

Hidden Data Embarrasses UN and Creates International Tension

Hidden data in the U.N. report on the Valentine's Day assassination of Lebanese Prime Minister Rafik Hariri raised international concern. The report by Detlev Mehlis, the German prosecutor in charge of the U.N. International Independent Investigation Commission, was published last week. On the surface, the electronic document did not appear to identify individuals by name, but the "Tracked Changes" in the document did reveal the names of several suspects, including the brother and brother-in-law of Syrian President Bashar al-Assad. The changes were apparently made after a meeting between Mehlis and U.N. Secretary-General Kofi Annan. Annan had previously stated that the investigation was an independent process and that the report would not be edited for public release. See the following URL for more information:

http://www.gcn.com/vol1_no1/daily-updates/37416-1.html

The UN Staff appeared to be surprised by the Tracked Changes feature, but long time readers of this newsletter already know the Microsoft Office XP enables Tracked Changes automatically by default. New readers need to know that PowerPoint and Excel also have a Tracked Changes feature. Once enabled, the Tracked Changes can be very difficult to remove, especially in PowerPoint and Excel. Even Microsoft's Remove Hidden Data (RHD) plug-in fails to remove this data reliably. The SRS Document Detective electronic document security toolkit can remove the Tracked Changes from Word, PowerPoint and Excel documents.

Document Reveals U.K. Home Secretary Doubts About New Anti-Terror Policy

A Word document attached to an email intended to support new Government anti-terrorism policies actually showed that Home Secretary Charles Clarke had doubts. Statements deleted in the final edits of the document showed his concern about detaining terror suspects for up to three months without trial. This is another case where the Tracked Changes feature was enabled. A deleted paragraph reportedly read, "The case for some extension is clear, though I believe there is room for debate as to whether we should go as far as three months. I'm still in discussion with the police on this point." For more information, see the following URLs:

<http://software.silicon.com/security/0,39024655,39152367,00.htm>

<http://www.scmagazine.com/news/index.cfm?fuseaction=newsDetails&newsUID=333b032e-eefa-498c-b9a8-2e893cb72b0c&newsType=Latest%20News&s=n>

We are seeing a very high incidence rate in the media because Microsoft Office enables the Tracked Changes feature by default. However, Tracked Changes is just one of many hidden data problems in electronic documents. A good electronic document review policy and software capable of doing a rigorous review is imperative when sharing electronic documents. The SRS Document Detective electronic document security toolkit was designed to meet Government standards for protecting National Security Information.

The President Demands More Information Sharing

Last week, the President signed Executive Order 13388 demanding more information sharing between Federal Agencies that fight terrorism. The order requires agencies to devise policies and implement computer systems that facilitate information sharing to the full extent allowed by law. The new Executive Order also requires that privacy and other rights. Sharing information is important, but knowing what information you are sharing is critical. High level formats allow information to be quickly and easily assimilated, but these high level formats usually include a lot more information than intended. The rush to share more information could result in the inadvertent disclosure of sensitive and classified information.

<http://www.fcw.com/article91216-10-27-05-Web>

NIAP Developing Redaction Tool Protection Profile

SRS recently talked with the Deputy Director of the National Information Assurance Partnership (NIAP), Ms. Pamela Yocum, regarding the testing and certification of the SRS Document Detective electronic document security toolkit. Many potential customers have asked about this certification because of NSTISSP-11 requirements. She acknowledged that there is nothing in Common Criteria for this type of application, but she also mentioned that they have received high level direction to develop a Protection Profile for redaction tools. Document Detective was primarily designed as an electronic document review tool, but it also includes features to sanitize electronic documents which will fit the redaction tool profile. As noted in the next article, Document Detective will soon have a redaction capability as well.

Document Detective Version 1.1 Scheduled for November Release

Document Detective will add Adobe Portable Document Format (PDF) review capability when it is released on 14 November, 6 weeks ahead of schedule. Because of this, the next maintenance release scheduled for mid October was cancelled and the changes were rolled into the new version.

The Phase I PDF capability shows the text of a PDF document arranged into pages. This review tool would have prevented the inadvertent disclosure of classified information when the Special Report on the shooting of the Italian journalist in Iraq was released. Document Detective clearly shows that the text in the redacted paragraphs is still visible in the document. Phase I also identifies the images that appear on the page, but can not yet display those images. PDF is an extremely complex format, and the SRS Phase I capability does not yet handle all of the objects that can be contained in a PDF document, but it does expose and organize all of the objects for consideration. Visually scanning the Objects collection in Document Detective will alert the user to problems that could compromise sensitive information. We have attached a PowerPoint Show to this newsletter showing screen shots of the new PDF review capability.

In addition to a PDF review capability, Version 1.1 adds a lot of new features. Document Detective now has simple editing features that allow the user to make minor modifications to the document under review. This editor is not a full featured editor intended to compete with Microsoft Office. The editor provides a capability to edit some text, to convert dangerous OLE objects and compound constructions into safer images, to compress images to 200 dpi as displayed in the document, and to delete objects. The new geometry algorithms in the PowerPoint and Excel reviewing engines can now find objects obscured by other objects, and objects that are off the page. The document flattener has been improved and now provides options to control and fine tune the flattening process. The document flatteners and the filters that remove Meta data, Tracked Changes, macros and other problems can now be run using a command line preprocessor. This allows users to pre-flatten and filter documents before the review, which usually reduces process to reviewing instances of keywords. The preprocessor could also be used to batch process documents for release; although, we still recommend a human review prior to release.

For brevity, we will not list all of the new features, but there is one more feature we need to mention. The Document Detective toolbar in Microsoft Word will now have a redaction capability. Unlike the recent Microsoft redaction tool, the SRS redaction tool will handle both text and images. The SRS text redaction algorithm does not keep the original spaces, and the actual length of the replaced text is also obscured. Obscuring the length of the redacted text is important when replacing short phrases, such as data in a table (e.g. yes, no, true, false, etc.). Knowing the length of the redacted text could allow an analyst to recover the data. Redaction tools for PowerPoint and Excel are in the works, but they will not be in the Version 1.1 release.

SRS Technologies is committed to continuing research and product development. While no process or procedure is perfect, we can assure you that Document Detective is the best available solution for reviewing and sanitizing electronic documents, especially when National Security Information is at stake.

Please feel free to forward this message to other Government personnel you believe would benefit from this information.

This newsletter is dedicated to raising awareness about the Desktop Publishing Threat to Information Security and to disseminating information that will help mitigate these risks. Our newsletter is published aperiodically when there is news to report. We will not clutter your inbox with idle chatter when there is nothing significant or useful to say. This newsletter is available free of charge to Government personnel and to certain contractors responsible for the generation and protection of National Security Information. Contractor subscriptions are at the discretion of SRS Technologies.

Suggestions and comments are always welcome. If you do not wish to receive this newsletter, please e-mail us and we will gladly remove you from the list. All EDS related correspondence should be sent to rhackett@stg.srs.com.

For a list of articles and information regarding EDS, please visit our web page: <http://www.stg.srs.com/eds>.

To subscribe or unsubscribe from this newsletter, send email to: rhackett@stg.srs.com

SRS Technologies
Systems Technology Group
500 Discovery Drive
Huntsville, AL 35806
Copyright 2005 SRS Technologies



Document Detective

Electronic Document Security Scanner



PDF Review Capability



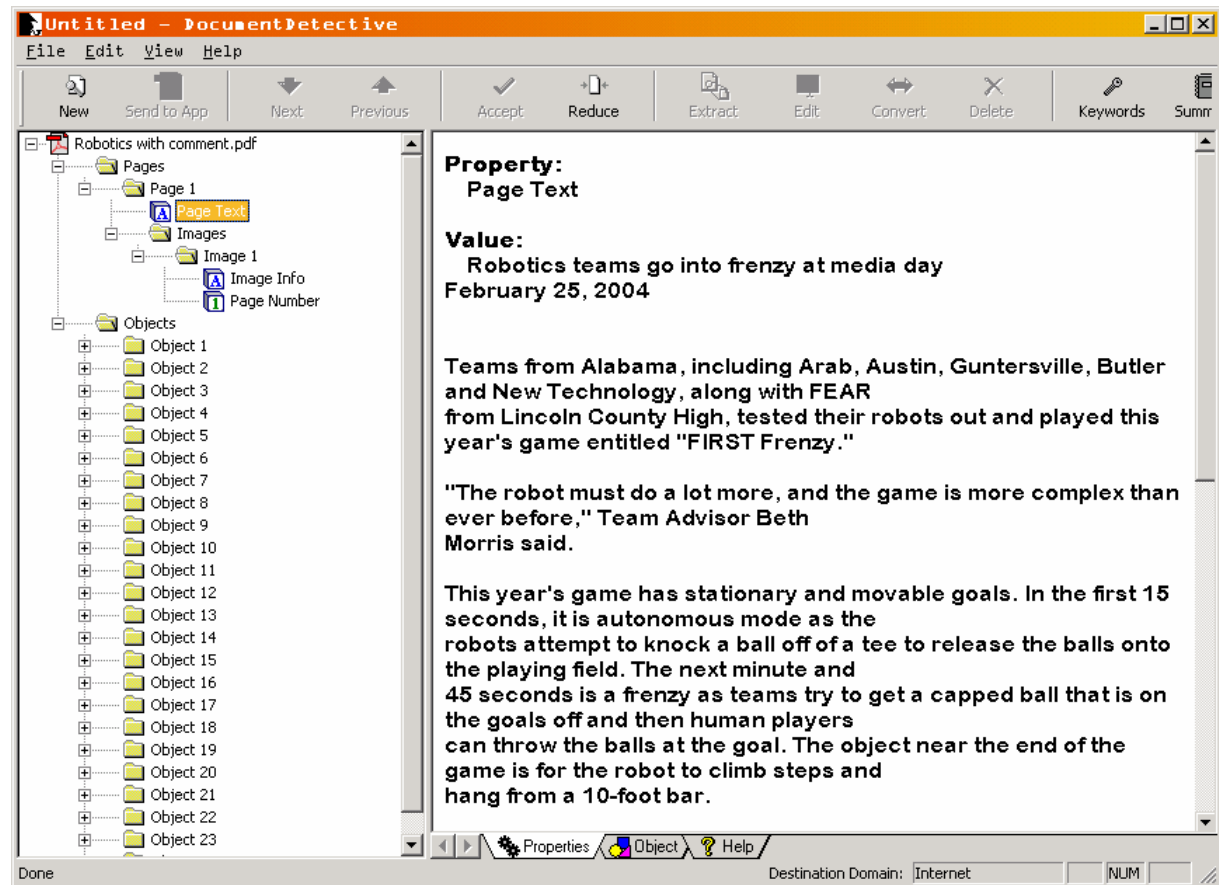
Introduction



Document Detective Version 1.1 now has an Adobe Portable Document Format (PDF) review capability. The Phase I PDF capability shows the text of a PDF document arranged into pages.

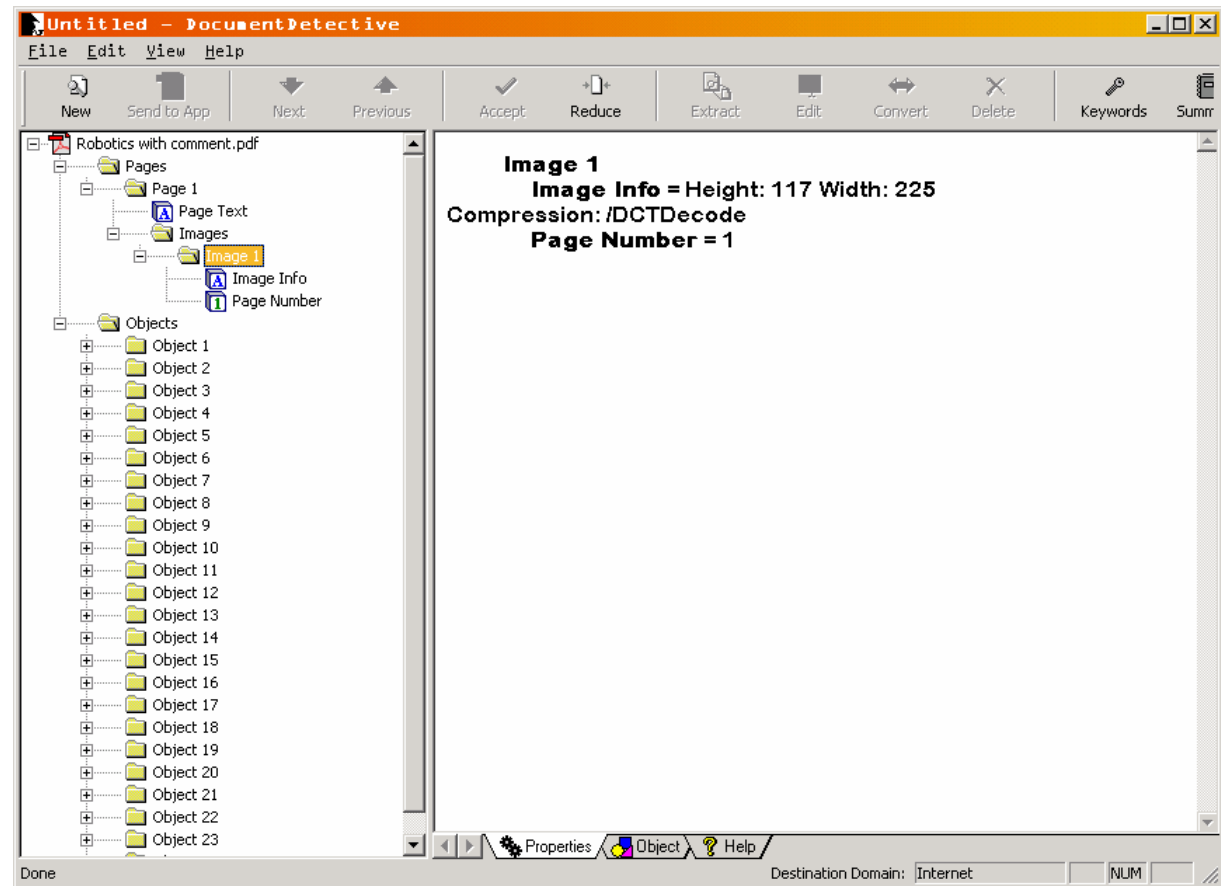
This review tool would have prevented the inadvertent disclosure of classified information when the Special Report on the shooting of the Italian journalist in Iraq was released. Document Detective clearly shows that the text in the redacted paragraphs is still visible in the document.

Phase I also identifies the images that appear on the page, but can not yet display those images. PDF is an extremely complex format, and the SRS Phase I capability does not yet handle all of the objects that can be contained in a PDF document, but it does expose and organize all of the objects for consideration. Visually scanning the Objects collection in Document Detective will alert the user to problems that could compromise sensitive information.

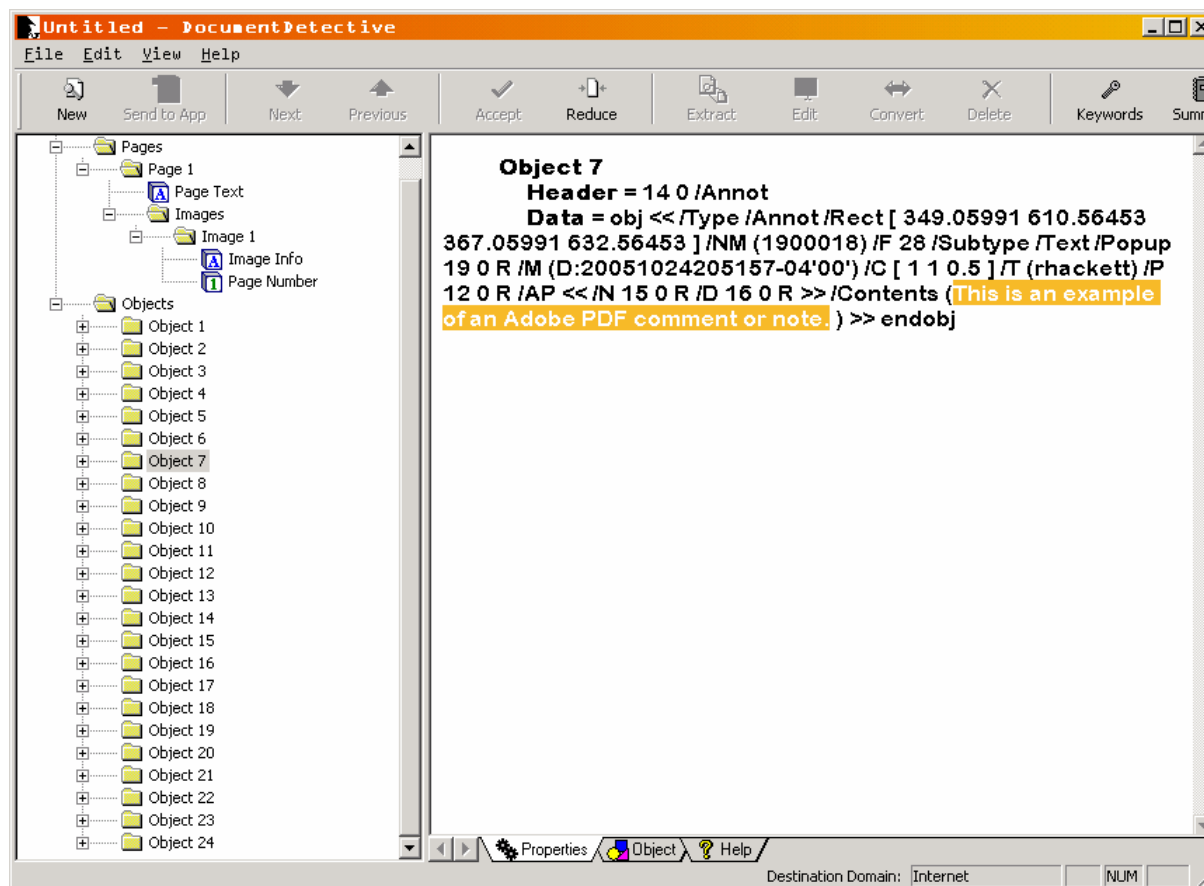


The text for each visible page in the PDF document is displayed in a specific Page folder under the Pages folder. Folders can be expanded or collapsed for the reviewer's convenience. All text is checked by the Document Detective regular expression keyword scanner, and keywords are highlighted for the reviewer's consideration.

The Document Detective PDF capability does not include an editor. If the user finds a problem in their PDF document, they will need to return to the original application to make corrections.



Images appearing on a page are found in the Images folder under the specific Page folder as shown above. The height and width of the image as it is stored in the document is shown along with the encoding or compression algorithm. Images may be displayed differently because of cropping and scaling. Document Detective can not display the images yet, but this allows you to compare the number and size of images you see on the page with what is really stored in the document. For example, if you see two images on the page in Adobe Acrobat, but Document Detective reports five, you may have some hidden images on that page. Just like Microsoft Office, Adobe PDF documents can have objects covering other objects.



Document Detective extracts all of the objects from a PDF document and then parses the objects needed for the individual page folders. The Objects folder contains all of the objects that were not used to create the individual page views in the Pages folder. A lot of this information is formatting data, font descriptors, and information that is not of interest to the reviewer, but some of the information needs to be reviewed. In this example, we see an Annotation (similar to a Microsoft Comment). The text of the annotation is clearly visible in the object data dump.

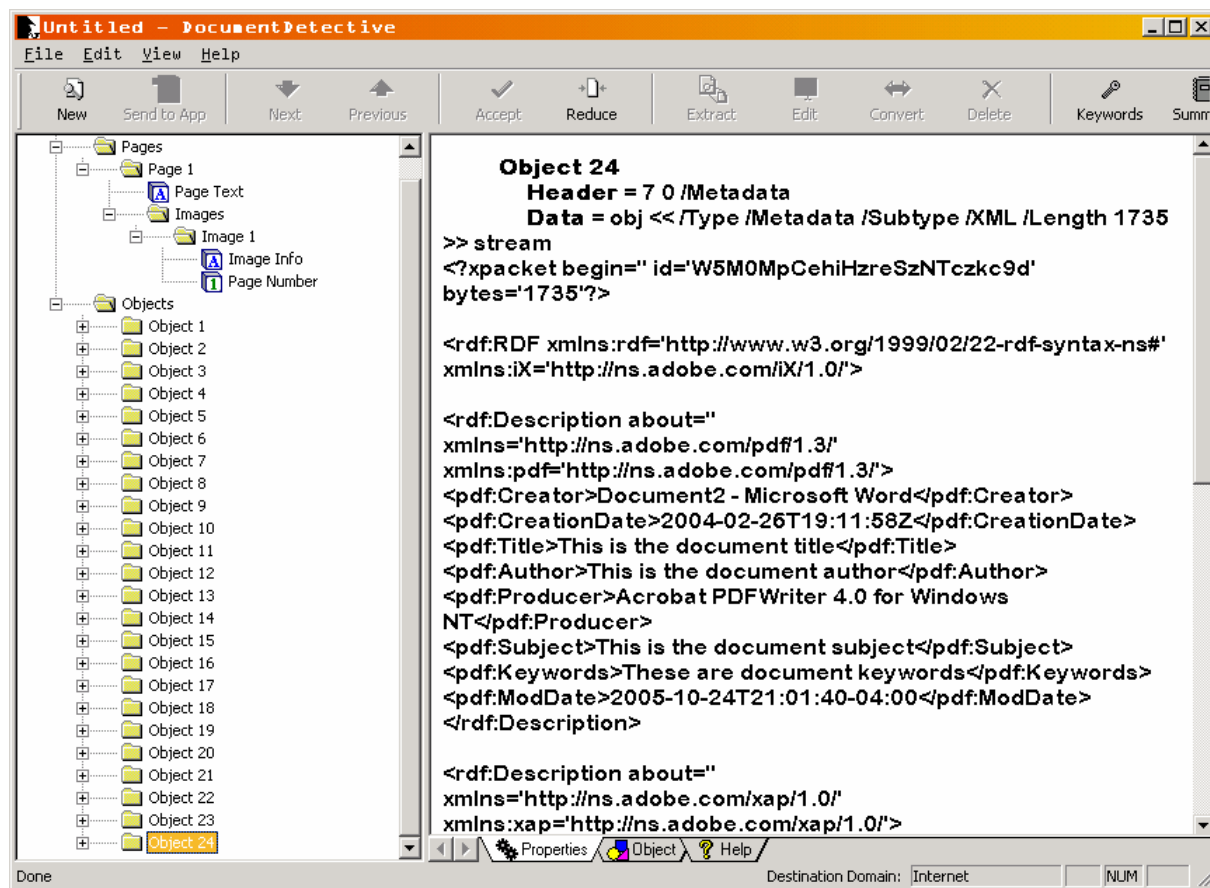


The screenshot shows the Document Detective interface. The left pane displays a tree view of PDF objects, with 'Object 8' selected. The right pane shows the raw data for this object:

```

Object 8
Header = 15 0 /XObject
Data = obj << /Filter /FlateDecode /Length 18 0 R /BBox [ 0
0 18 22 ] /Resources << /ProcSet [ /PDF ]
>> /Type /XObject /Subtype /Form /FormType 1 /Matrix [ 1 0 0 1 0 0 ]
>> stream
H%od=B1 f_ãDIUôgeabâ@@Ï·pÿ^Eá@KiÚÿU&Ã LÖpãCE
" $Ïb Ü40Ä@@èT·Äj... 'QYÜÛÿ_ ;N0?
AxúzLÛU:5'ù^%ã@W-y1b+#v,}
ö-@@0%Æ" <0$%óIG: Ö½ j@·fÏE&¶Tä;7_PP^@5±@@B}fö.:µ'^0 OMD
endstream endobj
  
```

The unused Objects folder should be checked carefully for objects that indicate the presence of significant security risks. The Header shows general information about the object, including the type of object if it is specified in the PDF document. You should be concerned about 'Page' objects which indicate the presence of an unused (deleted) page. A 'Pages' object indicates the document has been saved more than once and may be fragmented. The XObject shown above indicates a orphaned image. The Data property contains a complete copy of the object as it appears in the document. This is a raw data dump, so the object is not decoded or uncompressed. In this case, the object is very small. The user may decide it is an acceptable risk to release this document with this small hidden image.

The screenshot shows the Document Detective application window. The left pane displays a tree view of the PDF document's structure, including Pages, Page 1, Page Text, Images, Image 1, Image Info, Page Number, and a list of Objects from Object 1 to Object 24. Object 24 is selected and highlighted in orange. The right pane displays the content of Object 24, which is a Metadata object. The content is as follows:

```

Object 24
Header = 7 0 /Metadata
Data = obj << /Type /Metadata /Subtype /XML /Length 1735
>> stream
<?xpacket begin=" id='W5M0MpCehiHzreSzNTczkc9d'
bytes='1735'?">

<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:iX="http://ns.adobe.com/iX/1.0!">

<rdf:Description about=""
xmlns="http://ns.adobe.com/pdf/1.3"
xmlns:pdf="http://ns.adobe.com/pdf/1.3!">
<pdf:Creator>Document2 - Microsoft Word</pdf:Creator>
<pdf:CreationDate>2004-02-26T19:11:58Z</pdf:CreationDate>
<pdf:Title>This is the document title</pdf:Title>
<pdf:Author>This is the document author</pdf:Author>
<pdf:Producer>Acrobat PDFWriter 4.0 for Windows
NT</pdf:Producer>
<pdf:Subject>This is the document subject</pdf:Subject>
<pdf:Keywords>These are document keywords</pdf:Keywords>
<pdf:ModDate>2005-10-24T21:01:40-04:00</pdf:ModDate>
</rdf:Description>

<rdf:Description about=""
xmlns="http://ns.adobe.com/xap/1.0"
xmlns:xap="http://ns.adobe.com/xap/1.0!">

```

The presence of more than one Metadata object indicates the PDF document has been resaved and contains fragments. Fragments are deleted portions of the document that may still be recoverable. Document Detective does not parse the Metadata object, but the Metadata object contains readable text. The Document Detective regular expression keyword scanner does scan the unused objects, so keywords in clear text will be found.



Document Detective

For more information, please contact
Ronald D. Hackett, PE
Program Manager
(256) 971-7851
500 Discovery Drive
Huntsville, Alabama 35806
rhackett@stg.srs.com