



## Control Discovery

Proven Solutions for Enterprise Compliance

### Email Collections Guidelines

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.

### Generic Email Guidelines

#### Custodian Email Account Identification

Identify and document the email accounts each custodian may have used during the date range in this case, e.g. the earlier of their date of hire or January 2x, 1997 to present. This information is critical if any email backup tapes need to be restored and processed. Add this information to the Master Custodian List.

The Master Custodian list is an Excel worksheet that lists the employee name, email applications used (by date range), email account names (Notes, Exchange, GroupWise), and Email Aliases.

#### Custodian Email Alias Identification

Identify any email aliases the employee may have used within the date range of the case. Add this information to the Master Custodian List. Many users may have had multiple aliases; formal organizational changes, marriage, and the change of default company email addressing schemes are just a few of the reasons why email aliases can change.

#### NOTE:

Deduplication of email in electronic discovery uses data in email header fields. If the "from" address is different for a single employee, email is not deduplicated as well as it could be, and legal review costs increase as a result. It's very difficult for third parties to infer alias "similarity". Anything you can do to add "local knowledge" here will make a big difference. Here are some examples of what we mean by the term "aliases:"

<John B. Smith> jsmith@foo.com  
jsmith@foo.com  
<Rocket Scientist> jsmith@foo.com  
<John B. Smith> jsmith@gocanada.foo.com  
<John B. Wife-Smith> jsmith@foo.com



## Control Discovery

Proven Solutions for Enterprise Compliance

### Email Forwarding

Some organizations make a routine practice of forwarding email from departed employees to remaining employees. If this practice has occurred, the date the employee departed and the name of the new recipient should be recorded. The attorneys should be informed of these cases, as real confusion may otherwise result. Please document any such situations and raise them with your law department.

## Microsoft Exchange / Outlook

### Synthetic PST Creation Tools

An Outlook PST file containing the email for each individual custodian needs to be created from an exchange database. Use a standard tool like Microsoft ExMerge or Mailbox Merge Wizard. We have seen problems with some comparable 'shareware' utilities that do not export mail fields correctly. If you use a third-party utility to complete this task, be sure to conduct a quick QA check of the results to be sure header fields, body, and attachments have been correctly exported.

### Preferred Synthetic PST Version

The internal structure of the PST changed in Outlook 2003 from that previously used in Outlook 97, 2000, and XP. For maximum compatibility with electronic discovery processes, we'd prefer the PST's be extracted for compatibility with Office 2000 or XP. If your site is running Exchange2003, it would be helpful if you could folder these PST's separately under each employee folder in a subfolder named "O2K3" or similar.

### Local PST or MSG files

Occasionally custodians using Windows computers will locally archive email in either a PST or MSG format. MSG files are loose containers containing individual messages/attachments that may be foldered. If your site ever had a policy of restricting employee email storage, be on the lookout for PST and MSG files on the local machines.



## Control Discovery

Proven Solutions for Enterprise Compliance

### Synthetic PST Size Limits

While Microsoft has done a masterful job of hiding incidents of file corruption in its Office product line, errors are in fact present in many PST files, synthetic or otherwise. We strongly suggest that synthetic PST files larger than 1.0GB be split by date range (be careful to ensure that the date ranges are contiguous and the resulting files are named to indicate a split has been performed).

If a loose PST file larger than 2.0 GB is discovered on an employee desktop, be aware that data corruption and loss of email during electronic discovery will result. In this case, an effort should be made to ensure that corresponding email can be restored from either backup tape or your production Exchange server, if possible (breaking PST's into chunks less than 1.0GB).

UNIX Email Applications  
(Netscape, Mozilla, Pine, Kmail, etc.)

Our working assumption is that all UNIX email users connected to the outside world via SMTP email services. This would mean that the only archives of email would be located on local workstations and on backup tapes individual employees may have made and retained. We've seen cases where UNIX email accounts have been tied to Lotus Notes. If your site is one of these, a careful effort will need to be made to collect email from both Lotus Notes databases and local UNIX workstations.

### Local UNIX Email Files

UNIX workstation email will either be found in one "big text file" containing all emails along with Base64 encoded attachments, or in "lots of small text files," each containing one email and encoded attachment pair. You'll need to track down the path for each email application used by each employee. Don't assume the employee only used "Netscape." It's probably best to get the list of email applications each employee may have used and check for the email data associated with each.

### Local Backup and Email Retention

Be sure to ask employees about any tapes archive files ("TAR" or "GZIP") they may have created containing "old" email. Employees who may have used a personal tape backup unit in the normal course of business should be interviewed to identify their personal backup and retention habits, and a reasonable effort made to locate any resulting tapes.



## Control Discovery

Proven Solutions for Enterprise Compliance

### Windows (Downstream) Path Length Limitations

Windows 2000 and XP will be used by the electronic discovery processing technicians and attorneys to review certain data files directly. Windows 2000 and XP cannot properly handle file names longer than 255 characters. Typically, these files will be written correctly but then may not be able to be deleted, and also may have their content corrupted by application software. The average "path+file" in UNIX data can easily exceed 160 characters. When archive files are collected, downstream expansion of these archives can lead to path+file lengths that exceed 255 characters. For any directory names that you create, please try to be concise.

## Lotus Notes Email

### Lotus Notes Version Range

Most electronic discovery processes can handle Lotus Notes 4.5 - 6.5. (1996 - 2005). There are still sites we encounter that have data in Lotus Notes 4.0. Lotus Notes 4.0 presents some potential data recovery problems. If you know for certain that data you are retrieving is in Lotus Notes 4.0 format, please folder these NSF files separately in a subfolder "NSF4.0". Since all Lotus Notes databases share the same "NSF" file extension, it is otherwise difficult to identify these files before electronic discovery processing errors occur.

### Lotus Notes Distributed Links

Some sites using Lotus Notes configure their setup to allow messages in individual databases to "link" to other Lotus Notes databases. This can lead to significant difficulty in Notes email collection. If your site has done or now does this, contact us directly for guidance on this issue.

### Lotus Notes Field Names

If your site has renamed the default mail header fields corresponding to (from, to, date sent, cc, bcc, subject, etc), please contact us directly for guidance on this issue.



## Control Discovery

Proven Solutions for Enterprise Compliance

### Forwarding

Some organizations make a routine practice of forwarding email from departed employees to remaining employees. If you're one of these organizations, the date the employee departed and the name of the new recipient should be recorded. The attorneys should be informed of these cases, as real confusion may otherwise result. Please document any such situations and raise them with your law department.

## Eudora Email

### Eudora Attachments

Eudora stores email messages and attachments in separate locations on the custodian's computer. We have seen cases where attachments were not collected as the employee moved these attachments from the default Eudora attachments directory, or were even deleted. If you encounter Eudora, we recommend a check of the custodian's work habits relating to attachment handling.

## Further Information

For further information about this survey, 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.