


On PDF/A Conformance and Font Usage in PDF Documents Provided by Public Sector Organizations

Thomas Fischer, University of Skövde, Sweden*

 <https://orcid.org/0000-0003-0272-7433>

Björn Lundell, University of Skövde, Sweden

Jonas Gamalielsson, University of Skövde, Sweden

ABSTRACT

The use of appropriate fonts and file formats for long-term maintenance of digital assets is a challenge for organizations in the public sector. The article reports from a study which investigated the PDF/A conformance and font usage in PDF files provided by Swedish public sector organizations (PSOs). This article presents an analysis of the PDF files' properties and font usage including a categorization of fonts' licenses. This study is motivated by the PDF/A-1 standard's requirement that 'only fonts that are legally embeddable in a file for unlimited, universal rendering shall be used.' Analyzing PDF sets from three PSOs, the finding shows that the proportion of files that claim or succeed at conforming to PDF/A greatly varies among the sets despite similar backgrounds. Although the most popular way to make use of fonts is by embedding a subset of the font data, for some fonts expected to be 'always available,' a considerable proportion of PDF files does not include any font data. This puts the onus of locating this data on the PDF reader which is problematic for long-term archival.

KEYWORDS

Archival, Embedding, Font, ISO, License, Long-Term, Open Source, Standardization, Subset, Typeface

PDF/A CONFORMANCE AND FONT USAGE IN PDF DOCUMENTS PROVIDED BY PUBLIC SECTOR ORGANIZATIONS

Long-term maintenance and archiving of digital assets such as electronic office documents requires the consideration of how to prepare those digital assets for future use. Multiple challenges exist such as the choice of storage technology and file format. Development and use of file formats impose a number of technical and legal challenges (Lundell et al., 2019), and in particular when formats are to be implemented in software (Egyedi, 2007). To allow for a future use of digital assets, file formats that are clearly specified and provided under terms that allow for implementation and use by software

DOI: 10.4018/IJSR.329605

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

projects should be used (Lundell et al., 2023). This includes many formats which are made available as standards by standard-setting organizations such as the International Organization of Standardization (ISO¹) and the Organization for the Advancement of Structured Information Standards (OASIS²).

An example of where an existing file format was standardized is the Portable Document Format (PDF), which, somewhat simplified, allows to describe the content of pages that can be printed. The PDF file format has properties relevant for archiving, such as the read-only option. But the format has drawbacks: documents may refer to external data, which may not be available when reading a PDF file, and the format's specifications may rely on normative references which may be unavailable when implementing tools for those file formats. To address those limitations, a standard—commonly known as PDF/A—was specified by ISO (2005, 2011, 2012, 2020a) with the intent to define a self-sufficient subset of the PDF file format.

Adherence to the PDF/A standard is required by several national archives and national libraries (Bundesarchiv, 2010; Library and Archives Canada, 2015; Rog, 2007). For example, the Swedish National Archives mandate the use of PDF/A-1 if PDF files are to be archived (Riksarkivet, 2009). Determining the conformance to the PDF/A standard faces challenges both due to deficits in the technical specifications of the PDF/A standard and its normative references, and due to implementation deficits in the tools that may get employed to assess PDF/A conformance (Fischer et al., 2021).

To display a PDF file on screen or to print it, the text contained in such a file must be rendered (i.e., put into a graphical representation by using a so-called font program, commonly referred to as “font”). The font program must be available both to the system where the PDF file was originally created as well as to every PDF reader where the file is to be displayed. The technical specifications of each specific version of PDF outline various alternatives of how to make the font available to the PDF reader: embedding parts³ of the font program into the PDF file, relying on standard fonts that are expected to be generally available, and putting the onus on the PDF reader to locate or synthesize a suitable font when displaying the file.

Only embedding the font data into the PDF file allows to recreate the text's shapes and thus guarantees a faithful visual reproduction of the file on the PDF reader's side irrespectively whether the font is locally available to the reader or not. This, however, introduces the question whether the font can be for legal reasons (Groves, 1992) included into the PDF file (i.e., if the PDF file's author has the right to use and include copyrighted font data into the PDF file and then distributes the file). In the context of long-term archival, it is difficult to determine for a PDF file whether the author has legal right. It's difficult for an archiving organization to determine the legal status of used fonts, and if necessary, acquire the permission to use the font.

Addressing those legal challenges, the PDF/A-1 specification imposes the requirement on conforming PDF files that “only fonts that are legally embeddable in a file for unlimited, universal rendering shall be used.” (ISO, 2005, p. 10). The aspects of embedding fonts are further elaborated: “This part of ISO 19005 precludes the embedding of fonts whose legality depends upon special agreement with the font copyright holder. Such an allowance places unacceptable burdens on an archive to verify the existence, validity and longevity of such claims.” (ISO, 2005, p. 11).

Based on these arguments, we address the following research questions:

RQ 1: To what extent do PDF files from public sector organizations (PSOs) conform to the PDF/A standard and what characterizes those files?

RQ 2: How are different fonts used in the collected PDF files?

We investigate how well PSOs perform in providing PDF files that conform to the PDF/A standard and which fonts are used and into which licensing category those fonts can be put. The investigation is focused on PSOs as transparency laws apply to this type of organizations which further motivated our investigation: only file formats that are suitable for long-term archival provide citizens the best possible access to those files.

BACKGROUND

On the Evolution of PDF and PDF/A

The initial PDF format was introduced in 1993 by Adobe as a successor to the PostScript format from the same company (Perry, 1988). Since its inception, a number of revisions of the PDF format were published. Some revisions of the PDF format were the base for ISO standards such as ISO 32000-1:2008 (ISO, 2008) which is based on PDF 1.7 (Adobe, 2006). From PDF version 2.0 onwards, new PDF specifications themselves are published by ISO (2017, 2020b).

There exists an ISO standard describing subsets of regular PDF specifications, and which is designed for archiving, commonly referred to as “PDF/A” (Sullivan, 2006). There exist four parts of PDF/A, designated as PDF/A-1 (ISO, 2005), PDF/A-2 (ISO, 2011), PDF/A-3 (ISO, 2012), and PDF/A-4 (ISO, 2020a). Among the first three parts, PDF/A-1, based on PDF version 1.4 (Adobe, 2001), has the narrowest subset of features. PDF/A-2 and PDF/A-3, both based on version 1.7 (Adobe, 2006), allow for more features such as more image formats or embedding of attachments. A significant difference between PDF/A-2 and PDF/A-3 is that the former only allows for the embedding of documents that conform to PDF/A, whereas the latter does not have this limitation. The newest part of this standard, PDF/A-4 (ISO, 2020a), is based on ISO 32000-2:2020 (ISO, 2020b), also known as PDF version 2.0. Each of the four parts may contain several levels designated by letters such as “a” or “b.”

Adobe, as the original author of the PDF specification, and other organizations have claimed to hold standard-essential patents (SEP) on PDF and its normatively referenced standards (Lundell et al., 2019). Any patents which may have impinged on the PDF 1.4 format can be expected to have expired as the format was released in 2001, and a patent is only valid for 20 years (Thatcher & Pingry, 2007). For ISO 32000-1:2008 (i.e., PDF 1.7) Adobe issued a royalty-free Public Patent License (Adobe, 2008) for their SEPs for compliant implementations. However, Adobe reserves the right to revoke this license and states that it does not apply to updated versions of ISO 32000-1. This may limit the adoption of PDF/A-2 and later parts (Lundell et al., 2015, 2019; Lundell et al., 2022).

There are technical challenges involved in the process of generating PDF/A conformant files (Klindt, 2017; Koo & Chou, 2013; Suri & El-Saad, 2021). Determining whether a given PDF file conforms to PDF/A is far from obvious. A file may claim conformance to a specific standard part, but to verify this claim requires a detailed inspection and expert knowledge. Parts of the verification process can be delegated to specialized tools that check the technical details of a PDF file against codified rules derived from the PDF/A specifications.

Fonts, Their License Categories, and Embedding of Fonts in Other Documents

In order to reproduce the visual appearance of text in a PDF file, the necessary font data must be available to the PDF reader. A PDF specification, here exemplified by version 1.7 (Adobe, 2006), may provide several alternatives how this font data may get stored or acquired:

1. According to this specification, the “most predictable and dependable” result can be achieved by embedding the complete font program into the PDF file.
2. As embedding whole font programs may consume much space, only those parts— subsets—of the font program that are necessary to render the included text are embedded.
3. Embedding of fonts can be avoided completely if the font selection is restricted to size 14 standard fonts (Adobe, 2006, p. 416). PDF readers have then to provide size 14 fonts or substitutes.
4. A font can be referred to only by name, relying on the PDF reader to locate this font in its local environment.
5. Only the font metrics and styles are stored in a PDF file. Then, even if the font is not available to a PDF reader, it can try to locate a different font where the metrics match closely to the expected values.

General issues with font usage and embedding are discussed by Faroult (2016) and Royster (2011). As faithful visual reproduction is important for archival, PDF/A specifications such as PDF/A-1 (ISO, 2005) or PDF/A-4 (ISO, 2020a) only allow for the first two alternatives, where font data is embedded into the PDF file.

However, embedding the font data is not without problems, as Klindt (2017) writes: “Licensing problems may arise in converting to PDF/A for example if the copyright holder of digital typefaces does not allow embedding in documents.” In 2016, WIPO’s Standing Committee on the Law of Trademarks, Industrial Designs and Geographical Indications sent a questionnaire to its member states and intergovernmental organizations. When asked whether their jurisdiction provides protection for typefaces and type fonts, 76 out of 87 respondents confirmed such a protection (WIPO, 2017).

The legal protection of fonts is also highlighted in various PDF specifications—for example, PDF 1.4 (Adobe, 2001, Section 5.8), PDF 1.7 (Adobe, 2006, Section 5.8), and PDF 2.0 (ISO, 2020b). The latter expresses the relation between fonts and copyright as follows: “Font programs are subject to copyright, and the copyright owner may impose conditions under which a font program may be used. These permissions are recorded either in the font program or as part of a separate license.” The costs involved with not adhering to font license restrictions have been highlighted by Evans (2014, p. 316) as follows: “The size of the typeface industry and high fees involved in owning, commissioning, and using typefaces can lead to expensive disputes.”

Font vendors are aware of their customers’ interest in using their proprietary fonts in custom-generated PDF files. Adobe (2023) clarifies the legal use of its fonts as follows: “The fonts are licensed for embedding in any e-book format which protects the font data such as [...] PDF. Any e-book authoring workflow which requires the user to move the font files themselves is not allowed under the terms of use, however. As with any electronic document the fonts must only be used for viewing or printing existing content, not for creating new document variations, templates, or dynamic content.” The terms of use (Adobe, 2022) phrase those limitations technology-agnostic: “You may use desktop fonts to design and develop documents, and you may embed copies of the desktop fonts into your document for the purpose of printing and viewing the document. The font must be subset to include only the glyphs necessary for displaying the work, and the document must obfuscate or protect its embedded font data from deliberate or inadvertent discovery or misuse. No other embedding rights are implied or permitted under this license.” Another example is Microsoft’s license terms for Microsoft 365 consumer subscriptions (Microsoft, 2023), which tie the use of the fonts to the use of the software systems with which they were shipped: “While the software is running, you may use its fonts to display and print content. You may temporarily download the fonts to a printer or other output device to print content, and you may embed fonts in content only as permitted by the embedding restrictions in the fonts.”

The embedding restrictions in the fonts most likely refers to the “fsType” field in OpenType fonts (Microsoft, 2021b) which contain a number of flags where the font’s legal owner can describe the embedding rights for the font. This field has two parts. The first states which of the following four mutually exclusive embedding alternatives are allowed for this font:

Restricted license embedding (RLE): The font may not be embedded without first obtaining the explicit permission by the font’s legal owner.

Preview & print embedding (PPE): The font may get embedded in a way that allows reading and printing the document at the viewing user’s side, but it does not allow modifications to the document using this font.

Editable embedding (EE): Editing is allowed, for example to correct the text, but the font may not be extracted or used in other documents.

Installable embedding (IE): On the PDF reader’s side, the font may get permanently installed and used for other purposes. The viewing user is still subject to the same end-user license agreement as was the original user. In practice, this flag has limited value in addition to what “editable embedding”

allows, as confirmed by Microsoft's typography documentation (Microsoft, 2021a): "In practice all of the applications we are aware of treat these fonts the same as those set to 'editable embedding.'"

The second part of the field `fsType` describes two limitations for embedding the font which may be set independently of any other flag. The first limitation is subsetting the font is not allowed (i.e., the font must always be embedded completely). The second limitation is the font must be embedded as a bitmap instead of a scalable vector graphic.

To avoid potential legal and technical issues with embedding fonts into documents, an alternative is to skip including any font program and refer to it by name or identifier only. PDF readers then make use of locally available font programs to visualize the document. For example, Casario et al. (2012, p. 55) suggests this as a viable alternative for fonts seen as ubiquitously available: "Finally, we set the character "Arial" as the last alternative because it is available on practically any computer with any operating system."

The availability of those fonts may be tied to a specific operating system or other piece of software system the user must have installed and have a valid license for. Otherwise it would result in a situation as described by Bean (2015) who used an operating system that did not include the de-facto standard font Times New Roman, presumably to cut licensing costs. The author continues that this could not be easily mitigated: "And you can't even buy Times New Roman—not from Monotype, not from Google." Another example is provided by iText (2020) who wanted to use the font Arial from a Windows installation on a Linux machine, which is not allowed under the license Windows is distributed with. Instead, the authors resorted to fonts under permissive licenses. A popular example (Crossland, 2012) for such a permissive license is the SIL open font license (OFL). This license's introduction states that "it enables font authors to release their work under a common license that allows use, bundling, modification and redistribution." (Spalinger & Gaultney, 2007).

There exist further legal challenges with the distribution of fonts beyond the domain of licensing. A recent court ruling in Germany (LG München I, 2022) deemed embedding fonts in webpages from third parties (Google Fonts) as conflicting with the European general data protection regulation (GDPR) due to the unnecessary transfer of personal data.

RESEARCH APPROACH

In order to address the research questions, we devised a study and performed the following steps:

1. Identification and retrieval of PDF files.
2. Analysis of the PDF files, primarily to which extent they conform to PDF/A with the help of conformance checking tools, but also other metadata in order to answer RQ 1.
3. Analysis of which fonts are used in the PDF files as well as assessment of the licensing status of those fonts using a heuristic approach to answer RQ 2.

For the first step, we focused on PDF files published by public sector organizations (PSOs) in Sweden. We chose Sweden as the geographic and legal area for our study as it has been used in previous research (Fischer et al., 2021) and has a long history of transparency in the public sector allowing for easy access to documents from various organizations (Stenbeck et al., 2021). Public documents from PSOs (e.g., doctoral dissertations) are also required to be long-term archived (Regeringskansliet, 1993, 2012).

Sweden has several hundred governmental PSOs (Statistics Sweden, 2021) and even more on lower administrative levels which separately or jointly maintain public repositories of documents. From the available document repositories, we chose the following ones to retrieve data for our analysis.

First, building upon previous work (Fischer et al., 2021), we retrieved doctoral dissertations published at 28 public Swedish universities for the publication period of 2018 to 2021. We identified 9605 dissertations by their unique identifier assigned by the universities' libraries. For 9406

dissertations, the corresponding PDF files could be retrieved. The discrepancy can be explained by the lack of a reference to a PDF file corresponding to a dissertation in a library's catalog; the linked PDF files were clearly marked as only short summaries or cover pages, or dissertations were published in formats other than PDF. Doctoral dissertations may get written as part of projects spanning several universities and each involved university library may record the same dissertation separately. To identify such duplicates among the dissertations, we compared the titles. Among the dissertations we acquired PDF files for, we located 71 that were listed in two libraries' catalogs and one dissertation that was listed in three catalogs. As we were interested in unique PDF files, we checked each of the 145 concerned PDF files to discern if it would be identical to another PDF file in the set of duplicates and we removed it if that was the case. Sixty-five PDF files were removed, and 9,341 PDF files were kept.

Second, we retrieved official publications of laws and regulations from the Swedish Code of Statutes, sv: Svensk författningssamling, SFS as published by the Swedish government on a dedicated webpage (Regeringskansliet, 2023) since 2018. This document repository was selected due to its importance for the Swedish society. We collected 5931 PDF files, one for each SFS published on this webpage at the time of this study (from SFS 2018:160 to SFS 2021:1344).

Third, in September 2019, the Swedish government began an investigation to assess PSOs' need for secure and cost-effective IT operations⁴. In this process, 155 comments from various PSOs and 29 comments from various private actors were submitted in the form of PDF documents. As the focus of this study is on PSOs, only the 155 submissions from PSOs were used. This set of PDF files was chosen as it broadly samples PSOs.

For each of the collected PDF files from the three sets, we extracted the following metadata: PDF version, claimed PDF/A conformance (if any), and names and metadata (embedding status, permission flags) of used fonts.

For each file that claimed conformance to PDF/A, it was first tested by supplying this file to one conformance checker, veraPDF (Wilson et al., 2017, version 1.20.3). Following the recommendation that several independent conformance checking tools may give more reliable results (Fischer et al., 2021), files that passed veraPDF's conformance check were then analyzed by an online PDF validation tool called 3-Heights PDF Validator Online Tool (PDF Tools AG, 2022). This validator was chosen as it is an online service corresponding to the 3-Heights PDF Validator used in previous research (Fischer et al., 2021).

For fonts used in a PDF file, the PDF writer can set an arbitrary text as the fonts' name. However, in many cases those texts match or contain the fonts' original names, which allows to deduce the original names. In case that only a subset of a font was included in the PDF file, PDF specifications require a random, six letter long string to be prepended (Adobe, 2006). For our analysis, we normalized the extracted font names to the presumed original base name. For example, the extracted font name "QOZYTL+TeXGyreHeros-BoldItalic" got normalized to "TeXGyreHeros".

Based on the identified font names, we used a heuristic approach to determine the fonts' license status. We considered two approaches to identify under which license a font got released. First, based on the font's name, we searched publicly available data to locate information on license and original vendor. Second, as Linux distributions typically state under which license included fonts are available, one can check if the font under investigation is provided by a distribution and retrieve all legal data from there. Ultimately, we limited ourselves to the first approach, as the second did not provide information on non-open fonts and most inspected PDF files only made use of a limited selection of fonts making a manual investigation feasible.

As a variety of licenses for fonts exist, we categorized the legal conditions under which a font may be available into one of the following categories in order to simplify the following analysis:

Open: The font is available under an open license (i.e., a font license which does not restrict the modification or redistribution of the font program). One of the most prominent examples for an open

font license is the SIL open font license (Spalinger & Gaultney, 2007). An example of fonts released under this license is the font family Liberation.

Proprietary: A commercial license must be acquired or the font’s use and distribution is otherwise restricted such as “for personal use only.” An example is the font Calibri.

Ambiguous: There exist multiple independent fonts with the same name under open and proprietary licenses, respectively. An example is Symbol.

Unknown: The font’s name is inconclusive to determine its origin. An example is AdvOT40514f85.

CHARACTERIZATION OF PDF FILES AND CONFORMANCE TO PDF/A

This section addresses the first research question: To what extent do PDF files from public sector organizations conform to the PDF/A standard and what characterizes those files?

PDF files contain metadata, such as which PDF version the file follows and whether it conforms to PDF/A. Table 1 shows the distribution of PDF version numbers among the three PDF sets where considerable differences between the sets are visible. Among the doctoral dissertations, PDF versions 1.5, 1.6, and 1.7 are the most common versions, but out of those three versions, only 1.7 is base for a PDF/A standard part. Version 1.4, base for PDF/A-1, is used in only 5.9% of the files. For the PDF set of SOU submissions, PDF 1.7 has the largest proportion (39.4%), closely followed by PDF 1.6 (36.1%). PDF 1.4 has still a share of 23.2%. PDF files from the SFS set are clearly biased towards PDF 1.7, with only two files in a different version (1.6). Six PDF files were observed using PDF version 2.0: five files in the set of doctoral dissertations and one file from the SOU submissions.

PDF files may claim adherence to PDF/A, but such a claim does not imply conformance to the stated specification. Two conformance checking tools (veraPDF and 3-Heights PDF Validator Online Tool) were used to test every file’s claim of conformance. Table 2 [REMOVED REF FIELD] presents how many files from two of the PDF sets claim conformance to any of the PDF/A specifications and for how many files this claim could be confirmed by both tools. Data on files from the SFS set is not shown in [REMOVED REF FIELD] was no file of this set claimed conformance to the PDF/A standard.

For the doctoral dissertations, only 0.4% of the files claim to adhere to PDF/A and only 0.2% are conforming according to both conformance checking tools. Within this small collection of PDF files, PDF/A-1a is clearly the dominating specification. Those low numbers are in contrast to SOU submissions, where 40.0% of the files claim to adhere to PDF/A (dominating are PDF/A-3a and PDF/A-1b) and 29.0% are conforming to the claimed specifications according to both conformance checking tools. No files claiming conformance to PDF/A-4 were observed.

The main findings related to the first research question are as follows:

Table 1. Distribution of PDF versions across the three PDF sets by number of files

PDF Version	Doctoral Dissertations		SOU Submissions		SFS	
1.3	806	8.6%				
1.4	552	5.9%	36	23.2%		
1.5	2446	26.2%	1	0.6%		
1.6	3058	32.7%	56	36.1%	2	<0.1%
1.7	2474	26.5%	61	39.4%	5929	>99.9%
2.0	5	<0.1%	1	0.6%		
Total	9341		155		5931	

Note. Percent Values Refer to the total number Of PDF files per set.

Table 2. PDF files that claim conformance to some PDF/A Part/Level and PDF files for which conformance is affirmed by both conformance checking tools

PDF/A Part/Level	Doctoral Dissertations				SOU Submissions			
	Claims		Conforms		Claims		Conforms	
1b	3	<0.1%	1	<0.1%	26	16.8%	12	7.7%
1a	18	0.2%	11	0.1%	10	6.5%	8	5.2%
2b	8	<0.1%	3	<0.1%	2	1.3%	1	0.6%
2a								
2u	1	<0.1%	0	0.0%				
3b	4	<0.1%	3	<0.1%				
3a	5	<0.1%	4	<0.1%	24	15.5%	24	15.5%
3u								
With Claims	39	0.4%	22	0.2%	62	40.0%	45	29.0%
Total	9341				155			

1. The most popular PDF versions vary depending on the PDF set. Overall, common PDF versions are 1.4, 1.5, 1.6, and 1.7. Only six files out of 15,427 used PDF version 2.0
2. For doctoral dissertations, less than 0.4% of the files claim conformance to PDF/A. For SOU submissions, 40% of the files make such claims in their metadata. For SFS submissions, none of the files makes such claims.
3. Most PDF files which contain a claim for adherence to PDF/A in their metadata have this claim supported by the outcome of two conformance checking tools (67 out of 101 PDF files).
4. Preference for PDF/A specifications varies across PDF sets: PDF/A-1a for doctoral dissertations, PDF/A-1b and PDF/A-3a for SOU submissions, and no claims for PDF/A conformance for SFS publications.

FONT USAGE IN PDF FILES

To answer the second research question: How are different fonts used in the collected PDF files?, we analyzed which fonts were used in each PDF file and then used an heuristic approach to assess the license status for each font based on the font’s identified name.

Table 3 shows the number of PDF files that make use of the fonts that are in the top-10 for any of the three PDF sets. Two fonts, Times New Roman and Arial, dominate the font selection in each of the PDF sets (for SFS, only Times New Roman).

Those two fonts, as well as most other popular fonts, are categorized as proprietary, as their legal use requires acquiring dedicated licenses for the fonts or software systems which include those fonts. The font Garamond got classified as ambiguous, as variants in several license categories are available and the name alone does not allow to determine which variant was used. The most popular fonts under open licenses include Computer Modern, exclusively used in the PDF set of doctoral dissertations by PDF files, Open Sans used in seven files in the SOU set and 96 doctoral dissertations, Liberation Sans used in 57 doctoral dissertations and two SOU submissions, and Liberation Serif used in 42 doctoral dissertations and one SFS submission.

PDF files may make use of several fonts, where each may get separately classified into one of the four license categories introduced earlier. Figure 1 visualizes the number of files that exhibit possible combinations of font license categories. Figure 1a describes the distribution for doctoral dissertations. 1.2% of the PDF files make only use of fonts under open licenses and 22.7% of the

Table 3. The ten most popular fonts from each of the three PDF sets, merged and ordered by popularity among doctoral dissertations

Font name	License Category	Doctoral Dissertation		SOU Submission		SFS	
Times New Roman	Proprietary	8062	86.3%	103	66.5%	5931	100.0%
Arial	Proprietary	7789	83.4%	121	78.1%	13	0.2%
Calibri	Proprietary	5754	61.6%	52	33.5%	207	3.5%
Cambria	Proprietary	4091	43.8%	13	8.4%	7	0.1%
Symbol MT	Proprietary	3474	37.2%	13	8.4%	2	<0.1%
Helvetica	Proprietary	2850	30.5%	9	5.8%		
Symbol	Ambiguous	1615	17.3%				
Times	Ambiguous	1608	17.2%	11	7.1%		
Georgia	Proprietary	1531	16.4%	16	10.3%		
Computer Modern	Open	1501	16.1%				
Verdana	Proprietary	992	10.6%	13	8.4%	1	<0.1%
Garamond	Ambiguous	860	9.2%	21	13.5%	52	0.9%
Minion Pro	Proprietary	650	7.0%	12	7.7%		
Segoe UI	Proprietary	434	4.6%	4	2.6%	3	<0.1%
Microsoft Sans Serif	Proprietary	55	0.6%	20	12.9%		
Baskerville Old Face	Proprietary	47	0.5%			2	<0.1%
Liberation Serif	Open	42	0.4%			1	<0.1%
Constantia	Proprietary	33	0.4%			7	0.1%
<i>No Fonts Used</i>		15					

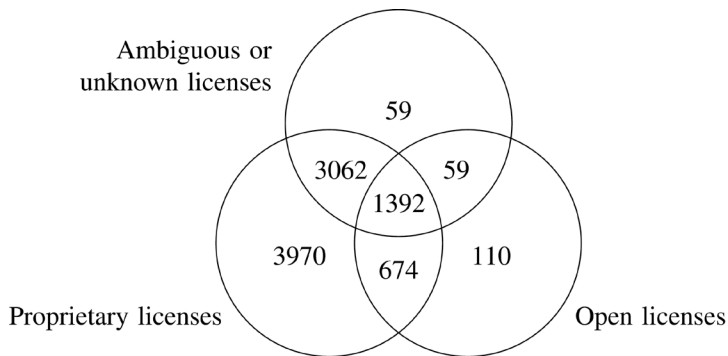
files combine fonts under open licenses with fonts under other licenses. In the SOU submissions (b) and the SFS set (c), most PDF files contain only proprietary fonts; for the SFS set more than 99% of the files use only fonts under such licenses. For the PDF set of doctoral dissertations, 15 files do not make use of any font, thus are outside of the diagram's circles. The single PDF file from the SOU submission set that uses only fonts under open licenses comes from the municipality of Katrineholm using the fonts Playfair Display and Open Sans. Out of the other 13 files that use both fonts under open licenses and under non-open licenses, four are from municipalities, and the remaining nine are from national government agencies. For the SFS set, the one file that makes use of a font under an open license is regulation SFS 2018:384. However, despite this font been referred to in the PDF file, it does not seem to be used for any visible text.

Table 4 puts PDF/A conformance in relation to used fonts' license categories. As discussed earlier, the majority of PDF files across all three PDF sets do not claim conformance to the PDF/A standard. Among those files that conform to PDF/A, none makes use of only open fonts. Among the 2,139 files that combine fonts under open licenses with fonts under non-open licenses, only 10 (0.5%) conform to PDF/A. Among the 13,162 files that do not make use of fonts under open licenses, only 57 (0.4%) conform to PDF/A.

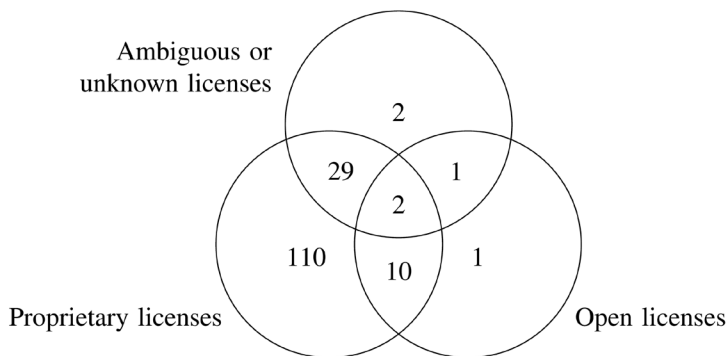
What has been previously referred to as font was often a family of fonts, where family members vary; for example, bold or italic. PDF writers when creating PDF files will only make use of those font family members that are used in the original document. Further, when creating font subsets, only the font data necessary to render the document's text may be embedded, but not the font data for unused characters. It is technically possible that one font variation may be the base for several

Figure 1. Combinations of the four font license categories (“Ambiguous” and “Unknown” merged) and the number of PDF files for each combination of font licenses in PDF files

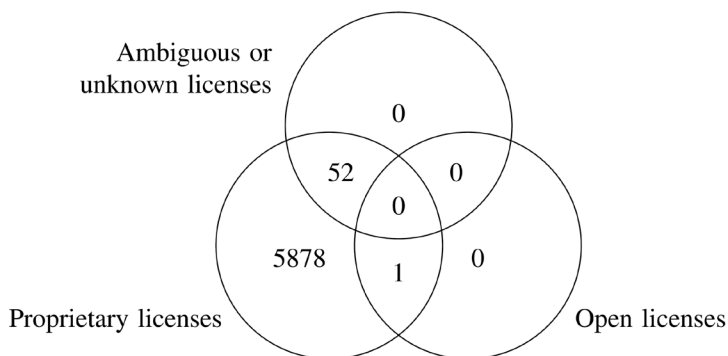
Note. One Venn Diagram Is Shown For Each PDF Set.



(a) Doctoral dissertations, 9341 documents



(b) SOU submissions, 155 documents



(c) SFS, 5931 documents

separate font subsets in a PDF file. Indeed, the specification for PDF 2.0 recommends (ISO, 2020b) to treat subsets originally based on the same font as independent entities. For the following discussion, we follow this recommendation counting each subset individually.

Table 5 presents the occurrences of different types of embedding of fonts as claimed by the PDF file. For the most popular fonts families in each of the three PDF sets, columns count the number of references across the chosen PDF set and how many of those font references are not embedded at

Table 4. Comparison of conformance to a PDF/a part as assessed by VeraPDF and 3-heights pdf validator versus font license categories

Used Fonts' Licenses	PDF/A-1		PDF/A-2 or A-3		Not PDF/A	
Doctoral Dissertations, 9341 documents						
No Fonts Used	0		0		15	0.2%
Only Open Licenses	0		0		110	1.2%
Both Open and Non-Open Licenses	1	<0.1%	3	<0.1%	2121	22.7%
Only Non-Open Licenses	11	0.1%	7	<0.1%	7073	75.7%
SOU submissions, 155 documents						
No Fonts Used	0		0		0	
Only Open Licenses	0		0		1	0.6%
Both Open and Non-Open Licenses	1	0.6%	5	3.2%	7	4.5%
Only Non-Open Licenses	19	12.3%	20	12.9%	102	65.8%
SFS, 5931 documents						
No Fonts Used	0		0		0	
Only Open Licenses	0		0		0	
Both Open and Non-Open Licenses	0		0		1	<0.1%
Only Non-Open Licenses	0		0		5930	>99.9%

all, only as a subset, or if a referenced font is fully embedded, respectively. The right-most column marked with “â” is discussed further below.

For the PDF sets of doctoral dissertations and SOU submissions, variants of font families Times New Roman, Arial, and Helvetica (the latter one only for doctoral dissertations) are more likely to be not embedded (i.e., missing font data) compared with other fonts. For example, in the PDF set of doctoral dissertations, variants of Times New Roman are referred to but not embedded in PDF files in 5,159 cases (2,394 PDF files) and subset embedded in 58,483 cases (ratio 1:11.3, 6,489 files), whereas for Calibri in only 12 cases (eight files) the font is referred to but not embedded and subset embedded in 19,226 cases (ratio 1:1,602, 5,708 files). This special pattern for Times New Roman could not be observed for PDF files from the SFS set where the ratio is 1:3,063, similar to what was observed for Calibri in the doctoral dissertation PDF set.

The numbers for font references that are listed as “fully embedded” in Table 5 are based on the PDF files’ data. However, the numbers of fonts that are actually fully embedded is considerably lower. First, we observed cases where font references claimed to be fully embedded had six-letter tags attached to their font name that are meant to mark an embedded font to be a subset only (Adobe, 2006). Second, we observed fonts that were fully embedded, but did not contain a representation (glyph) for the letter “A,” which we deem as unlikely for a fully embedded, non-symbol, Western font. The occurrence of font references that claim to be fully embedded but still match either of above two criteria are counted in the right-most column in Table 5 marked with “â”. For example, for the previously mentioned font family Times New Roman used in doctoral dissertations, 1,400 occurrences of this font family’s variants out of 1,677 occurrences contain the special tag in their names or lack a glyph for the letter “A”. Those occurrences correspond to 563 out of 775 PDF files.

Table 6 lists the number of instances of embedded font data which were of type OpenType and for which the values of the field fsType (Microsoft, 2021b) could be extracted. Across all three PDF sets, only a single PDF file contained a single font that had the flag for “restricted license embedding” set. The two dominating alternatives for flags set by fonts in the doctoral dissertations PDF set were

“editable embedding” (EE) and “installable embedding” (IE). For some fonts such as Times New Roman, the number of cases for each alternative was similar (25,589 vs 29,397), for other fonts, like Computer Modern, are very different (6 vs 25,158). For the other two PDF sets, only the alternative of “editable embedding” was clearly dominating.

Table 5. Number of font references for the most popular fonts in the three PDF file sets for different variants

Font	# Files	Number of font references per embedded status			
		Not Embedded	Embedded as Subset	Claims to be Fully Embedded	*
Doctoral Dissertations, 9341 Documents					
Times New Roman	8062	6609	58483	1677	1400
Arial	7789	5159	33657	896	699
Calibri	5754	12	19226	239	235
Cambria	4091	4	16225	178	167
Symbol MT	3474	1	4085	1078	1078
Helvetica	2850	2532	10218	175	108
SOU Submissions, 155 Documents					
Arial	121	68	143	93	92
Times New Roman	103	65	85	207	207
Calibri	52	0	73	34	34
Garamond	21	4	35	0	0
Microsoft Sans Serif	20	0	0	41	41
Georgia	16	0	44	0	0
SFS, 5931 Documents					
Times New Roman	5931	10	30634	608	33
Calibri	207	0	213	0	0
Garamond	52	0	52	0	0
Arial	13	0	13	0	0
Constantia	7	0	9	0	0
Cambria	7	0	8	0	0

The last alternative, “preview & print embedding,” occurred only in 4,570 doctoral dissertations (12,690 cases of font embedding) and not at all for the SOU and SFS sets.

Subsetting is common when embedding font data. This is mirrored in the “no subsetting” flag, which was only observed for 32 cases of font embedding across 29 different PDF files, most often for font Wingdings (24 cases). Contrary to the flag’s requirement, the font data was still embedded as a subset. The flag requiring embedding only bitmap fonts was not observed at all. The main findings related to the second research question are as follows.

1. The most often used fonts are Times New Roman and Arial, which are both proprietary like most of popular fonts. Among doctoral dissertations, only about 1.2% make use of only open fonts, but 22.7% combine fonts under open licenses with fonts under other types of licenses. In the set

Table 6. The most popular fonts in every PDF file set and how many font data embeddings have certain flags in the font data's fsType field set

Font	# Files	fsType Flag Occurrences Per Font Variant			
		RLE	EE	IE	PPE
Doctoral Dissertations, 9341 Documents					
Times New Roman	8062	0	25589	29397	93
Arial	7789	0	10668	24756	25
Calibri	5754	0	10775	6607	1
Cambria	4091	0	4874	8423	1
Symbol MT	3474	0	2334	2190	0
Helvetica	2850	0	932	9209	72
Computer Modern	1501	0	6	25158	20
All Cases of Embedding		3	73407	192099	12690
SOU Submissions, 155 Documents					
Arial	121	0	153	59	0
Times New Roman	103	0	94	60	0
Calibri	52	0	73	0	0
Garamond	21	0	35	4	0
Microsoft Sans Serif	20	0	0	0	0
Georgia	16	0	44	0	0
All Cases of Embedding		0	475	177	0
SFS, 5931 Documents					
Times New Roman	5931	0	31225	10	0
Calibri	207	0	213	0	0
Garamond	52	0	52	0	0
Arial	13	0	13	0	0
Constantia	7	0	9	0	0
Cambria	7	0	8	0	0
All Cases of Embedding		0	31530	11	0

of SOU submissions, only a single file was using only fonts under open licenses, but about 8% of the other files combined fonts under open licenses with fonts under other types of licenses. Among the SFS files, one file combined open font licenses with other types of licenses.

2. No PDF file combines claiming conformance (or even adhering to it) to the PDF/A standard with the use of fonts only under open licenses, 10 PDF files (out of 2139, 0.5%) achieve conformance with a mixture of open and non-open fonts, and 57 (out of 13,162, 0.4%) achieve conformance with non-open fonts only.
3. Embedding subsets of font programs is the preferred choice for making use of font data. Not embedding font data is most often chosen for fonts seen as ubiquitously available like Times New Roman or Arial.
4. Embedded fonts' metadata ("fsType") allows embedding of said font data in the vast majority of cases.

DISCUSSION

PDF files' claims to conform to PDF/A and the conformance as assessed by two checking tools vary greatly between the different sets of PDF files. Legal texts from the SFS set do not claim any conformance, and only a very limited subset of doctoral dissertation files claim conformance and even fewer achieve conformance, but among the submissions to SOU 2021:1 and SOU 2021:97, almost one third of the files achieve conformance. The results for the latter set of PDF files demonstrate that it is indeed possible to achieve PDF/A conformance in the context of PSOs. A possible explanation for the low adoption of PDF/A among both doctoral dissertations and legal texts may be that the paper copy is seen as more important and eventually archived, whereas the PDF version is seen as ephemeral only. Our observation of the low adoption among doctoral dissertations is in line with previous findings by Fischer et al. (2021). Universities failing to comply with PDF-related standards was also observed by Kopel (2022) where veraPDF was used to assess PDF files' compliance to PDF/UA (ISO, 2014). Only about 15.6% out of 24,366 files did not violate any rules related to PDF/UA.

The Swedish National Archives mandate the use of PDF/A-1, not any later part (Riksarkivet, 2009). Findings in this study suggest that this requirement is not generally considered in Swedish PSOs, as more than half of all files in the SOU set that conform to PDF/A do not conform to PDF/A-1. A possible explanation is that the choice which PDF/A specification a PDF adheres to is essentially determined by the used PDF-generating tool.

For most PDF files, only fonts under non-open licenses are used. An exception is doctoral dissertations, where 23% of the files combine fonts under open licenses with fonts under other licenses. Doctoral dissertations also use more often fonts classified as ambiguous or unknown than the other two PDF sets. Possible reasons include that dissertations are more diverse in their origin and make use of special fonts such as symbol and mathematical fonts or organizations' own fonts.

Embedding font data as subsets of the original font program (i.e., only font data necessary to render the document's text) is the dominating alternative of including font programs, but there is still a substantial body of PDF files that lack font data for fonts that are assumed to be ubiquitously available. This negates the ability to visually reproduce a PDF file if the PDF reader lacks those fonts.

Based on the four flags of font usage permissions found in OpenType font data, a substantial number of PDF files included font data where the "preview and print embedding" restriction was set, which may hinder further modification of those PDF files—for example, when preparing for long-term archival. This restriction is closely related to the choice of using fonts under proprietary licenses; fonts under open licenses cannot have such restrictions.

We acknowledge inherent limitations in the transferability of our findings on the three investigated PDF sets to PDF sets acquired from other types of organizations: PSOs are not a homogeneous group as our data shows, which is documented in the large spread of PDF files' properties like conformance to the PDF/A standard.

CONCLUSION

Regarding our first research question: "To what extent do PDF files from public sector organizations conform to the PDF/A standard and what characterizes those files?" findings show that conformance greatly varies across organizations. Whereas universities achieve only a very limited adoption of PDF/A, other PSOs at different levels of government achieve better results.

The differences in the adoption of PDF/A between universities and general PSOs such as national government agencies is surprising, for both types of organizations operate within similar legal frameworks and have comparable technical and financial capabilities to, for example, acquire and setup tools for the generation of PDF files. The general PSOs' adoption of PDF/A suggests that the challenges are not only technical, but also organizational. Starting with the procurement of PDF-related hardware and software, organizations must adjust their PDF workflows, raise awareness

among their employees (and students, in the case of universities), and provide incentives and support for the adoption of PDF/A.

With regard to our second research question: How are different fonts used in the collected PDF files?” findings from our study show that most dominating are proprietary fonts. This is problematic as it is not possible, based only on a given PDF file, to determine if the fonts’ usage was legal. Another problem is when PDF files do not include font data for fonts, they make use of, assuming that the standard fonts are ubiquitously available. However, without the necessary font data, it is impossible to visually reproduce the PDF file as it was intended. Both problems may prevent further processing of the PDF file, for example in the context of long-term maintenance, reuse, and archiving of files. Only fonts under open licenses allow for the unrestricted redistribution and archiving as required by the PDF/A standard. Organizations must make an active choice in providing open fonts and adjusted templates to their users. The use of fonts under open licenses can be part of a migration to open source software and avoids limitations for font distribution in heterogeneous IT environments (different operating systems, BYOD).

Findings suggest that future research may motivate an investigation of additional sets of PDF files to have a broader base to perform analyses on. The effects of attempts to convert PDF files to PDF/A may get investigated in line with Lehtonen et al. (2018). Alternatively, a complementary study can investigate the reasons why some organizations do not strive for PDF/A conformance whereas other (types of) organizations achieve this to a considerably higher degree.

FUNDING INFORMATION

This research has been financially supported by the Swedish Knowledge Foundation (KK-stiftelsen) and participating partner organizations in the SUDO project.

ACKNOWLEDGEMENTS

The authors are grateful for the stimulating collaboration and support from colleagues and partner organizations.

REFERENCES

- Adobe. (2001). *PDF reference: Adobe portable document format version 1.4* (3rd ed.). Adobe. <https://opensource.adobe.com/dc-acrobat-sdk-docs/pdfstandards/pdfreference1.4.pdf>
- Adobe. (2006). *PDF reference: Adobe portable document format version 1.7* (6th ed.). Adobe. https://www.adobe.com/content/dam/acom/en/devnet/pdf/pdf_reference_archive/pdf_reference_1-7.pdf
- Adobe. (2020). *Adobe service additional terms*. Adobe. <https://www.adobe.com/go/adobe-fonts-terms>
- Adobe. (2023). *Font licensing*. Adobe. <https://helpx.adobe.com/fonts/using/font-licensing.html>
- Bean, J. (2015). The times new roman lie. *Interactions (New York, N.Y.)*, 22(3), 24–25. doi:10.1145/2742938
- Bundesarchiv. (2010). *Aussonderung digitaler unterlagen und deren archivierung im bundesarchiv—ein leitfaden* [Version 1.2, in German].
- Casario, M., Wormser, N., Saltzman, D., Bradford, A., Reid, J., Improta, F., & Congleton, A. (2012). *Fonts, text, and color. CSS3 Solutions: Essential Techniques for CSS3 Developers*. Apress. doi:10.1007/978-1-4302-4336-6
- Crossland, D. (2012). Moral rights and the SIL open font license. *Libre Graphics Magazine*, 1(4), 12–13. https://libregraphicsmag.com/files/libregraphicsmag_1.4_highquality.pdf
- Egyedi, T. M. (2007). Standard-compliant, but incompatible?! *Computer Standards & Interfaces*, 29(6), 605–613. doi:10.1016/j.csi.2007.04.001
- Evans, E. N. (2014). Fonts, typefaces, and IP protection: Getting to just right. *Journal of Intellectual Property Law*, 21(2). <https://digitalcommons.law.uga.edu/jipl/vol21/iss2/4>
- Faroult, S. (2016). *Getting the message across: Using slideware effectively in technical presentations*. Apress., doi:10.1007/978-1-4842-2295-9
- Ferro, N., Silvello, G., Buelinckx, E., Doubrov, B., Fresa, A., Geber, M., Jadehlans, K., Justrell, B., Lemmens, B., Martinez, J., Munoz, V., Oliveras, S., Prandoni, C., Rice, D., Rohde-Enslin, S., Tarrés, X., Verbruggen, E., Yousefi, B., & Wilson, C. (2018). Evaluation of Conformance Checkers for Long-Term Preservation of Multimedia Documents. In J. Chen, M. A. Gonçalves, J. M. Allen, E. A. Fox, M.-Y. Kan, & V. Petras (Eds.). *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries* (pp. 145–154). Association for Computing Machinery. <https://doi.org/doi:10.1145/3197026.3197037>
- Fischer, T., Lundell, B., & Gamalielsson, J. (2021). Achieving conformance to document standards: Can PDF files conform to the PDF/A-1b specification? [IJSR]. *International Journal of Standardization Research*, 19(1), 1–32. doi:10.4018/IJSR.288523
- Groves, P. J. (1992). Copyright in digitized typograph. *Computer Law & Security Report*, 8(3), 125–129. doi:10.1016/0267-3649(92)90054-D
- Infrastrukturdepartementet. (2021a). *It-driftsutredningens delbetänkande Säker och kostnadseffektiv it-drift—Rättsliga förutsättningar för utkontraktering*, Infrastrukturdepartementet. <https://www.regeringen.se/remisser/2021/02/remiss-sou-20211-it-driftsutredningens-delbetankande-saker-och-kostnadseffektiv-it-drift--rattsliga-forutsattningar-for-utkontraktering>
- Infrastrukturdepartementet. (2021b). *Remiss av It-driftsutredningens slutbetänkande Säker och kostnadseffektiv it-drift—förslag till varaktiga former för samordnad statlig it-drift*. Infrastrukturdepartementet. <https://www.regeringen.se/remisser/2021/12/remiss-av-it-driftsutredningens-slutbetankande-saker-och-kostnadseffektiv-it-drift--forslag-till-varaktiga-former-for-samordnad-statlig-it-drift/>
- ISO. (2005). *Document management – Electronic document file format for long-term preservation—Part 1: Use of PDF 1.4 (PDF/A-1)*. ISO.
- ISO. (2008). *Document management—Portable document format – Part 1: PDF 1.7*. ISO 32000-1:2008.
- ISO. (2011). *Document management—Electronic document file format for long-term preservation – Part 2: Use of ISO 32000-1 (PDF/A-2)*. ISO.

- ISO. (2012). *Document management – Electronic document file format for long-term preservation – Part 3: Use of ISO 32000-1 with support for embedded files (PDF/A-3)*. ISO.
- ISO. (2014). *Document management applications – Electronic document file format enhancement for accessibility – Part 1: Use of ISO 32000-1 (PDF/UA-1)*. ISO.
- ISO. (2017). *Document management – Portable document format – Part 2: PDF 2.0*. ISO 32000-2:2017.
- ISO. (2020a). *Document management – Electronic document file format for long-term preservation – Part 4: Use of ISO 32000-2 (PDF/A-4)*. ISO 19005-4:2020.
- ISO. (2020b). *Document management – Portable document format – Part 2: PDF 2.0*. ISO.
- iText. (2020). *Why can't I embed a font due to licensing restrictions?* iText. <https://kb.itextpdf.com/home/it7kb/ebooks/itext-7-converting-html-to-pdf-with-pdfhtml/chapter-7-frequently-asked-questions-about-pdfhtml/why-can-t-i-embed-a-font-due-to-licensing-restrictions>
- Klindt, M. (2017). PDF/A considered harmful for digital preservation. In *Proceedings of iPres – 14th International Conference on Digital Preservation*. C Seas. <https://www-archive.cseas.kyoto-u.ac.jp/ipres2017.jp/wp-content/uploads/15.pdf>
- Koo, J., & Chou, C. C. H. (2013). PDF to PDF/A: Evaluation of converter software for implementation in digital repository workflow. *New Review of Information Networking*, 18(1), 1–15. doi:10.1080/13614576.2013.771989
- Kopel, M. (2022). Towards making university pdfs conform to universal accessibility standard. In N. T. Nguyen, Y Manolopoulos, R., Chbeir, A. Kozierkiewicz, & B. Trawiński (Eds), *Computational Collective Intelligence, ICCCI 2022*. Springer, Cham. doi:10.1007/978-3-031-16014-1_57
- Lehtonen, J., Helin, H., Kylander, J., & Koivunen, K. (2018). PDF mayhem: Is broken really broken. In *Proceedings of iPres–15th International Conference on Digital Preservation*. iPres. <https://ipres2018.org/proceedings>
- LG München I. (2022). Endurteil vom 20.01.2022, 3 O 17493/20 (REWIS RS 2022, 1892).
- Library and Archives Canada. (2015). *Guidelines on file formats for transferring information resources of enduring value*. LAC. <https://www.bac-lac.gc.ca/eng/services/government-information-resources/guidelines/Documents/file-formats-irev.pdf>
- Lundell, B., Gamaliellson, J., & Katz, A. (2023). Implementing the HEVC standard in software: Challenges and recommendations for organisations planning development and deployment of software. *Journal of Standardisation*, 2. Advance online publication. doi:10.18757/jos.2022.6695
- Lundell, B., Gamaliellson, J., & Katz, A. (2015). On implementation of open standards in software: To what extent can ISO standards be implemented in open source software? [IJSR]. *International Journal of Standardization Research*, 13(1), 47–73. doi:10.4018/IJSR.2015010103
- Lundell, B., Gamaliellson, J., & Katz, A. (2019). Implementing IT standards in software: Challenges and recommendations for organizations planning software development covering IT Standards. *European Journal of Law and Technology*, 10(2). <https://ejlt.org/index.php/ejlt/article/view/709>
- Lundell, B., Gamaliellson, J., Katz, A., & Lindroth, M. (2022). Use of commercial SaaS solutions in Swedish public sector organizations under unknown contract terms. In M. Janssen et al. (Eds.), *Electronic Government* (pp. 73–92). Springer., doi:10.1007/978-3-031-15086-9_6
- Microsoft. (2021a). *Font redistribution FAQ*. Microsoft. <https://docs.microsoft.com/en-us/typography/fonts/font-faq>
- Microsoft. (2021b). *OS/2 and Windows metrics table*. Microsoft. <https://docs.microsoft.com/en-us/typography/opentype/spec/os2>
- Microsoft. (2023). *Microsoft license terms, supplement to Microsoft services agreement, office service and software in Microsoft 365 consumer subscription*. Microsoft. https://www.microsoft.com/en-us/Useterms/Retail/OfficeinMicrosoft365/Personal/Useterms_Retail_OfficeinMicrosoft365_Personal_English.htm

- PDF Tools AG. (2022). *PDF tools online – Validate PDF*. PDF Online. <https://www.pdf-online.com/osa/validate.aspx>
- Perry, T. S. (1988). PostScript prints anything: A case history. *IEEE Spectrum*, 25(5), 42–46. doi:10.1109/6.4550
- Regeringskansliet. (1993). *Lag (1993:1392) om pliktexemplar av dokument*. *Svensk författningssamling*. <https://rkrattsbaser.gov.se/sfst?bet=1993:1392>
- Regeringskansliet. (2012). *Lag (2012:492) om pliktexemplar av elektroniskt material*. *Svensk författningssamling*. <https://rkrattsbaser.gov.se/sfst?bet=2012:492>
- Regeringskansliet. (2021a). *Säker och kostnadseffektiv it-drift–förslag till varaktiga former för samordnad statlig it-drift*. *SOU 2021:97*. <https://www.regeringen.se/rattsliga-dokument/statens-offentliga-utredningar/2021/12/sou-202197>
- Regeringskansliet. (2021b). *Säker och kostnadseffektiv it-drift – rättsliga förutsättningar för utkontraktering, Delbetänkande av It-driftsutredningen, SOU 2021*. <https://www.regeringen.se/rattsliga-dokument/statens-offentliga-utredningar/2021/01/sou-20211>
- Regeringskansliet. (2023). *Svensk författningssamling*. <https://www.svenskförfattningssamling.se>
- Riksarkivet. (2009). *Riksarkivets föreskrifter och allmänna råd om tekniska krav för elektroniska handlingar (upptagningar för automatiserad behandling)*. Riksarkivets författningssamling, RA-FS 2009:2. <https://riksarkivet.se/rafs?pdf=rafs/RA-FS%202009-02.pdf>
- Rog, J. (2007). *PDF Guidelines–Recommendations for the creation of PDF files for long-term preservation and access*. Koninklijke Bibliotheek/National Library of the Netherlands.
- Royster, P. (2011). Word 2010 Font embedding issues resolved. *Digital Commons/Institutional Repository Information*, 70. https://digitalcommons.unl.edu/ir_information/70/
- Spalinger, N., & Gaultney, V. (2007). *SIL open font license (OFL)*. <https://scripts.sil.org/ofl>
- Statistics Sweden. (2023). *Myndighetsregistret*. <http://www.myndighetsregistret.scb.se>
- Stenbeck, M., Fält, S. E., & Reichel, J. (2021). Swedish law on personal data in biobank research: Permissible but complex. In S. Slokenberga, O. Tzortzatou, & J. Reichel (Eds.), *GDPR and biobanking: Individual rights, public interest and research regulation across Europe* (pp. 379–394). Springer International Publishing., doi:10.1007/978-3-030-49388-2_21
- Sullivan, S. J. (2006). An archival/records management perspective on PDF/A. *Records Management Journal*, 16(1), 51–56. doi:10.1108/09565690610654783
- Suri, R. E., & El-Saad, M. (2021). Lost in migration: Document quality for batch conversion to PDF/A. *Library Hi Tech*, 39(2), 337–351. doi:10.1108/LHT-10-2017-0220
- Thatcher, M. E., & Pingry, D. E. (2007). [Software patents] The good, the bad, and the messy. *Communications of the ACM*, 50(10), 47–52. doi:10.1145/1290958.1290961
- Williams, G. (2012). *CSS3 for Mobile. Learn HTML5 and JavaScript for Android*. Apress., doi:10.1007/978-1-4302-4348-9
- Wilson, C., McGuinness, R., & Jung, J. (2017). VeraPDF: Building an open source, industry supported PDF/A validator for cultural heritage institutions. *Digital Library Perspectives*, 33(2), 156–165. doi:10.1108/DLP-08-2016-0031
- World Intellectual Property Organization. (2017). *Compilation of the replies to the questionnaire on graphical user interface (GUI), icon and typeface/type font designs*. Standing Committee on the Law of Trademarks, Industrial Designs and Geographical Indications, Thirty-Sixth Session, SCT/36/2, revision 2. https://www.wipo.int/edocs/mdocs/geoind/en/sct_36/sct_36_2_rev_2.pdf

ENDNOTES

¹ <https://www.iso.org>

² <https://www.oasis-open.org>

³ A font can be embedded completely (ISO, 2020, p. 351) or only in parts (“subset”, ISO, 2020, p. 355)

⁴ The investigation started in September 2019 and its initial report (sv: delbetänkande) was presented in January 2021 as SOU 2021:1 (Regeringskansliet, 2021b). Thereafter, 120 stakeholders were asked to submit comments (Infrastrukturdepartementet, 2021a). Stakeholders included both PSOs and private actors. Before the deadline in May 2021, 99 stakeholders answered: 86 PSOs (64 national, 3 regional, and 19 municipal) and 13 private actors. In December 2021, the final report (sv: slutbetänkande) designated as SOU 2021:97 was published (Regeringskansliet, 2021a) and this time 99 stakeholders (a different set of stakeholders than the 99 who answered to SOU 2021:1) were asked to submit comments (Infrastrukturdepartementet, 2021b). Before the deadline in March 2022, 85 stakeholders answered: 69 PSOs (exclusively national) and 16 private actors.

Thomas Fischer majored in computer science at the Technical University of Darmstadt in 2003 and received a PhD from the Technical University of Kaiserslautern in 2008. He is a senior lecturer at the University of Skövde and is a member of the Software Systems Research Group. His research interests include open source and open standards, in particular file formats, lock-in, and interoperability. Teaching interests center around programming, system administration including virtualization, sustainability, and cloud computing. Complementing his research interests, he is also an active member in a number of open source and open data communities.

Björn Lundell received a PhD from the University of Exeter in 2001. He is a professor at the University of Skövde where he leads the Software Systems Research Group, and has been a staff member and researcher since 1984. Professor Lundell’s research contributes to theory and practice in the software systems domain and centres on different aspects of openness (in particular open source and open standards) related to development, use, and procurement of software systems. His research addresses fundamental socio-technical challenges concerning software systems, and focuses on different aspects of lock-in, interoperability, and longevity of systems. His research is reported in over 100 papers in a variety of international journals and conferences. Professor Lundell has been active in a number of international and national research projects which have lead to significant scientific and societal impact, and has been an expert advisor and contributed to guidelines and policies in the field.

Jonas Gamalielsson is a researcher at the University of Skövde, Sweden. He has worked with software development in industry from 1990 to 2001, and has thereafter (since 2001) been involved in teaching and research at the University of Skövde. He received his Ph.D. in computer science from Heriot-Watt University (Edinburgh) in 2009. He is a member of the Software Systems Research Group at University of Skövde and has conducted research related to open source software in a number of national and international projects since 2008. Focus has been on challenges related to lock-in, interoperability and long-term maintenance of software systems in various contexts. His research is reported in publications in a variety of international journals and conferences.