PDF REDACTIONS AT SCALE





Meet the Presenter



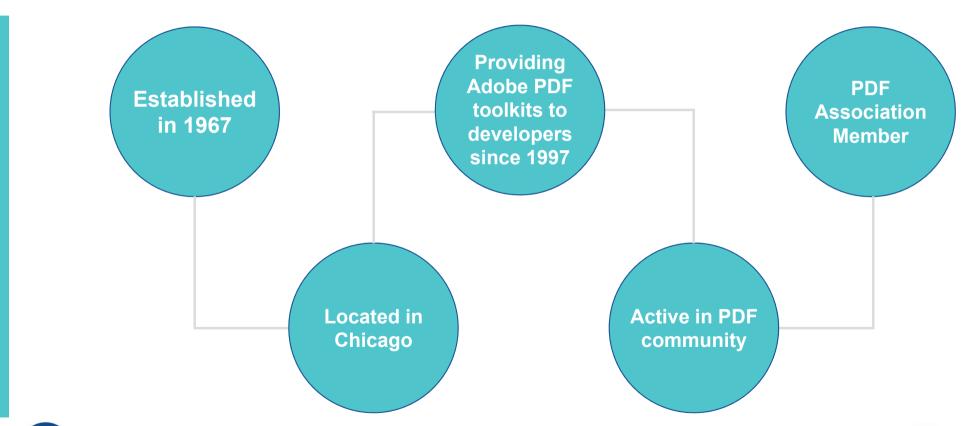
Patrick Gallot

Senior Support Engineer Datalogics





About Datalogics







What We'll Cover

- Redaction in the News
- Introduction to Redaction
- Privacy Concerns
- Technical Concerns
- Challenges Beyond Current Capabilities
- Conclusions & Key Takeaways







REDACTION IN THE NEWS









Atlantic Popular Latest Sections > Magazine > Mo

POLITICS

Manafort's Own Lawyers May Have Hastened His Downfall

The initial failure to redact a sensitive document was the latest in a series of missteps by Paul Manafort's lawyers.

NATASHA BERTRAND JAN 9, 2019



President Trump's former campaign manager Paul Manafort (*left*) and his attorney Kevin Downing depart a U.S. district court after a motions hearing in Alexandria, Virginia, on May 4, 2018. (JONATHAN ERNST / REUTERS)



Grow with Google

MORE STORIES

The Atlantic Politics & Policy Daily: Obama Is .. Biden His Time

MADELEINE CARLISLE , ELAIN GODFREY , AND OLIVIÁ PASCHAL





Case 1:17-cr-00201-ABJ Document 471 Filed 01/08/19 Page 5 of 10

D. The Areas Identified by the Government

1. Mr. Manafort's Interactions with Konstantin Kilimnik

It is accurate that after the Special Counsel shared evidence regarding Mr. Manafort's meetings and communications with Konstantin Kilimnik with him, Mr. Manafort recalled that he had – or may have had – some additional meetings or communications with Mr. Kilimnik that he had not initially remembered. The Government concludes from this that Mr. Manafort's initial responses to inquiries about his meetings and interactions with Mr. Kilimnik were lies to the OSC attorneys and investigators.

It is not uncommon, however, for a witness to have only a vague recollection about events that occurred years prior and then to recall additional details about those events when his or her recollection is refreshed with relevant documents or additional information. Similarly, cooperating witnesses often fail to have complete and accurate recall of detailed facts regarding specific meetings, email communications, travel itineraries, and other events. Such a failure is unsurprising here, where these occurrences happened during a period when Mr. Manafort was managing a U.S. presidential campaign and had countless meetings, email communications, and other interactions with many different individuals, and traveled frequently.

5

D. The Areas Identified by the Government

1. Mr. Manafort's Interactions with Konstantin Kilimnik

It is accurate that after the Special Counsel shared evidence regarding Mr. Manafort's meetings and communications with Konstantin Kilimnik with him, Mr. Manafort recalled that he had – or may have had – some additional meetings or communications with Mr. Kilimnik that he had not initially remembered. The Government concludes from this that Mr. Manafort's initial responses to inquiries about his meetings and interactions with Mr. Kilimnik were lies to the OSC

attorneys and investigators. (See, e.g., Doc. 460 at 5 (After being shown documents, Mr. Manafort "conceded" that he discussed or may have discussed a Ukraine peace plan with Mr. Kilimnik on more than one occasion); id. at 6 (After being told that Mr. Kilimnik had traveled to Madrid on the same day that Mr. Manafort was in Madrid, Mr. Manafort "acknowledged" that he and Mr. Kilimnik met while they were both in Madrid)).

It is not uncommon, however, for a witness to have only a vague recollection about events

that occurred years prior and then to recall additional details about those events when his or her recollection is refreshed with relevant documents or additional information. Similarly, cooperating

witnesses often fail to have complete and accurate recall of *detailed* facts regarding specific meetings, email communications, travel itineraries, and other events. Such a failure is unsurprising

here, where these occurrences happened during a period when Mr. Manafort was managing a







BIZ & IT TECH SCIENCE POLICY CARS GAMING & CULTURE ST

LONDON CALLING -

Facebook pondered, for a time, selling access to user data

Ars simply copy-pasted a 2017 court document into a text editor to reveal redactions.

CYRUS FARIVAR - 11/28/2018, 7:10 PM



Enlarge / Facebook co-founder, Chairman, and CEO Mark Zuckerberg testifies before a combined Senate Judiciary and Commerce Committee hearing in the Hart Senate Office Building on Capitol Hill, April 10, 2018, in Washington, DC.



A failure to adequately redact a public court.

document from February 2017 shows that, back in
2012, Facebook considered charging companies at
least \$250,000 for access to one of its primary.

troves of user data, the Graph API.







In April 2014, Facebook changed the way the previously permissive Graph API works. The social media giant restricted some data access and eliminated all access to the earlier version by June



commercial statements intentionally designed to mislead companies and consumers to participate in Facebook's operating system and purchase advertisements in the operating system, which Plaintiff and many other businesses in fact did, fall squarely under the purview of Section 17500 as its prohibition "extends to the use of false or misleading oral statements" and to recover under Section 17500 "it is necessary to show only that members of the public are likely to be deceived." See People v. Dollar Rent-A-Car Systems, Inc., 211 Cal. App. 3d 119, 129 (1989); see also Moss v. Infinity Ins. Co., 2015 U.S. Dist. LEXIS 158059; Kasky, 27 Cal. 4th at 951 ("Thus, to state a claim under either the UCL or the false advertising law, based on false advertising or promotional practices, it is necessary only to show that members of the public are likely to be deceived") (quotations omitted).

d. It is Premature to Determine the Full Nature of Potential Antitrust Claims as

Discovery in Superior Court is Ongoing, but Ample Evidence of Potential

Cartwright Act Violations Has Been Found in Facebook's Files.

Facebook filed its removal petition on the eve of its deadline to serve its motions for summary judgment and mere days before the Superior Court's ruling on Plaintiff's discovery motions to obtain information from key Facebook executives, including Chief Executive Zuckerberg, regarding the decision to close Graph API that shut down Plaintiff's business and many others.

10

11

23

Plaintiff has yet to receive information regarding this decision that shut down its business. Rather, Facebook has produced documents only from low-level employees that Facebook unilaterally selected as custodians and who clearly had no involvement in the decision that shut down Plaintiff's business.

Cam No. 3-17-cq-00359-Wha REPLY TO OPPOSITION TO MOTION TO REMAND





Improper Redaction Causes Military Snafu

O May 12, 2005 / 🛓 Judi / 🖿 In the News, Privacy, Redaction, Security

via PDFZone

Military's Snafu Highlights PDF Redaction

By Robyn Weisman

On April 30, the U.S. military command in Baghdad posted a PDF in which confidential portions had been redacted—or so officials thought.

John Landwehr, group manager for Adobe Acrobat security solutions and strategy, had a good look at the document after the media got hold of the confidential information, which included names of soldiers manning Baghdad-area checkpoints, training procedures and other security items.

The U.S. military clearly did not understand what it was doing when it used the black highlighter function in a Microsoft Word document that it then converted into a PDF, Landwehr said. Readers could easily separate the black text from the black highlighter and read the classified text.

This matter has not only called attention to the increasing ubiquity of PDF, it also has underscored the importance of proper redaction tools for situations that call for them.

"To be fair, the Department of Defense's first concern is not publishing documents. They are dealing with things that are much more stressful. But an incident like this creates a ripple effect, [so that] greater attention gets paid to the publishing and approval process and people see there is technology available to handle this," Landwehr said.

Adobe does not offer a specific redaction tool in Acrobat, although it does offer other security measures,





INTRODUCTION TO REDACTION

What is Redaction?

Redaction is the process of removing **sensitive** or **classified information** from a document prior to its publication.













The Redaction Process



Search terms or images to redact



Create redaction annotations



Review, modify as needed



Apply redactions





PRIVACY CONCERNS

Important Takeaway

Some tools only put black bars over the text instead of actually removing it.

This is <u>NOT</u> redaction as the underlying text and data is still visible and searchable.

Gaulme, P. and Mosser, B. 2005, Icarus, 178, 84
Gaulme, P. et al. 2011, A&A, 531, 104.
Guillot, T. 2005, Annual Review of and Planetary Sciences, 33, 493
Hedman, M.M. & Nicholson, P.D. 2013, A&A, 146, 12.
Lederer, S.M., Marley, M.S., Mosser, B., Maillard, J.P., Chanover, N.J., & Beebe, R.F. 1995, Icarus, 114, 269
Leibacher, J.W., & Stein, R.F. 1981, The Sun as a Star, 263

Letbacher, J.W., & Stein, R.F. 1981, The Sun as a Star, 263 Lissaner, J.J., & Stevenson, D.J. 2007, Protostars and Planets V, 591 Marley, M.S., & Porco, C.C. 1993, Icarus, 105, 508 Marley, M.S., Gomez, P., & Podolak, M. 1995 JGR 100, 23349 Marley, M.S. 1999 Ph.D. thesis, Univ. of Arizona Marley, M.S. 1990, Icarus, 94, 420 Mosser, B. 1990, Icarus, 87, 198

Mosser, B., Maillard, J.P., & Mekamia, D. 2000, Icarus, 144, 104Rowe, J.F., Matthews, J.M. 2006, ApJ, 646, 1241

Protostars and Planets V, 607

Schmider, F.X., Fossat, E., & Mosser, B. 1991, Astron. & Astrophys., 248, 281

Vorontsov, S. V., V. N. Zharkvov, and V.M. Lubimov 1976, Icarus, 27, 109.



Mark Marley

Mark.S.Marley@NASA.gov

Jack Lissauer

Jack.Lissauer@NASA.gov

Jason Rowe

JasonFRowe@gmail.com





Consequences of Incorrect Redaction



Users can still search for content



Annotations can be moved, causing the content to still be visible



Bookmarks, comments, metadata, or embedded search indexes can still contain sensitive information





Potential Liability

Strictly-enforced rules apply to documents that contain either classified or sensitive personal information and identifiers.

- Consumer privacy
- Legal
- Finance
- Healthcare

Non-compliance can be extremely costly!





Document Sanitation Removes:

- Metadata who created it and how, location and time created
- Attached content
- Comments from a review process
- Previous versions these occur due to incremental saving and could be restored
- Embedded search indexes these could have references to sensitive content





TECHNICAL CONCERNS

Redaction Gone Wrong

Note the redacted sections, however, large chunks of text were also inadvertently removed



Introduction

This white paper suggests a potentially high-reward secondary science target that may be appropriate to include during a revised Kepler planet search observe a field on or near the ecliptic plane we suggest that view. Assuming an appropriate field is chosen there should dition the planet. A longterm (three months or longer) photometric series taken of could potentially detect internal oscillation modes of the planet and open a new windo probing the interior structure of an ice giant. Kepler has demonstrated both that ice giants are common in the galaxy and the exceptional value of continuous photo of electring and interpreting stellar oscillations. A Kepler observation of appropriately combine these two successes to perhaps similarly dissecone of our own ice giants.



al. 2013). It is thus important to understand the interior structure of these worlds in order to better model their formation and evolution. Unfortunately our best data on the interior structure of these worlds comes from the gravitational harmonics measured during the single flybys of Voyager 2 about 25 years ago. Given the uncertainties in those harmonics, a number of possible interior structures and compositions are possible (Guillot 2005).

Guillot (2005) outlines a number of important questions surrounding the ice giants, including the size of their cores and the composition of their deep envelopes. Marley et al. (1995) employed a Monte Carlo method for the construction of interior models and demonstrated that the uncertainties in interior structure were sufficiently great that both distinctly layered structures (e.g., envelope-mantle-core) as well as continuously varying models (a deep envelope with varying composition overlying a core) were possible. Without new constraints on the interior structure, little more can be gleaned from the available data.

Seismology is "by far" (Lissauer & Stevenson 2007) the most promising technique for constraining the core mass of a giant planet, independent of the uncertainties that plaque interior model inversion. In principle, for a fixed spherical harmonic degree, acoustic oscillation modes of sequentially higher order n penetrate progressively less deeply into the interior. Some oscillation modes thus "see" the core while others do not. The progression of mode frequencies—if observed—uniquely delineates the size of the planet's core as well as the structure of the envelope. The basic theory for computing giant planet oscillations has be discussed for thack a — tsov et al. (1976) and includes work by Mosser (1990) for

structure of the envelope. The basic theory for computing giant planet oscillations has be discussed far back a tsov et al. (1976) and includes work by Mosser (1990) for have bee ber of searches for giant planet oscillations, primarily for





Matching Acrobat Behavior

- Repeating replacement text
- Sizing text to fit
- Vertical alignment







Additional functionality from Datalogics PDDoc Functions

PDDocEmbedFontsFromFontArray

Header: DLExtrasProcs.h:231

Related Methods PDDocEmbedFonts

Routine to embed fonts in a document. This will embed only the fonts listed in an array. The parameters to this routine are them same as described for PDDocEmbedFonts. With the exception of the additional fonts and nFonts parameters described below.

Note: If the font has information indicating that it cannot be embedded for print and preview, then the font will not be embedded in the document.

Syntax

void PDDocEmbedFontsFromFontArray(PDDoc doc, const PDFont *fonts, ASUns32 nFonts, ASUns32 flags, ASStatusMonitorProcs statusMon);

Parameters

doc

PDDoc to operate on. The fonts will be embedded in this document. To keep the changes the document must be saved. Fonts on the user's system that match the original font definition will be embedded. For the Times-Roman and Helvetica and corresponding styles, if displayed with a font alias, then the font alias will be embedded in the file. If the font has information that indicates the font cannot be embedded for print and preview, then the font will not be

	embedded in the document.
fonts	An array of PDFont containing fonts to be embedded.
nFonts	The number of fonts in the fonts array.
flags	One of the following:

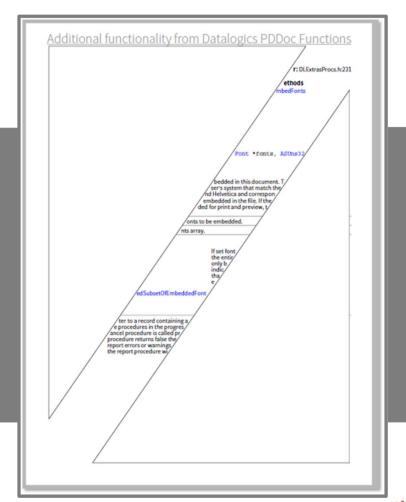
kDocEmbedSubset

If set fonts will be subset when embedded. Otherwise the entire font will be embedded. Full fonts should only be embedded if the information in the fonts indicates it can be used for editible embedding. Fonts that have more than 2048 characters will be subset embedded regardless of the setting of this flag.

kDocEmbedSubsetOfEmbeddedFont If set, fonts that have already been embedded but are not subset fonts, will be reembedded as a subset font.

statusMon

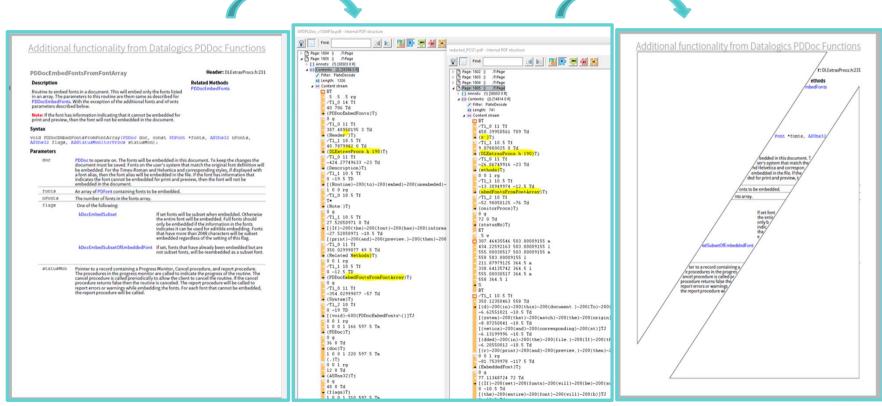
Pointer to a record containing a Progress Monitor, Cancel procedure, and report procedure. The procedures in the progress monitor are called to indicate the progress of the routine. The cancel procedure is called preciocially to allow the client to cancel the routine. If the cancel procedure returns false then the routine is canceled. The report procedure will be called to report errors or warnings while embedding the fonts. For each font that cannot be embedded, the report procedure will be called.







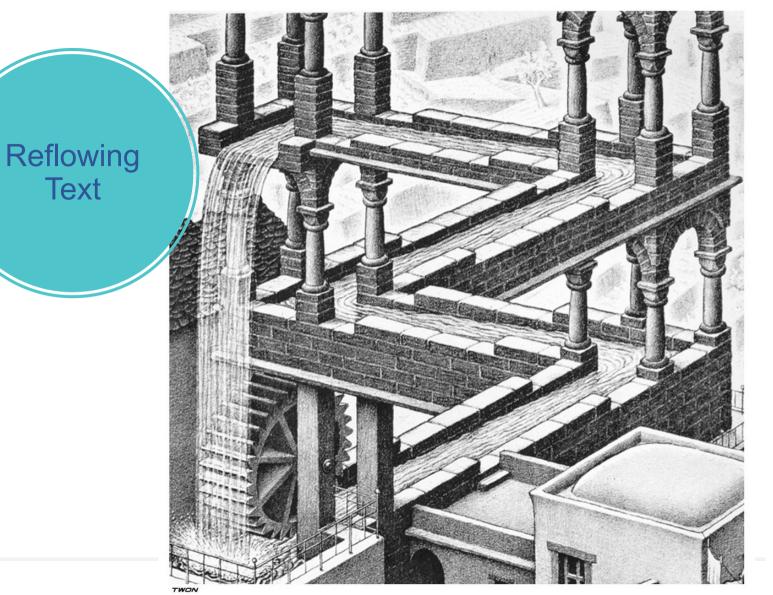
Redaction: Before & After







CHALLENGES BEYOND CURRENT CAPABILITIES



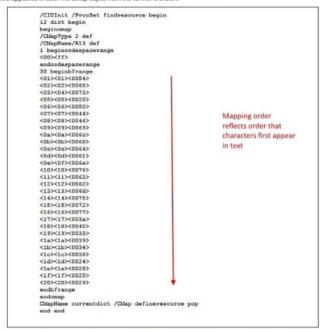








In contrast, the PDF documents created with CutePDF (using open source Ghostscript), or open source LibreOffice Writer, both embedded ToUnicode CMap objects where the order of mappings reflected the order that the characters first appeared in text. The CMap object from the former is shown.



Apart from facilitating the mapping of character codes to character selectors, the CMap created an artefact where the order of the mappings itself encoded a text string. The string contained the password from the PDF document which is highlighted in red below.

Test PDFilhdanvcbmurw:@3946\$(%)



Single-factor authentication

A significant threat to the compromise of user accounts is offline passphrase cracking tools. When an adversary gains access to a list of usernames and hashed passphrases from a system, they can attempt to recover passphrases by comparing the hash of a known passphrase with the hashes from the list of hashed passphrases that they obtained. By finding a match, an adversary will know the passphrase associated with a given username. Combined, this often forms a complete set of authentication information for an account. In order to reduce this security risk, organisations can implement multi-factor authentication. Alternatively, an organisation may attempt to increase the time on average it takes an adversary to compromise a passphrase by increasing both its complexity and length while decreasing the time it remains valid.

Security Control: 0417; Revision: 4; Updated: Sep-18; Applicability: O, P, S, TS; Priority: Must A numerical password is not used as the sole method of authenticating a user.



Security Control: 0421; Revision: 5; Updated: Sep-18; Applicability: 0, P, S; Priority: Must Passphrases used as the sole method of authentication enforce a minimum of 13 alphabetic characters; or a minimum of 10 characters consisting of at least three of the following character sets:

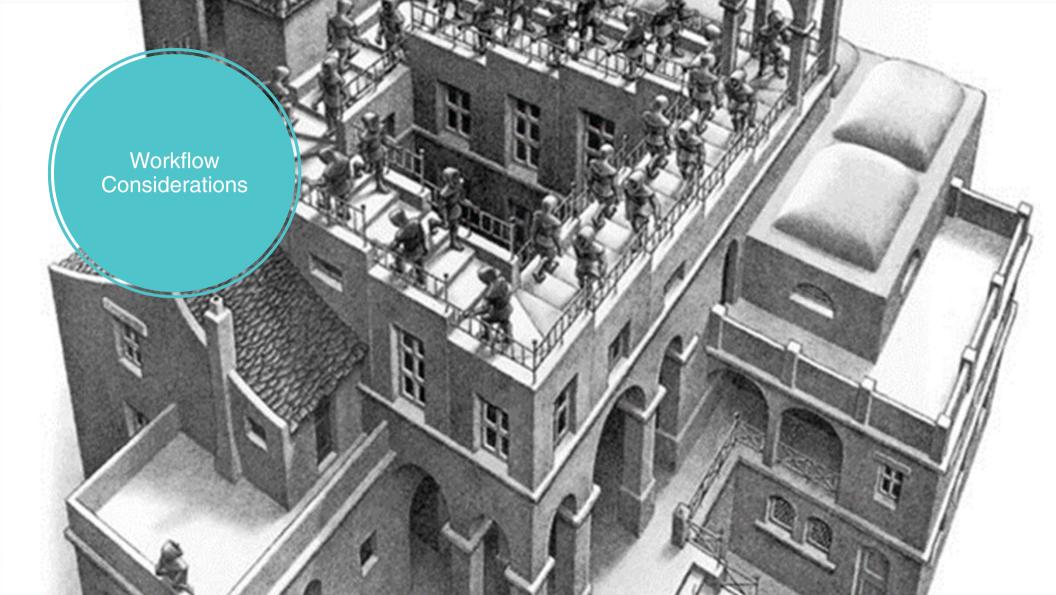
- lowercase alphabetic characters (a-z)
- uppercase alphabetic characters (A-Z)
- numeric characters (0-9)
- special characters.

Security Control: 0422; Revision: 5; Updated: Sep-18; Applicability: Ts; Priority: Must
Passphrases used as the sole method of authentication enforce a minimum of 15 alphabetic characters; or a minimum
of 11 characters consisting of at least three of the following character sets:

- lowercase alphabetic characters (a-z)
- uppercase alphabetic characters (A-Z)
- numeric characters (0-9)
- special characters.







CONCLUSIONS AND KEY TAKEAWAYS

In Summary

- Redaction is a vital tool for repurposing documents for a wider audience, while still being sensitive to privacy concerns
- Redactions work best when the tools work consistently

 It's important to have confidence in the redaction tool you use, and to assured that your documents are fully & completely redacted





Datalogics: Comprehensive PDF Solutions Provider

SDKs ideal for SaaS and OFM implementations





ADOBE PDF LIBRARY

PDF JAVA TOOLKIT

Easy-to-use scriptable server tools seamlessly integrate into existing workflows







PDF Checker



PDF2PRINT



AFRE



FLIP2PDF



We are active in the **PDF** and **Developer** community!









Free evaluations available for all products

www.datalogics.com





QUESTIONS?

Visit us at www.datalogics.com

Patrick Gallot

Senior Support Engineer,
Datalogics



