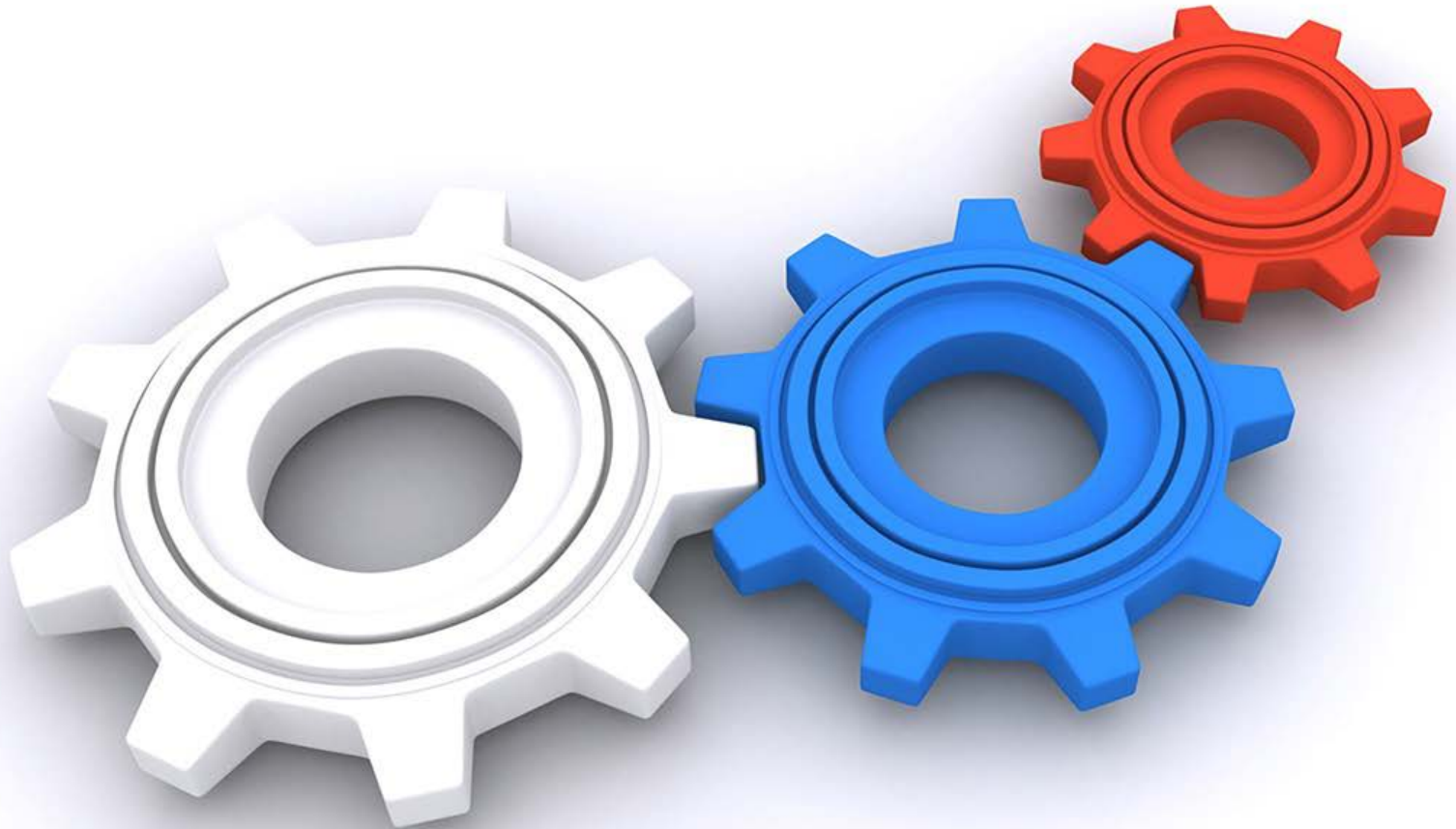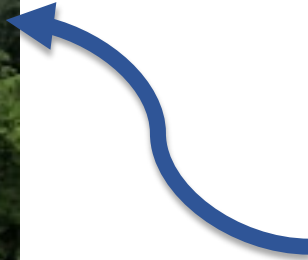# Pitfalls of Machine-Generated PDFs

Still a work in progress
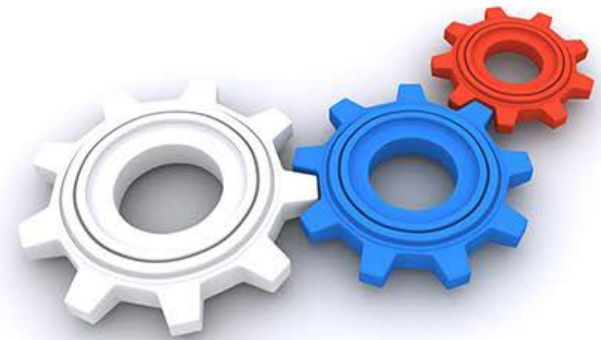
The Presenter

Richard U. Steinberg
rsteinberg@austinchronicle.com
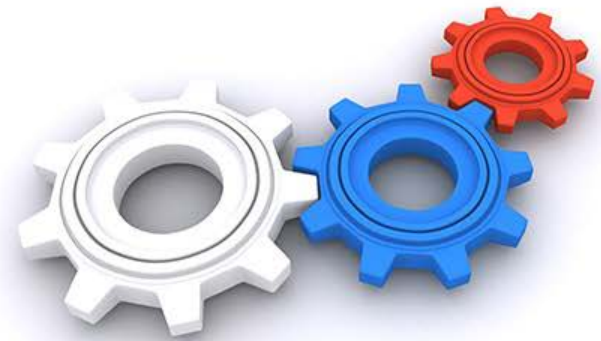richard.steinberg@wellsfargo.com

Outline

- Overview of PDFs

- What are PDF tags and their importance

- Converting HTML to PDF
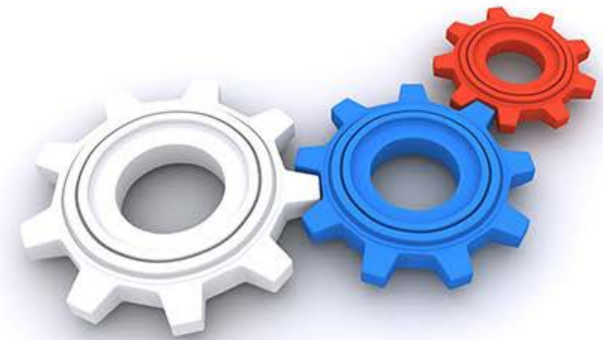
Overview of PDFs

- History (including my history)
- History with assistive technology users
- Different ISO standards
- The future of PDFs (?)

"State of the Art" Web 1997

Children Available for ADOPTION

Girls Available for Adoption
(Last updated March 31, 1997)

Boys Available for Adoption
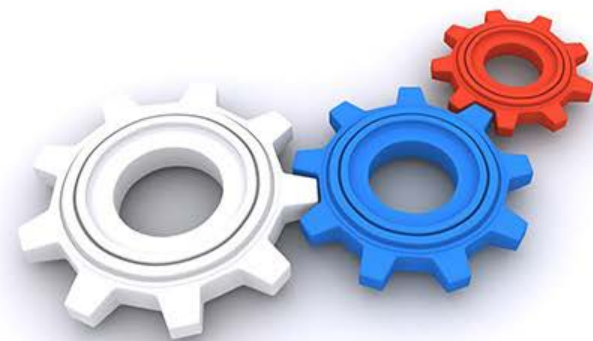(Last updated April 3, 1997)

Brothers and Sisters Available for Adoption
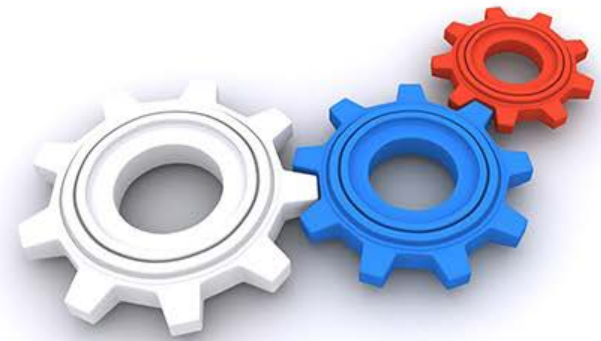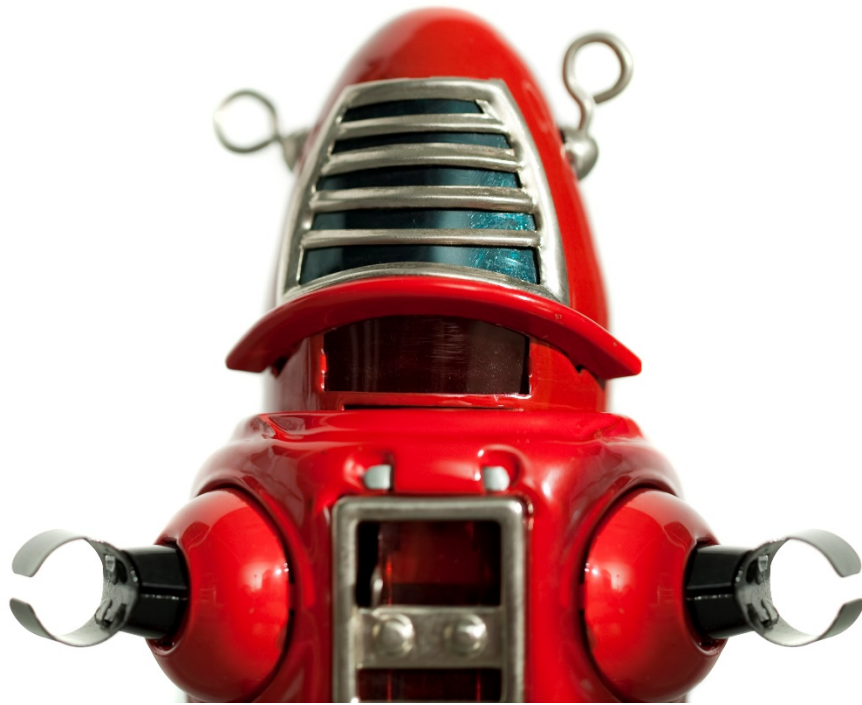(Last updated April 10, 1997)

General Information

You may E-mail your inquiries and questions to the
Texas Adoption Resource Exchange (TARE) Coordinator
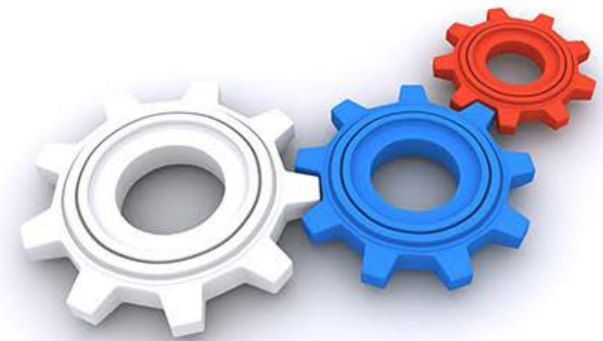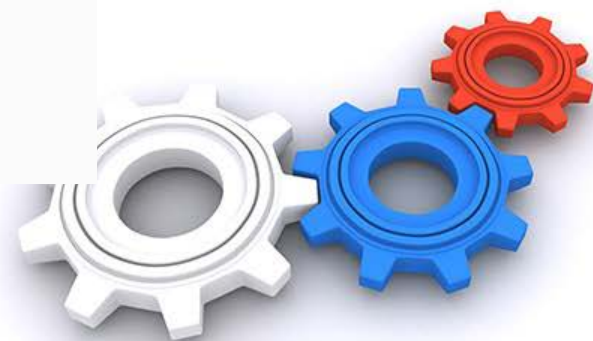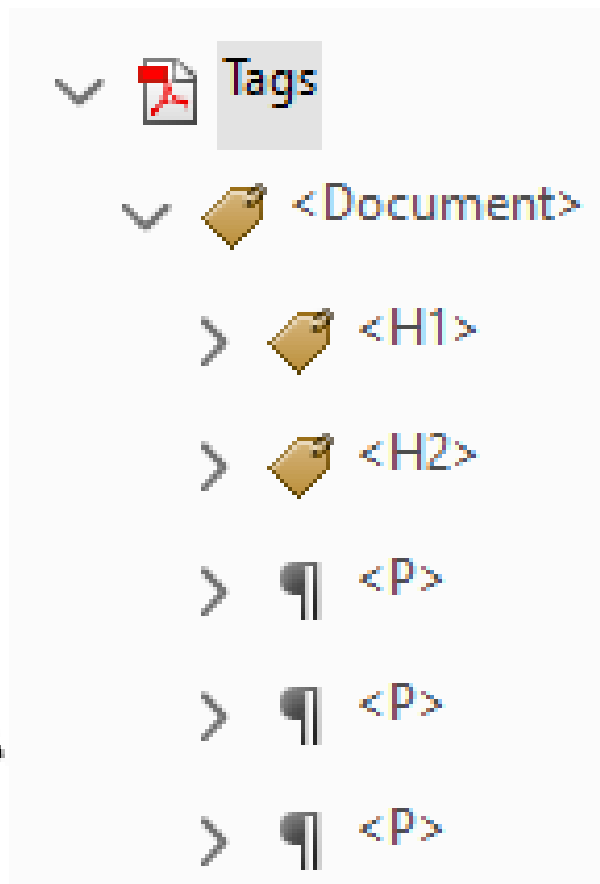tovarhg@auste654c.aust.tdprs.state.tx.us

The Evolution of PDFs

The PDF Association

Reputation of PDFs

PDFs in Academia

Proceedings of the 53rd Hawaii International Conference on System Sciences | 2020

PDF Accessibility of Research Papers:
What Tools are Needed for Assessment and Remediation?

Aravind Jembu Rajkumar[1], Jonathan Lazar[1], J. Bern Jordan[1], Alireza Darvishy[2], Hans-Peter Hutter[2]
[1]Trace Center, HCIL, College of Information Studies, University of Maryland, USA
[2] ICT-Accessibility Lab, Zurich University of Applied Sciences, Switzerland
(aravind@umd.edu; jlazar@umd.edu; jbjordan@umd.edu; dvya@zhaw.ch; huhp@zhaw.ch)

Abstract

Trillions of documents online are in PDF format, but only a small amount of these PDF documents include the necessary markup to make them accessible for people with disabilities. This paper presents the results of three related data collection efforts: a survey (with 61 participants), interviews (with 6 participants), and usability testing (with 6 participants), to learn more about what tools are needed for content contributors, to assist them in the assessment and remediation of accessibility in PDF documents. The paper provides suggested features and usability needed for software tools to support PDF document accessibility, as well as implications for content creators, scientific publishers, as well as the creator of the PDF format, Adobe.

1. Introduction

The topic of web accessibility has received much attention in the news recently, due to the large number of lawsuits on the topic [16]. A related topic, which has received much less attention, is the accessibility of PDF files. While many scholars, researchers and practitioners are aware of Web Content Accessibility Guidelines WCAG 2.1, few are aware of the PDF accessibility

information in their research paper templates and have accessibility chairs for conferences. Many technology companies are increasing the availability of accessibility features in their tools [9], including those involved in STEM education. A recent report states that about 10% of employed scientists and engineers report one or more disabilities, including limitations in hearing, vision, cognitive ability, or independent living [8]. For all of these reasons, this paper will focus on understanding the needs of content creators in STEM fields, related to software tools for making their PDF documents more accessible for people with disabilities.

2. Previous work

2.1 About PDF

PDF was first created for printing purposes by Adobe in the early 1990s. PDF is not really designed to be a content presentation tool. It is a graphical layout model, analogous to using CSS but without the HTML semantic markup. There are over 2.5 trillion PDF documents online [11], and we do not know how many of them are accessible. According to the World Health Organization, there are 36 million people who are blind and 217 million people who have moderate to severe
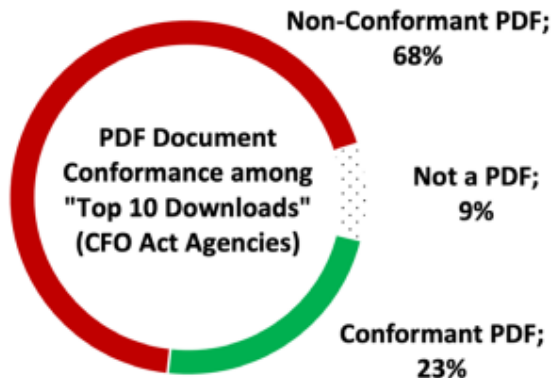
- [Link to paper brief](#)

- [Full text PDF](#) (ironically, the PDF fails Acrobat's accessibility checker)
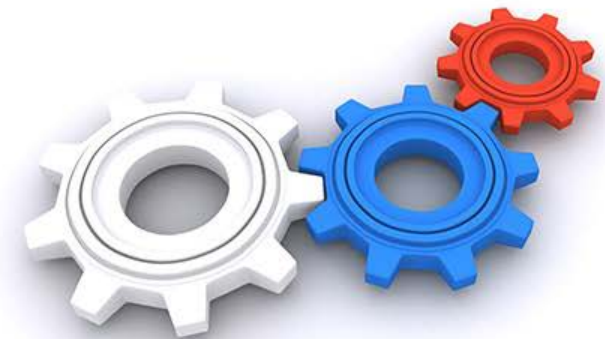
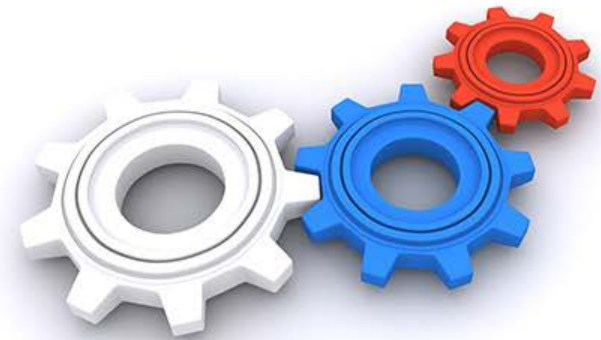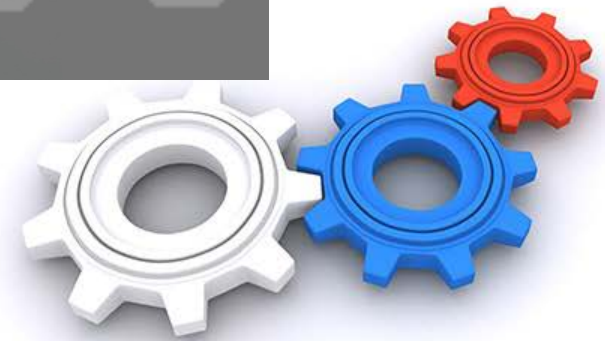# PDFs in Government

January 2023, "Section 508 Report to the President and Congress," U.S. Department of Justice
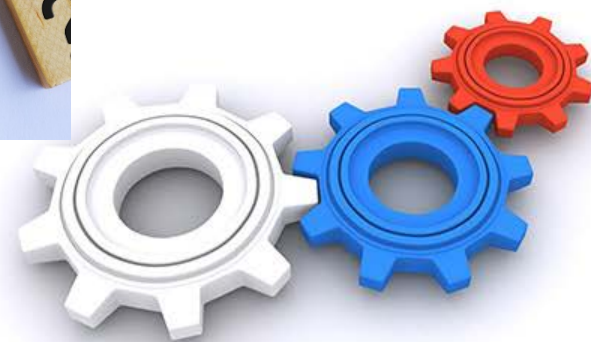

Non-Conformant PDF; 68%

PDF Document Conformance among "Top 10 Downloads" (CFO Act Agencies)

Not a PDF; 9%

Conformant PDF; 23%

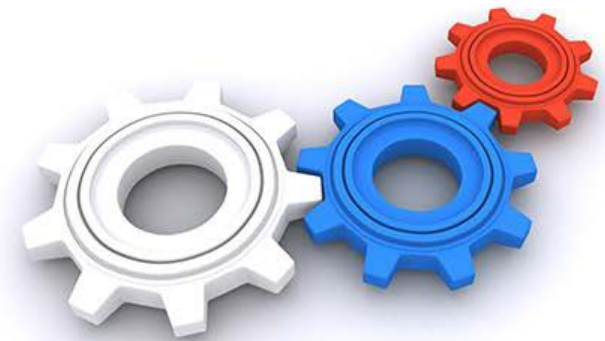**1/3 of PDFs are Accessible** among top 10 downloaded files for CFO Act agencies

The Future of PDFs

What are PDF tags?

Creating PDFs with Tags

Some of the Worst Ways to Create PDFs Are From Web Pages...

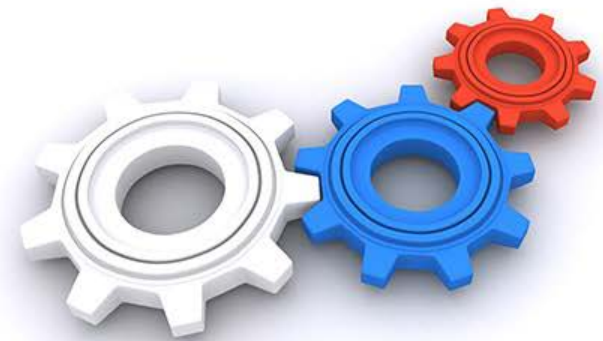Select a destination

Q Search destinations

📄 Save as PDF

🖨 Microsoft Print to PDF
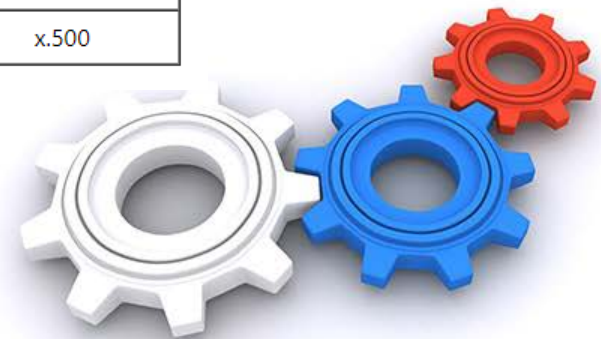
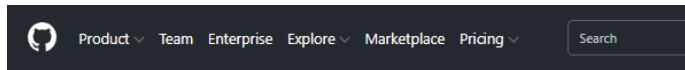🖨 Adobe PDF

…But Often the Worst

Check Out Our Rates!

`Save as PDF`

30-Yr Fixed Rate	30-Yr Fixed 30-Day	30-Yr Fixed 60-Day	20-Yr Fixed Rate	20-Yr Fixed 30-Day	20-Yr Fixed 60-Day
x.500	x.500	x.500	x.500	x.500	x.500
x.500	x.500	x.500	x.500	x.500	x.500
x.500	x.500	x.500	x.500	x.500	x.500
x.500	x.500	x.500	x.500	x.500	x.500
x.500	x.500	x.500	x.500	x.500	x.500
x.500	x.500	x.500	x.500	x.500	x.500
x.500	x.500	x.500	x.500	x.500	x.500
x.500	x.500	x.500	x.500	x.500	x.500
x.500	x.500	x.500	x.500	x.500	x.500
x.500	x.500	x.500	x.500	x.500	x.500

A Quick Search of GitHub

PDF Association

Supporting PDF/UA

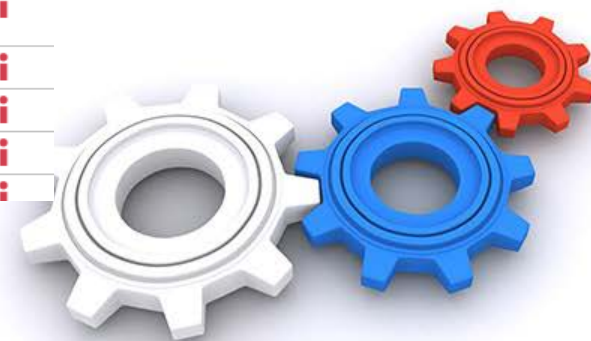The PDF Association recognizes its members supporting ISO 14289 (PDF/UA)! Members offering PDF/UA services are also recognized.

See the key for more information about this table.

NOTE: All claims are made by the software publisher or service provider, and have not been vetted by the PDF Association. Please see each vendor's "More info" link for additional information.

Vendor	Product	Create Tagged PDF	Create PDF/UA	Remediate	Verify	Viewer	More
ABBYY	ABBYY FineReader PDF for Mac	✔	✔	☐	☐	✔	i
	ABBYY FineReader PDF 15 for Windows	✔	✔	☐	☐	✔	i
Antenna House, Inc.	Antenna House Formatter	☐	✔	☐	☐	☐	i
axaio software GmbH	axaio MadeToTag	✔	✔	☐	☐	☐	i
BFO (Big Faceless Organization)	Big Faceless Java PDF Library	✔	✔	✔	✔	☐	i
callas software GmbH	pdfGoHTML	☐	☐	☐	✔	✔	i
	callas pdfaPilot CLI	✔	✔	✔	✔	☐	i
	callas pdfaPilot SDK	✔	✔	✔	✔	☐	i
	callas pdfToolbox CLI	☐	☐	☐	✔	☐	i
	callas pdfToolbox SDK	☐	☐	☐	✔	☐	i

- [Link to this resource](#)

Further Reading

General PDF History

- Nearly every accountancy system, sales system, enterprise resource planning, or invoicing software in the world can generate a PDF. https://support.lightyear.cloud/portal/en/kb/articles/what-the-heck-is-a-system-generated-pd

- PDF was created by Adobe in 1992 to present documents, including text formatting and images, in a manner independent of application software, hardware, and operating systems
https://en.wikipedia.org/wiki/PDF
https://en.wikipedia.org/wiki/History_of_PDF

Future of PDFs

- https://www.techradar.com/news/adobe-announces-biggest-change-to-the-pdf-in-a-generation

- https://www.pdftron.com/blog/pdf-format/a-vision-for-the-future-of-pdf/

- https://www.protocol.com/adobe-future-of-pdf

Vendors that Support PDF/UA

- https://www.pdfa.org/supporting-pdf-ua/

Index of PDF Syntax-related and ISO Publications

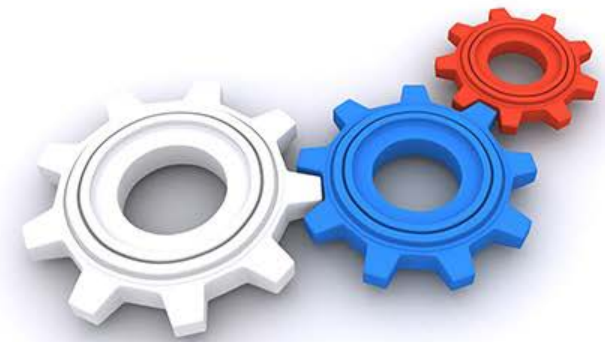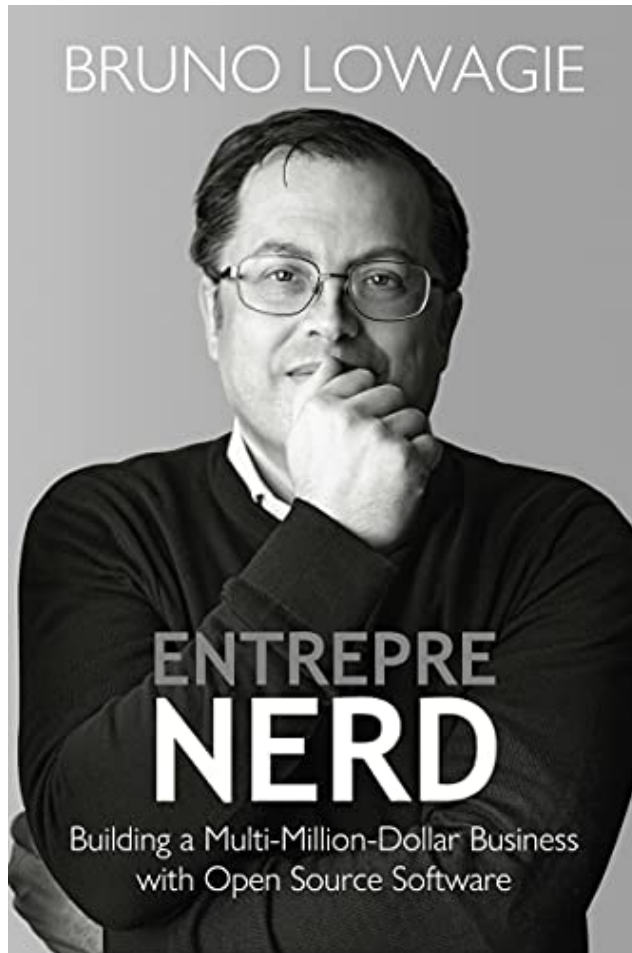- https://www.pdfa.org/index-of-pdf-related-iso-publications/

- https://www.pdfa.org/wp-content/until2016_uploads/2013/08/PDFUA-in-a-Nutshell-PDFUA.pdf

- https://www.pdfa.org/resource/tagged-pdf-best-practice-guide-syntax/

Further reading (continued)

Thank you!