

PLANET DATA™ EXEGO™ EARLY COST ASSESSMENT TOOL FEATURES

Planet Data developed our proprietary Exego Early Cost Assessment approach utilizing the Cerulean Engine to provide our clients with best practice methodologies to reduce the amount of data processed and reviewed in litigation. Our internally designed workflow reduces document review time and costs. The ability to search, filter, cull, de-dupe and provide metrics early in the discovery process allows a litigation team to be ahead of opposing counsel at the outset of a case.

FEATURE	DESCRIPTION
Who developed Exego Early Cost Assessment tool?	The Planet Data development team created this tool and source code as part of the Cerulean Engine suite of processing software. There was no outside or contract development for this tool.
Who owns the source code for Exego?	Planet Data owns all source code.
Are any off-the-shelf or outside tools embedded within Exego as part of the processing? For example, Stellent, etc?	No, Planet Data does not embed off-the-shelf or outside technology within the tool to process the data.
Is Planet Data SAS70 Certified?	Yes, Planet Data is SAS 70 Type II Certified and can provide the certification report upon request.
Provide examples of Exego's defensibility/chain of custody.	<p>Planet Data developed and owns the source code and because of this has the ability to precisely know how data is processed. Specific features of our chain of custody include:</p> <ul style="list-style-type: none"> • Retaining original version of source data during processing. • Step by step ability to know what was and was not processed. • Validation and verification at each step in the process. • Ability to customize requirements on a project by project basis. • Reporting through each phase of the process. • QC/QA tools that identify issues with the data that may be unknown to the legal team. • Thorough exception reporting which details files not processed and the reason.
If custom feature sets are required, how is this accomplished?	The Planet Data development team will work with the client to define the requirements and develop the new feature sets.
Is the ECA offering a standalone tool or is it incorporated into an ESI workflow?	The Planet Data Exego ECA tool is a part of the Cerulean Engine workflow.
What search technology is used?	Planet Data integrated DTSearch, which provides advanced searching techniques for a user to cull and receive reports on their data.



FEATURE

DESCRIPTION

<p>Are there any restrictions to providing all metadata for searching?</p>	<p>No, Planet Data processes and delivers all metadata within an email or edoc for searching. It is the client’s decision as to whether they want to view or search all metadata.</p>
<p>What type of searching is available?</p>	<p>The following are examples of the types of searching available within the Planet Data tool:</p> <ol style="list-style-type: none"> 1. Stemming - Common forms of words are found. 2. Synonym Search - Several levels of looking for similar terms using DT Searches, built-in synonym checking or Princeton’s Cognitive Labs WordNet (somewhat broader search). 3. Phonic Searching - Mostly useful for OCR and spelling mistakes. Typically not helpful for ESI, as it is too broad. 4. Fuzzy searching - Mostly useful for OCR and spelling mistakes. Can be controlled and is narrower than phonic searching.
<p>How are embedded objects/documents processed?</p>	<p>Planet Data processes and captures all embedded objects (regardless of level) within an email or edoc. The following are examples:</p> <ol style="list-style-type: none"> 1. Outlook email - Planet Data captures embedded objects within Outlook email. Many embedded objects in Outlook e-mails are image only types and would not otherwise be searchable. 2. Text from Embedded Objects - For example, an email may have a MS Excel document inserted into the middle of the text that could contain text, graphs or other items. When getting the text, most, if not all solutions will only get the text of the email without the visible text from the embedded object. 3. Lotus Notes E-Mail - Planet Data processes Lotus Notes natively and we are able to process embedded objects (which is very complex due to the nature of Lotus Notes).
<p>How many levels of de-duplication are completed?</p>	<p>Planet Data uses a combination of MD5 and SHA1 to ensure the accuracy of our de-duping.</p> <p>Research has shown that files that are different can have the same MD5 value at higher rates than previously thought possible. In addition, there is freely available software that can be used to create different files that have the same MD5 hash signature. As a consequence, we are using SHA1 and MD5 hash signatures to avoid this issue and to detect any attempts at creating files with the same hash signature.</p> <p>De-duping is done based on MD5 and SHA1 hash values for each document. An MD5 and an SHA1 hash is applied at the document level as each file is processed. E-mail MD5’s are calculated based upon the following meta-data fields: To, From, Subject, Send Date/Time, CC, BCC and Body (additional fields can be used as required). The MD5’s/SHA1’s from any attachments are then added to the original value to provide the final hash values. For example, if an email has an Excel file “A.XLS” as an attachment, and later on that same excel file appears as a stand-alone E-DOC, the stand alone E-DOC will be a duplicate. Attachments are never removed from an email under any circumstance.</p>

FEATURE	DESCRIPTION
How are exceptions tracked and managed through the process?	Planet Data provides an exception report that details the document(s) that were slip-sheeted and the applicable reason.
How are image types (tiff, jpeg, etc) processed?	The Planet Data software identifies all image types and then OCRs each image, thereby making it searchable.
How are encrypted files decrypted?	Planet Data applies three different decryption software tools to decrypt data. If after a set period the data can not be decrypted, it is tracked in the exception report and the client is contacted.
How are multiple time zones managed?	Planet Data retains all original time zone settings and can process the data in the time zone of the client's choice and normalize all date/times to that time zone. This is important ,especially when tracking sent dates and times of email.
How is document protection handled?	Planet Data's process automatically identifies and removes password protection.
How are foreign language documents handled?	Planet Data's software is Unicode compliant for metadata and text and can process LOTE (Languages Other Than English). Planet Data can deliver back the text as UTF-8 or UTF-16 depending upon the needs of your review platform.
How is Microsoft Outlook/ Exchange Server handled?	Planet Data processes Outlook and Exchange natively without converting the data, which maintains chain of custody.
How is Lotus Notes handled?	Planet Data processes Lotus Notes natively without converting the data, which maintains chain of custody. We are able to process non-email types of data for Lotus Notes. Many solutions do not process anything other than the e-mail forms.
How is Bloomberg email handled?	Planet Data processes Bloomberg email natively without converting the data, which maintains chain of custody.
How is AOL email handled?	Planet Data processes AOL mail natively without converting the data, which maintains chain of custody.
How are HTML emails processed?	<p>Planet Data processes HTML emails from the following sources:</p> <ul style="list-style-type: none"> PST Item MSG Items EML MBOX Formats AOL PFC Formats

FEATURE	DESCRIPTION
How are system files managed?	System files will be culled and removed from the gross data received for processing utilizing the National Software Reference Library (NSRL) developed by the National Institute of Standards and Technology (NIST).
How is a document type identified?	Planet Data uses multiple methods for determining file types. First, we look at the data internal to the file. We also then compare the extensions to the expected values and if both match, we give it the highest confidence of file typing. At times, due to corruption or where signatures do not give enough confidence, we will also use the extension to validate the document type identification. We then create a confidence level in our document type identification so that we can track how a particular document was doc typed. We are aware that in some circumstances vendors are doing doc-typing by the extension only.
What type of reporting is available?	<p>The following are examples of the Planet Data reports which are provided during the ECA process:</p> <ul style="list-style-type: none"> • Manifest Report - details the source data to be processed and will include the File Name, File Type, and the Size of the file before processing. • Status Report - The daily status report details information such as daily and cumulative documents and pages processed for each job/custodian. • Delivery Manifest Report - details the document types and the counts for each. • Exception Report - details the document(s) that were slip-sheeted and the reason for it. • Delivery Volume Report – details the contents of each delivery including volume names, Control ID information, and document and page counts. • De-duplication Reports – details which documents had duplicates and their location. • Search Reports including: <ul style="list-style-type: none"> - Details total documents hit for query set. - Total number of hits. - Description of query set. - Query Analysis: Document returned by actual query. Number of hits by query. Unique number of documents returned by the query. - Detailed Query Analysis: Number of hits per term/phrase within the query plus any permutations in the search word (i.e. if stemming is turned on all the various forms of the word that were hit). - Indication of where hit was (Specific Meta-data field or Body Text).

