

PDF についての考察

水越一郎^{†1}

概要: アドビシステムズが 1993 年に開発し、2008 年 7 月には ISO 32000-1 として標準化された PDF(Portable Document Format)は学術論文や公的資料など情報公開時のスタンダードとなっている。その機能は単なる印刷フォーマットではなく、電子署名やメタデータなど情報処理に適したものを有している。国内では官報のように電子署名・タイムスタンプを付与した利活用がされる一方で、墨塗り処理のミスに対する対策として画像化したものを改めてスキャンして公開するといった、情報の利活用に反した事例も見られる。本報告では、1)誤った墨塗りによる PDF 公開の事例の調査結果を報告し 2)学術論文の構造に注目したメタデータの活用について提案する。

キーワード: PDF、墨塗り、メタデータ、日本語処理

1. はじめに

アドビシステムズが 1993 年に開発し、2008 年 7 月には ISO 32000-1 として標準化された PDF(Portable Document Format)は学術論文や公的資料など情報公開時のスタンダードとなっている。その機能は単なる印刷フォーマットではなく、電子署名やメタデータなど情報処理に適したものを有している。国内では官報のように電子署名・タイムスタンプを付与した利活用がされる一方で、墨塗り処理のミスに対する対策として画像化したものを改めてスキャンして公開するといった、情報の利活用に反した事例も見られる。本報告では、1)誤った墨塗りによる PDF 公開の事例の調査結果を報告し、2)学術論文の構造に注目したメタデータの活用について提案する。

2. PDF

PDF はアドビシステムズが 1993 年に開発した。当初から仕様は無償公開されていたが、2008 年 7 月には ISO 32000-1 として標準化された [1] [2]。ISO 化された現行の PDF には電子署名やメタデータを付与することが可能である。

2.1 電子署名

1999 年 11 月以来、国立印刷局は PDF による官報デジタルデータの提供を行っており、2003 年 7 月からは PDF に電子署名とタイムスタンプが付与されている [3]。2012 年に NISC は Adobe 社と協力し GPKI 証明書が同社の Acrobat 等に自動配信されると発表するとともに、政府機関発行文書への電子署名を推奨している [4]。NISC が公開している文章については 2012 年当初から署名がされている [5]。しかし、官公庁の公開文書に対する電子署名の状況調査は筆者が確認した限りでは見つけることができなかった。詳細な調査は完了していないが、筆者による予備調査では官公庁・民間企業・学術論文などで電子署名がされているものは先の例(官報・NISC 公開文書)にしか見つけることができなかった。

った。

2.2 メタデータ

PDF には XML 形式の文書メタデータが含まれており、拡張したり修正することが可能である [6]。またメタデータを抽出するコマンドラインツールも存在する [7]。これらのメタデータは検索などを容易にするといった使い方でなく、デジタルアセットの管理といった用途にも使用される [8]。しかし、メタデータによる情報漏洩の危険性を指摘しこの削除を推奨するものが散見される [6] [7] [8]a。しかし、[9]が主張するように、メタデータを消去することは情報処理の利便性を損なうものである。セキュリティ原理主義とも言える闇雲な消去では無く、メタデータによる適切な情報提供が肝要である。

2.3 誤った墨塗り

PDF 自体に原因はないが、ユーザによる誤った墨塗りに起因する情報漏洩が断続的に発生している。日経・毎日・読売・朝日の各新聞社のデータベースと www.security-next.com (以下 SN) の検索結果 (keyword:PDF 塗、マスキング等)に加え高木浩光氏のページで国内の誤った墨塗りに寄る情報漏洩について調査したところ以下の事例が確認できた。

表 1 誤った墨塗りについての報道^b

報道日付	墨塗り組織	ソース
2003/7/29	総務省	高木氏のページ [10]
2006/8/1	千葉市教育委員会	日経、読売、朝日、毎日
2008/9/2	内閣府国民生活局	日経、
2010/3/25	大阪市水道局	SN
2010/12/11	大阪市	読売
2012/6/14	原子力安全委員会	SN
2013/7/17	防衛省	毎日
2013/7/18	UR 都市機構	SN

^{†1} 東日本電信電話株式会社/情報セキュリティ大学院大学

a 国内での具体的な情報漏洩についての事例はこれらの主張の中に見つけ

ることはできなかった

b 複数報道された場合は初報日を記載

2015/5/27	文部科学省	SN
2015/9/17	一宮市	毎日
2016/7/15	東京都	朝日
2016/8/5	多摩市 図書館	読売, SN
2018/4/26	法務省入国管理局	日経
2018/5/1	神戸市教育委員会	日経、読売、SN

これらの事例はいずれも、画面上では墨塗りがされているように見えるが、実際はペイントツールなどで黒塗り対象レイヤーに重ねて墨塗り画素を置いただけであり、この画素を容易に取り除ける状態になっているものであった。

現行の Acrobat8 には墨塗り用のコマンドが用意されている。これは対象と同じレイヤーで黒塗りされる(対象部分が消去され、墨塗りと置き換わる)ため元の部分を見ることはできない [11]。しかし、2018 年 5 月に大阪市教育委員会が発表した再発防止策 [12]は” 今後、黒塗りした紙媒体を再度 PDF ファイルにしたものをホームページ資料とすることとし、ファイルの情報処理についての知識を局内でしっかりと共有し、データの情報を適切に取り扱うよう取り組んでまいります。”であった。検索除けとも揶揄されるこのような対策の背景には前節のメタデータの闇雲な消去と同様なセキュリティ原理主義があると考えられる。

また、PDF の問題ではないが、墨塗りについては報道[1]にあるように、手作業による漏れや他の情報と付き合わせることで容易に推測可能なものが対象にされているといった課題も存在する。

3. 先行研究

情報を構造化することで処理を容易にしようという研究は多い。上本ら [13]は脅威情報を構造化記述する方式である STIX(Structured Threat Information eXpression)において自由記述が許される Description タグに JSON 形式での記述を採用することで処理が容易になると主張している。山城ら [14]は syslog などの多様な出力フォーマットを処理するためにソースコードからテンプレートを抽出することを検討している。

4. 考察

メタデータの闇雲な消去や、印刷したものを scan して画像を PDF として公開する行為をセキュリティ原理主義と批判することは容易い。また、先行研究にあるように、情報処理のためには情報を構造化することが肝要である。しかし、情報処理の専門家である我々の現状はどうであろうか？

自分たちの PDF に

- 電子署名を付与しているか？

- タイトルや著者といった基本的な情報をメタデータとして提供しているか？
- 概要・キーワード・参考文献といった学術論文に特徴的な情報をメタデータとして提供しているか？

紺屋の白袴という言葉でこの状況を片付けてしまっただけとはいえないだろう。暴論ではあるが、論文投稿サイト側で投稿された PDF のメタデータから概要、キーワードなどを登録する、あるいは登録された概要、キーワードなどを公開時の PDF に付与するなどの仕組みを検討しても良いのではないだろうか。

また、日本語処理という点では一般的な用法に適合していくことも必要であろう。

参考文献

- [1] 山城裕陽、小林論、福田健介、江崎浩，“ソースコードからのネットワークログテンプレート自動生成に関する検討,” 電子通信学会 IA 研究会 2018 年第二回研究会, 2018.
- [2] 上本悠貴、岡村耕二, “STIX を用いた多様化する脅威情報の表現拡張に関する研究,” マルチメディア, 分散, 協調とモバイル(DICOMO2018)シンポジウム, 2018.
- [3] 石. 竜仁, “自衛隊「イラク日報」——ブラック職場が生んだ意味なきぬ黒塗り | 文春オンライン,” 23 7 2018. [オンライン]. Available: <http://bunshun.jp/articles-/8249>. [アクセス日: 4 9 2018].
- [4] 大阪市, “大阪市: 報道発表資料 教育委員会会議議案等にかかる個人情報の漏えいについて,” 2 5 2018. [オンライン]. Available: <http://www.city.osaka.lg.jp/hodoshiryoku/kyoiku/0000434551.html>. [アクセス日: 5 9 2018].
- [5] 齊藤貴之 日経 xTECH, “森友文書でも発生、墨塗り PDF から漏洩 間違った対策 : 日本経済新聞,” 13 6 2018. [オンライン]. Available: <https://www.nikkei.com/article/DGXMZO31095690Z20C18A5000000/?df=2>. [アクセス日: 5 9 2018].
- [6] 高木浩光, “高木浩光@自宅の日記 - 高度バーチャル化社会における墨塗りテクノロジー、はてなダイアリーへの要望,” 29 7 2003. [オンライン]. Available: <http://takagi-hiromitsu.jp/diary/20030729.html>. [アクセス日: 5 9 2018].
- [7] I. Kuksov, “メタデータが漏らす情報 | カスペルスキー公式ブログ,” 24 3 2017. [オンライン]. Available: <https://blog.kaspersky.co.jp/office-documents-metadata/14790/>. [アクセス日: 5 9 2018].
- [8] 上. 宮. 佐. 江. 吉. & 松. 梅澤淳子, “インターネッ

c 本気で検索除けを狙っているとは思いたくはないが...

ト官報データ提供サービスにおけるデジタル証拠性の保証,” 情報処理学会論文誌, 45(8), 1954-1965., 2004.

- [9] NISC, “PDF ファイルの改ざんによるサイバー攻撃への対策について,” 25 4 2012. [オンライン]. Available: https://www.nisc.go.jp/press/pdf/pdf_kaizan_press.pdf. [アクセス日: 5 9 2018].
- [10] NISC, “公開ウェブサーバの脆弱性に関する注意喚起,” 19 1 2012. [オンライン]. Available: https://www.nisc.go.jp/press/pdf/websv_zeijkusei_press.pdf. [アクセス日: 5 9 2018].
- [11] ISO, “ISO 32000-1:2008 - Document management -- Portable document format -- Part 1: PDF 1.7,” 7 2008. [オンライン]. Available: <https://www.iso.org/standard/51502.html>. [アクセス日: 5 9 2018].
- [12] 垣内郁栄, “担当者が振り返る PDF1.7 標準化の道のり - ITmedia エンタープライズ,” 28 10 2008. [オンライン]. Available: <http://www.itmedia.co.jp/im/articles/0810/28/news130.html>. [アクセス日: 5 9 2018].
- [13] so-net, “公開文書の「黒塗り」から情報漏えい相次ぐ——隠すべき情報を守るには:セキュリティ通信: So-net ブログ,” 14 5 2018. [オンライン]. Available: <https://security-t.blog.so-net.ne.jp/2018-05-14>. [アクセス日: 5 9 2018].
- [14] kamuy, “知らぬ間にファイルに摺り込まれる情報、気にしてますか? | kamuy の日記 | スラド,” 13 11 2006. [オンライン]. Available: <https://srad.jp/~kamuy/journal/382313/>. [アクセス日: 5 9 2018].
- [15] toybox, “メタ情報を消すことがセキュリティと勘違いしている人へ | ToyBox - トイボックス,” 14 7 2014. [オンライン]. Available: https://jtoybox.net/note/meta_information.html. [アクセス日: 5 9 2018].

正誤表

参考文献との対応に誤りがあったため以下の通り訂正させていただきます。

2.2 メタデータ

訂正前

れる [8]。しかし、メタデータによる情報漏洩の危険性を指摘しこの削除を推奨するものが散見される [6] [7] [8]a。し

訂正後

れる [8]。しかし、メタデータによる情報漏洩の危険性を指摘しこの削除を推奨するものが散見される [A][B][C]a。し

2.3 誤った墨塗り

訂正前

また、PDF の問題ではないが、墨塗りについては報道[1]にあるように、手作業による漏れや他の情報と付き合わせる

訂正後

また、PDF の問題ではないが、墨塗りについては報道[D]にあるように、手作業による漏れや他の情報と付き合わせる

追加参考文献

- [A] I. Kuksov, “メタデータが漏らす情報 | カスペルスキー公式ブログ,” 24 3 2017. [オンライン]. Available: <https://blog.kaspersky.co.jp/office-documents-metadata/14790/>. [アクセス日: 5 9 2018].
- [B] so-net, “公開文書の「黒塗り」から情報漏えい相次ぐ——隠すべき情報を守るには:セキュリティ通信: So-net ブログ,” 14 5 2018. [オンライン]. Available: <https://security-t.blog.so-net.ne.jp/2018-05-14>. [アクセス日: 5 9 2018].
- [C] kamuy, “知らぬ間にファイルに摺り込まれる情報、気にしてますか? | kamuy の日記 | スラド,” 13 11 2006. [オンライン]. Available: <https://srad.jp/~kamuy/journal/382313/>. [アクセス日: 5 9 2018].
- [D] 石. 竜仁, “自衛隊「イラク日報」——ブラック職場が生んだ意味なさぬ黒塗り | 文春オンライン,” 23 7 2018. [オンライン]. Available: <http://bunshun.jp/articles/-/8249>. [アクセス日: 4 9 2018].