A Lawyer’s Guide to eDiscovery Processing

What You Should Know to Competently Handle Your Case
eDiscovery Webinar Series

About our Webinars

● If you have technical issues or questions, please email webinars@lexbe.com

● Webinars take place monthly and cover a variety of relevant e-Discovery topics

● Lexbe webinars are available for viewing (streaming video), and downloadable as a PDF Presentation or an MP3 podcast. This Webinar and a complete listing of other onDemand webinars is part of the: Lexbe eDiscovery Webinar Series

● For notices of future live and on-Demand webinars as part of this series please email us at webinars@lexbe.com or: Follow us on LinkedIn
eDiscovery Webinar Series

About Lexbe

We are an Austin, TX based eDiscovery software and services provider that achieved over 120% growth in 2016. We, specialize in serving boutique law firms and organizations involved in complex litigation. We provide:

- Full-featured, cloud-based eDiscovery software and services
- A massively scalable and lightning fast eDiscovery platform
- Experienced eDiscovery specialists and expert consultants

Lexbe Sales
sales@lexbe.com
(800) 401-7809 x22

---

‘Cost-effective eDiscovery’
‘Secure, easy-to-use and a great review tool for consideration’
“A powerful litigation document management service”

---

Lawyer's Guide to eDiscovery Processing | What You Should Know to Competently Handle Your Case
Christian Detrude bio

- eDiscovery Solutions Senior Director at Lexbe, a leading provider of cloud-based litigation processing, review and document management software & eDiscovery services
- Certified eDiscovery Specialist, by the Association of Certified E-Discovery Specialists (ACEDS)
- Specializes in working with firms without a full in-house department handling eDiscovery which are involved in the type of complex litigation that requires a high level of precision and eDiscovery expertise to gain the advantage in the discovery phase of trial.

Christian Detrude
800-401-7809 x55
cdetrude@lexbe.com
A Lawyer's Guide to eDiscovery Processing

Agenda

- Growth of Volume of ESI
- Goals of eDiscovery Processing
- Processing in Overall EDRM Workflow
- Scalable Processing
- Processing Steps Overview
- Processing Steps Details
- QC and Security
- Review
A Lawyer's Guide to eDiscovery Processing

Exponential ESI Growth

Source: IDC Digital Universe Study (2012)  ESI = Electronic Stored Information
* 1 Zettabyte = 1 Trillion Gigabytes
In 2016, the number of business and consumer emails sent and received per day will total over 215.3 billion, and is expected to grow at an average annual rate of 4.6% over the next four years, reaching over 257.7 billion by the end of 2020.

<table>
<thead>
<tr>
<th>Daily Email Traffic</th>
<th>2016</th>
<th>2017</th>
<th>2018</th>
<th>2019</th>
<th>2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Worldwide Emails Sent/Received Per Day (B)</td>
<td>215.3</td>
<td>225.3</td>
<td>235.6</td>
<td>246.5</td>
<td>257.7</td>
</tr>
<tr>
<td>% Growth</td>
<td>4.6%</td>
<td>4.6%</td>
<td>4.6%</td>
<td>4.5%</td>
<td>4.5%</td>
</tr>
</tbody>
</table>
**A Lawyer's Guide to eDiscovery Processing**

**Average Case Size & Collection Rising**

**GBs of ESI in a Typical Commercial Case**

- **High**
- **Low**

<table>
<thead>
<tr>
<th>Year</th>
<th>GBs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>Low</td>
</tr>
<tr>
<td>2000</td>
<td>Low</td>
</tr>
<tr>
<td>2005</td>
<td>Low</td>
</tr>
<tr>
<td>2010</td>
<td>Low</td>
</tr>
<tr>
<td>2015</td>
<td>Low</td>
</tr>
</tbody>
</table>

**Enron Criminal Trial (2005)**
- Source ESI: 100M pages (~4 TBs)
- Brought to Trial: 1M pages (~40 GBs)
- Extraordinary at time
- Not now

**Microsoft (2011)**
- Microsoft collects 45 custodians per matter average (2011)
- Almost 1 TB per matter, average
A Lawyer's Guide to eDiscovery Processing

eDiscovery Processing Goals

- Extract metadata & support data reduction (deNIST, culling, dedup)
- Standardize all documents into a review format & allow integrated review or load to a review platform
- Comply with production requirements or other guidelines for standard production formats and loadfiles
- Create high quality search indexes & support data analysis
- Legally defensible processes
- Fast & cost-effective
A Lawyer's Guide to eDiscovery Processing

Processing in the EDRM

Electronic Discovery Reference Model

Electronic Discovery Reference Model / © 2014 / v3.0 / edrm.net
Data volume moves from larger to smaller and less relevant to more relevant along EDRM model as irrelevant data is removed.

Processing is needed during the review and analysis phase to promote more precise data reduction, speed review, and to support legally defensible practices.

Review in Native, Near-native, HTML, PDF or TIFF; Choice driven by review platform capabilities.

Processing facilitates analysis of data and formulation of review strategy.

Information gleaned by processing and analysis of data allows for a faster, more efficient, review which saves time and controls cost.
A Lawyer's Guide to eDiscovery Processing

Processing Workflow

1. Log In to Secure Web Account

2. Upload Native ESI
   - Accessible from all popular browsers on PC and MAC systems
   - Receive email notice upon completion

3. QC & Cull ESI Upload
   - Cull based on document content rather than just metadata

4. Export as Processed Natives, PDF, or TIFF
   - Email notification when complete

5. Load to Review Platform
   - If using Lexbe eDiscovery Platform, immediate access available
A Lawyer's Guide to eDiscovery Processing

Manage the ESI Data Funnel

High Volume | Low Relevance

ESI ID & Collection

EDA & Culling/Filtering

Review & Production

Use (Depos, Motions, Trial)

Low Volume | High Relevance

<< Early Efforts Here Result in
<<

<< Improved Quality and Reduced Costs Here
<<

Lawyer's Guide to eDiscovery Processing | What You Should Know to Competently Handle Your Case
Quality -- Processing must deliver usable and complete outputs to support accuracy and efficiency in data reduction and review, and to support the subsequent stages of discovery.

Speed and Scalability -- Available capacity needs to meet demand. The faster collections can begin and finish processing the sooner review can begin.

Budget -- eDiscovery processing expenditure should be predictable and within budget, and should result in data reduction and review efficiency that cuts overall project costs.

Integration -- Output data should move smoothly into ECA and litigation review platforms to avoid additional time delays and expenses.
Scalable Processing Engine

- Available on demand, as needed, with no costly set up or wait times
- ESI collections can be broken into smaller pieces and processed simultaneously in parallel server environments
- Scalable, proprietary architecture allows for instant access to near unlimited computing power.
- This means faster processing, hours and days vs. weeks.
A Lawyer's Guide to eDiscovery Processing

Processing Key Steps / Features

- Archive/Container Expansion
- File Repair
- Metadata extraction & fielding
- MD5 hash code generation
- System file identification & DeNIST
- Deduplication
- Near Duplicate Identification
- Email attachment extraction & parent email association
- Custodian Assignment
- Native text extraction
- OCR of images
- Indexing of extracted & OCR-ed text
What are Hash Values
○ An MD5 hash (also SHA) is a 128-bit number that (like a fingerprint) uniquely identifies an electronic file.
○ MD5 example: 417BCBDG845179C10D9BBBD1C23294198

How are Hash Values Used
○ Chain of Custody and authentication
○ DeNIST ID and removal
○ Exact Duplicate ID for deduplication

Special Issues with Emails
○ MS Outlook email container files (PSTs) change the included email MSG files every time separated.
○ So standard hashing does not work (different numbers)
○ Instead Hash Values are created from Metadata strings, which are more stable through transmission.
Metadata Use in Processing

Metadata is field-level file information used in review and often delivered with a production as part of a load file (Concordance/Relativity DAT, Summation DII and Lexbe XLSX)

Key Metadata Fields & Use

<table>
<thead>
<tr>
<th>Metadata Field Name</th>
<th>Type</th>
<th>Use in Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date/Time Sent, Date/Time Received</td>
<td>Email</td>
<td>Show when emails sent and received</td>
</tr>
<tr>
<td>Sender, Recipients, CC and Bcc</td>
<td>Email</td>
<td>Show who sent and received emails</td>
</tr>
<tr>
<td>Doc Source, File Path, Custodian</td>
<td>Email &amp; Native</td>
<td>Custodian and chain of custody</td>
</tr>
<tr>
<td>Date Last Modified</td>
<td>Native</td>
<td>Usually best date field for files collected normally (without a forensic collection)</td>
</tr>
<tr>
<td>File Extension / File Type</td>
<td>Email &amp; Native</td>
<td>Show type and quantity of ESI produced</td>
</tr>
</tbody>
</table>

For a full listing of standard loadfiles: [http://lexbe.com/support/technical-resources/](http://lexbe.com/support/technical-resources/)
Fielding Metadata

Email Metadata (sender, receiver, date, time, subject, etc.) is extracted and then fielded to the litigation database for review.

Email Metadata in Outlook Header

Fielded to the Litigation Database
With a Dual-Index approach the search engine indexes both text extracted from Native files (email, attachments, spreadsheets, etc.) and imaged file OCR text (TIFF, JPG or PDF).

Most comprehensive approach minimizes potential for lost and unsearchable data, finds more privileged documents, more PII, and improves the accuracy and quality of culling.

<table>
<thead>
<tr>
<th>Index Method</th>
<th>Captures Embedded Text</th>
<th>Captures Text Excluded From Print</th>
<th>Captures Hidden Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imaged/OCR</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Native Extraction</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Dual Index</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
A Lawyer's Guide to eDiscovery Processing

Benefits of Dual Indexing

HTML Text

OCR Text

Both searchable
Reduce Docs with Culling

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Defensibly remove files from process that are unlikely to lead to responsive documents</th>
</tr>
</thead>
</table>
| Culling Processes | - DeNIST, deduplication  
- Filter by file type & date  
- Keyword filtering  
- Linear vs. dynamic culling |
| Issues | Keyword selection & testing, concept searching, process documentation, repeatability, culled file retention |
| Reduction | ESI may reduce 95% at this stage from raw data size |
Why Duplicates Exist
○ Collection from multiple sources (e.g. Outlook and Gmail)
○ Duplicates between custodians
○ Email attachments
○ Email chains (near duplicates)

Types of Duplicates
○ Exact Duplicates
○ Near Duplicates

Role of Hash Values
○ Separates exact dups from near dups

Vertical vs. Horizontal Dedup
○ Within custodian vs. across custodians
Deduplication

- **Vertical dedupe**
  - dedupes only within custodian
  - May result in less data reduction
  - Unlikely to cause data holes or production gaps

- **Horizontal dedupe**
  - dedupes across all custodians
  - May result in more data reduction
  - Can cause data holes and production gaps
Near Duplicate Detection

What Does It Do?

NearDup technology will group similar documents, even though not exactly the same. Examples include:

- Separately scanned documents.
- Multiple versions of a Word document that are slightly different due to minor edits, reformatting, etc.
- An original document and one with handwritten notes on it.
- Emails and responses that continue a conversational ‘chain’ or ‘thread’.
Near Duplicate Detection

**What Is It?**

- NearDup technology automatically recognizes similar documents within an e-discovery document collection.
- Algorithm analyzes, evaluates and compares the actual text content of the documents to each other.

![Diagram of Unstructured Documents and NearDup Groupings](image-url)
A Lawyer's Guide to eDiscovery Processing

Export Processing Key Steps / Features

- Control numbers/Bates Stamping
- PDF & TIFF creation
- Placeholder creation
- Native extracted, PDF and TIFF loadfile generation in multiple formats:
  - XLSX (Lexbe)
  - DAT/OPT (Case Logistix, Concordance, iPro Allegro, Ringtail, Relativity)
  - DII (Summation)
Latest security technologies and best practices:

- **Encryption**
  Data encrypted (256-bit or above) in-place and in-transit.

- **Data Center Certifications**
  U.S. data centers are certified and follow industry best standards, etc.

- **Clear Ownership Rights**
  Service agreements clearly acknowledge client data ownership.

- **Redundant Back-Ups; Recovery**
  Robust and redundant backup & recovery protocols.
Summary

Processing is an essential element of discovery and the foundation of high quality reviews.

Processing procedures must be legally defensible and result in outputs which comply with the requirements for your case.

Scalable processing solutions are crucial for meeting deadlines and controlling costs.

Processing done right should include a number of basic steps such as metadata extraction and fielding, MD5 hashing, and dual-indexing in order to facilitate fast, accurate, and defensible data reduction and review.

Processing done right saves time and money in the overall eDiscovery process by facilitating data reduction and review efficiency.
Thank You For Attending

We’ll be making the following available to webinar attendees:

- A recorded streaming version
- MP3 podcast
- PDF

Please let us know if you have any questions or comments about this webinar or suggestions for future topics. This webinar is part of the Lexbe eDiscovery Webinar Series. For notices of future live and on-Demand webinars as part of this series please email us at webinars@lexbe.com or Follow us on LinkedIn.

**Presenter:**
Christian Detrude
cdetrude@lexbe.com
(512) 554-5714

**Moderator:**
Jeff Fugitt
jfugitt@lexbe.com
(512) 653-8295

**Webinar Questions:**
webinars@lexbe.com
The Most Advanced eDiscovery Index

Four Critical Data Sources are Processed and Concatenated in the Lexbe Uber Index℠ Resulting in the Industry’s Most Complete Index and Search Platform.
Lexbe eDiscovery Platform

Learn More About Lexbe

- The Lexbe eDiscovery Platform, is our cloud-based processing, review and production tool. Designed for Attorneys/legal staff to be DIY and easy to use, with no users fees or case fees. Free standard loading with annual plans.

Get a Demo   Get a Quote

‘Cost-effective eDiscovery’  ‘Secure, easy-to-use and a great review tool for consideration’  “A powerful litigation document management service”

451

‘Lexbe cost advantages, SaaS convenience and search capabilities appeal to many small firms’  “Because of the Lexbe software, the entire playing field has been leveled for my firm.”  “Lexbe is the easiest eDiscovery software I have ever used”

Lexbe Sales
sales@lexbe.com
(800) 401-7809 x22

Lawyer’s Guide to eDiscovery Processing | What You Should Know to Competently Handle Your Case