



Capturing MetaData

An overview of Dojo services is available upon request. What follows is a detailed summary of the services, software compatibility and output formats offered through our automated, proprietary processes.

Metadata Production

Metadata refers to the information available about an electronic file. This information is only available from the original electronic version of the file so it is imperative the information is captured at file extraction.

Dojo captures and stores metadata in a relational database format. The database format permits the client to search specific fields using keywords or date ranges. The fields available are detailed below.

Email Solutions

Dojo Technology offers a full range of options regarding the production of email files. Dojo has solutions for all the standard email applications including Microsoft Outlook, Lotus Notes, Groupwise, Netscape and others.

All email files are converted to a standard format prior to production. The conversion is done by using proprietary applications.

Email conversion is a three-step process:

- Indexing
- Email and File Extraction
- Conversion

Indexing

The specific data of email and attachments are indexed into a SQL database. Dojo's indexing of email captures the following fields:

Turning Electronic Data Into Evidence

www.dojotechnology.com
866-365-6835

- Control Field
- Dojo tracking name
- Creation Date and Time
- Sent On Date and Time
- Received On Date and Time
- EntryID
- Folder Path
- Recipient(s)
- CC(s)
- BCC(s)
- Sent By
- Number of Attachments
- Names of Attachment
- Subject
- Body of the Email

Information regarding the attachments is also recorded in a database. These fields, excluding Dojo tracking and reference fields, are:

- Email Folder Path
- File Name
- Date and Time Created
- Date and Time Last Modified
- MediaVolume – this is a field used to track the relationship to the original media. The media volume is typically a CD or media source label name.
- Relation (relation of Attachment to Email). This field is used to verify the association between Emails and their Attachments. It is a redundant field to other associations designed into the database structure.

In addition, the following document properties can be retrieved from Microsoft Office files (*the following is a small portion of the total document properties available*). These are often referred to as document properties:

- Document Title
- Author
- Date and Time Created (this can be different than the file property!)
- Date Last Saved
- Last Edited By
- Subject
- Company
- Category
- Key Words
- Comments
- MD5 Hash (used to identify unique documents for document deduping)

Email and File Extraction

During the extraction process the following occurs:

- Emails are examined for embedded objects (i.e. embedded graphics, charts or documents). Emails with embedded objects are extracted as an email file (i.e. .msg file) and examined by Dojo staff. Secondary processes will batch convert all acceptable files.
- Attachments are extracted in their native file format and renamed in correspondence to their Dojo tracking number. This provides for accurate tracking of all files, maintains association to the parent email and assists in the verification of production accuracy.
- The indexing of loose files is done similar to email attachments. The files are extracted in their native file format and renamed in correspondence to their Dojo tracking number. This provides for accurate tracking of all files and assists in the verification of production accuracy.

File Conversion

Extracted files are converted to images and/or the client desired format. This can be in a single or multi-page format of text, tiff, or PDF. The file conversion process includes:

- Extensive formatting of Excel
- The automated removal of date macros in MS Office documents
- Removal of blank pages
- File de-duping (upon request)
- Examination by Quality Control staff of output files
- Any document handling requests of the client
- Final Production – including:
 - Exclusion of document ranges, duplicate documents or or file types (upon request)
 - Single/Multi-Line Branding
 - Bates Labeling
 - Slipsheets – blank or printed with file name/filepath
 - Load File
 - Other. Please see Dojo Work Order for additional custom services (available online at www.dojotechnology.com).