of Electronic Discovery Processing Errors by Attributed Cause

<i>Attributed Cause</i>	<i>Percent of Incidents</i>
Enterprise IT Practices	54%
Custodian Data Practices	19%
Vendor Capability Gaps	14%
Vendor Software QA	11%
Custodian Collections	3%
<b>Total</b>	<b>(100%)</b>

*Source: Control Discovery AMLAW 100 Electronic Discovery Quality Survey*

### Data Extraction

Discovery vendors generally use one of three types of software “filter” technology to extract data and metadata from custodial email messages and native files.

- Commercial Filters (Stellent, Keyview)
- Public Domain Filters (Open Source)
- Proprietary Filters (Vendor developed)

**Q. Please identify the software supplier your firm uses to perform data and metadata extraction. (Example: Keyview or Stellant filters from Autonomy, or public domain, etc.).**

Autonomy is the only player in the industry that holds a development license from Microsoft. This license allows Autonomy to develop data extraction technology filters based on the unpublished file format specifications used by Microsoft Office applications.

**Q. Is the data extraction for any of your firm’s supported file formats performed using public domain software components? If so, please identify the file formats and public domain software components involved.**

Public domain filters used to extract data from Microsoft Office applications may contain numerous defects as most were designed by “trial and error” reverse engineering. None of these public domain filter packages has been licensed by Microsoft. Providers who use public domain filters in their discovery processing operations may be liable for patent infringement claims by Microsoft and others.

**Q. During development of your firm’s electronic discovery processing software, was testing done to ensure accuracy of “document text” extraction for each version of each supported file type? (Example: Microsoft Excel 97, 2000, XP/2002, and 2003, etc.) If so, please provide a copy of the test results identifying each supported file type, version, and test results.**

Neither the Autonomy Stellant nor Keyview filters were designed specifically for the rigorous requirements of electronic discovery processing. While both vendors claim to have performed basic testing for accurate “document data” extraction, neither will publicly release those test results nor provide access to the data files used for internal tests. Extraction software technologies are complex and can be implemented by vendors to varying levels of completeness and quality. We suggest that Electronic discovery vendors who have implemented either of these software technologies be asked to provide evidence that they performed their own validation testing to ensure accurate document processing.

**Q. During development of your firm’s electronic discovery software, was testing done to ensure accuracy of “document metadata” extraction for each version of each supported file type? (Example: Microsoft Excel 97, 2000, XP/2002, and 2003, etc.) If so, please provide a copy of the test results identifying each supported file type, version, and test results.**

Neither Autonomy Stellant nor Keyview has identified or documented the specific metadata their software extracts from supported file formats. While both vendors do extract a number of items of document metadata, there is no documentation from either vendor that tells us what items of metadata we should expect from any specific file of a given format and version. Extraction software technologies are complex and can be implemented by vendors to varying levels of completeness and quality. We suggest that Electronic discovery vendors who have implemented either of these software technologies be asked to provide evidence that they performed their own validation testing to ensure accurate document processing.

**Q. Does your firm maintain a test file library, a collection of files of each supported file type and version for use in testing and retesting the accuracy of processing software? If so, please provide a list of the files, identified by file type and version.**

We believe that it is critical that electronic discovery vendors maintain their own library of native files for test purposes. Both Keyview and Stellant development teams update their software once or twice a year. Implementing these updates increases the risk that errors will be introduced. To ensure accurate processing, testing of all supported file formats and versions should be completed after relevant areas of the vendors processing software have been changed or updated. If the vendor does not maintain this library, the ongoing accuracy of their discovery processing must be viewed with caution.

### **Exception Handling**

As mentioned in the Overview, there are a number of Enterprise IT practices, unique to each client, as well as Custodian-specific software usage practices, that can lead to electronic discovery processing errors. Taken together, we know of between 75 and 100 situations where the content of client data will lead to errors or exceptions during processing.

**Q. Does the firm maintain a list of processing exceptions that can be identified automatically through discovery processing or through quality assurance efforts? If so, please provide a list describing each exception and indicating at what point in the process the exception is identified.**

Exception situations should extend well beyond the simple “virus” or “password” cases, where it is clear that files should not be processed. Focus should be placed on situations where the data is arguably useful to the case and the data errors are not so serious as to preclude processing.

**Q. What criteria does the firm use to distinguish between an exception that is deemed minor and should not be brought to the client’s attention, and one that should?**

In some cases, we have found the error volumes are low, accounting for 0.1% - 0.5% of the total document volume. The error may be related to the idiosyncrasies of one or a few custodians, or a corner case issue affecting a single email server. Many vendors and Lit Support people have told me they believe these “corner case” errors can safely be ignored.

In other situations, the error pattern turns out to be more extensive. What was presumed to be attributable to a specific custodian is found across a group of custodians, geographic locations, or an entire date range of restored backup tapes. We have seen error rates in these situations unexpectedly “explode” to 10% to 40% or even higher. Through experience we now advise clients that it is very difficult to initially differentiate benign errors from those that will eventually erupt and cause major problems.

**Q. Does your firm provide a “root cause” analysis of each identified data error or exception? If so, please provide a sample error analysis template or report showing the methods used.**

The best practice is to understand the underlying cause of each error case as it arises during processing, and to track the cumulative number of each error from the point each exception has been identified. That’s why we believe it is important to review exception handling capabilities of a prospective discovery processing firm. Whatever exceptions are not caught during processing are at risk for growing undetected until they are flagged during the review.

**Q. Does your firm provide an Exceptions Report detailing the files that could not be processed, and the reasons why, for each delivery? If so, please provide a sample report.**

We’ve found no clear pattern describing how troublesome errors emerge. Some errors may not surface until a certain delivery, while others appear sporadically throughout the case until someone notices the aggregate number of errors has grown unacceptably large. Exception reporting helps ensure known problems don’t grow to disabling proportions.

**Q. Does your firm conduct quality assurance inspections on extracted data and metadata (database tables), as well as on TIFF images for each delivery? If so, please provide a summary description of each test, when each test is performed, and what criteria are used to determine “pass” or “failure” for each delivery.**

The exact number of exceptions that the discovery processing vendor can identify through its processing system and internal QA varies widely. Posing the “hundred questions” to the vendor is ultimately not useful until the corresponding issues are pre-identified in client data. Then, comparing vendor capabilities to the issues in client data tells us where the real case vulnerabilities will be.

### **Processing Software**

The remaining 25% of all discovery processing errors are related to the vendor’s processing software. This category includes both software quality failures as well as gaps in processing capabilities.

Vendors who develop and maintain custom electronic discovery processing software may not properly test or deploy changes to that software, leading to errors. Processing software is typically changed and updated continually in response to reported defects, updates to integrated software components maintained by third parties, and the addition of processing capabilities. Cases expected to last 6 months or more will face increased risk due to processing software changes.

**Q. How does your firm respond when it discovers that certain items of client data cannot be accurately processed?**

Vendor capability gaps accounted for 14% of all discovery processing errors. Without full knowledge of the content of client data, and with tight deadlines, situations in which the discovery vendor is caught unaware occur with disturbing regularity. We have learned that the reality is that every case, and especially every large case, may involve client data that a specific discovery vendor cannot process accurately. We want to hear discovery vendors acknowledge this potentiality and commit to early client communication of any such situation.

**Q. In your firm, what is the ratio between quality assurance analysts and software development engineers?**

Defects in processing software account for 11% of all discovery processing errors. As a rule of thumb, there should be two software quality analysts for every development engineer responsible for developing or maintaining discovery processing software. This headcount should not include additional analysts who have operational responsibility for performing quality checks on processed data.

**Q. Does the firm have a software defect tracking system? If so, please provide a current summary report of outstanding software defects showing the number, description, severity by classification, and “days open” since first identified.**

Having a software defect tracking system is mandatory for any discovery vendor engaged in the development of custom discovery processing software. It should be clear that such a system is used effectively to identify, analyze, and repair software defects. High-severity defects should be corrected within 15-30 days. Longer delay times indicate more serious issues in software design, development staffing, or quality assurance. There should be few medium-severity issues older than 90 days. Look for concentrations of defect reports relating to file types of critical importance to your case. If the discovery vendor refuses to release this reporting, or offers only a “sample” report, find another discovery vendor.

**Q. Does the firm have a physically separate hardware and software test environment on which updated discovery processing software can be tested separately from those used to process client data?**

Software change management is often a complex process that should be handled in a systematic fashion to avoid introducing errors into “production” systems responsible for processing client data. If the discovery vendor does not test software changes on physically separate “test” environment, your case data may well become the test bed.

**Additional Assistance**

The field of electronic discovery quality assurance is evolving rapidly. We have only provided what we believe to be the most critical “early indicator” questions relating to the current most problematic areas of electronic discovery processing. Please contact us to learn more. We are always happy to help.

## Further Information

Control Discovery offers a wide range of electronic discovery services designed to help law firms and corporate law departments control the cost, quality, and accuracy of electronic discovery.

For further information about Control Discovery, please contact:

Jim Barrick  
President & CEO

Control Discovery, Inc.  
355 Melrose Avenue  
San Francisco, CA 94127  
T 415-334-9700  
F 415-358-9913

[jim.barrick@controldiscovery.com](mailto:jim.barrick@controldiscovery.com)  
[www.controldiscovery.com](http://www.controldiscovery.com)

© 2005-2009 Control Discovery, Inc. Microsoft Office (logo), OneNote, Outlook, PowerPoint, and Windows are registered trademarks of Microsoft Corporation in the United States and/or other countries worldwide. Adobe and Acrobat are registered trademarks of Adobe Systems, Inc. in the United States and/or other countries worldwide. Lotus and Lotus Notes are registered trademarks and Notes and Domino are trademarks of IBM in the United States and/or other countries worldwide. WinZip is a registered trademark of WinZip Computing. All other trademarks are the property of their respective owners. Specifications are subject to change without notice.