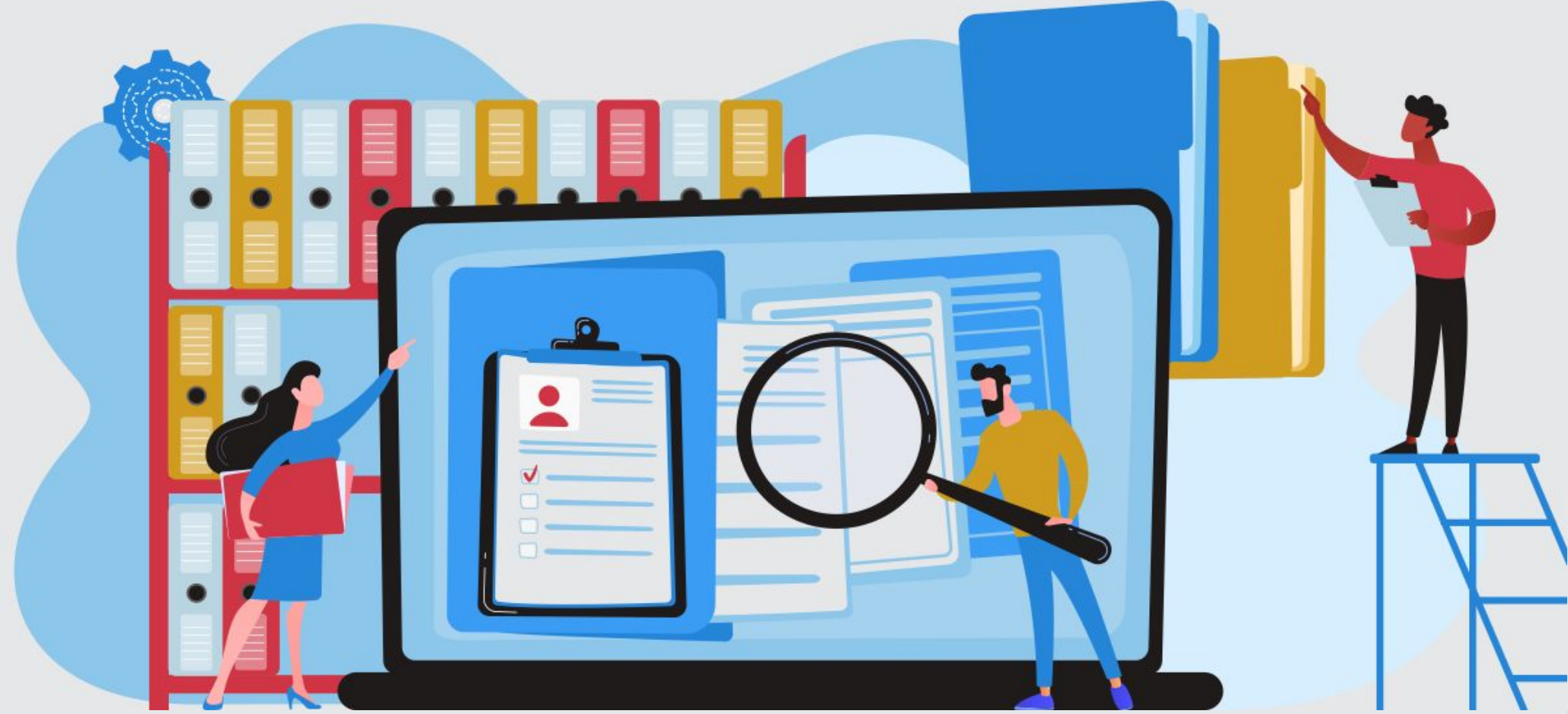




ORPALIS
imaging technologies



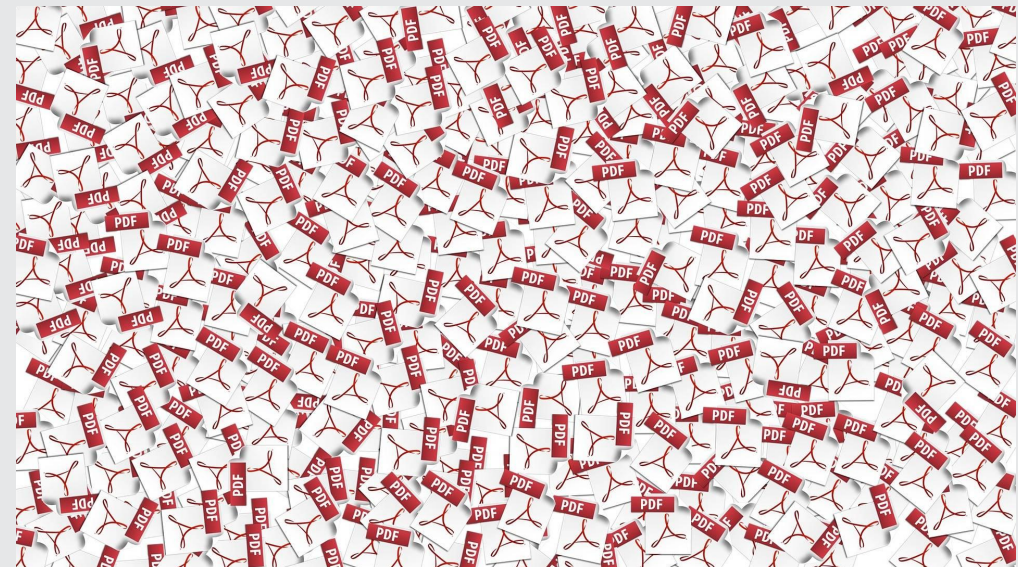
Le format PDF, un outil essentiel pour la transformation numérique

De l'acquisition à l'archivage des documents électroniques

OctoberPDFest
ONLINE

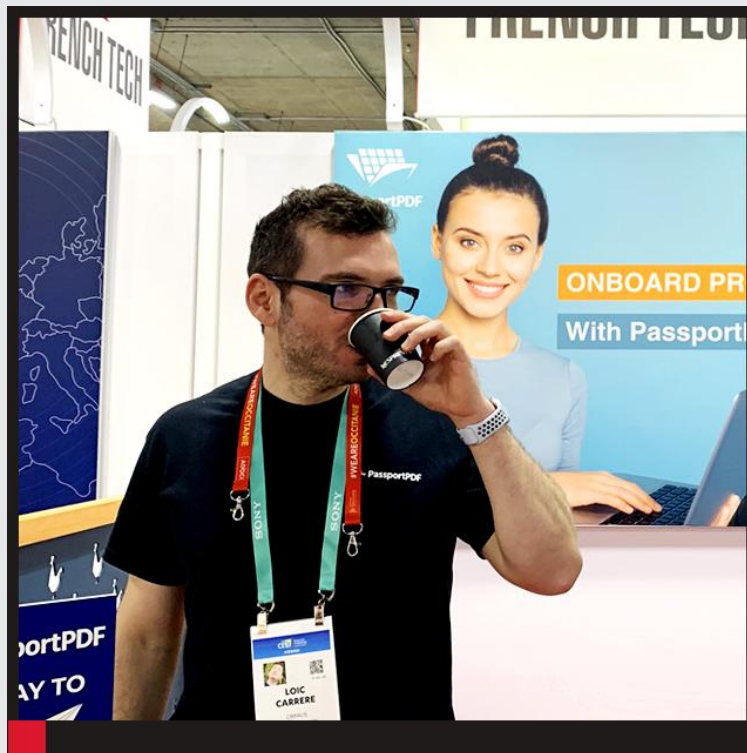
- Le PDF est universel
- Le PDF est normalisé
- Le PDF est flexible
- Le PDF fait gagner du temps
- Le PDF facilite la collaboration
- Le PDF est sûr
- Le PDF est accessible
- Le PDF est durable
- Le PDF est écologique

Présentation de nos solutions
Questions !



Loïc CARRERE, CEO/CTO

spécialiste de l'analyse de documents (OCR, hyper-compression) depuis 2003

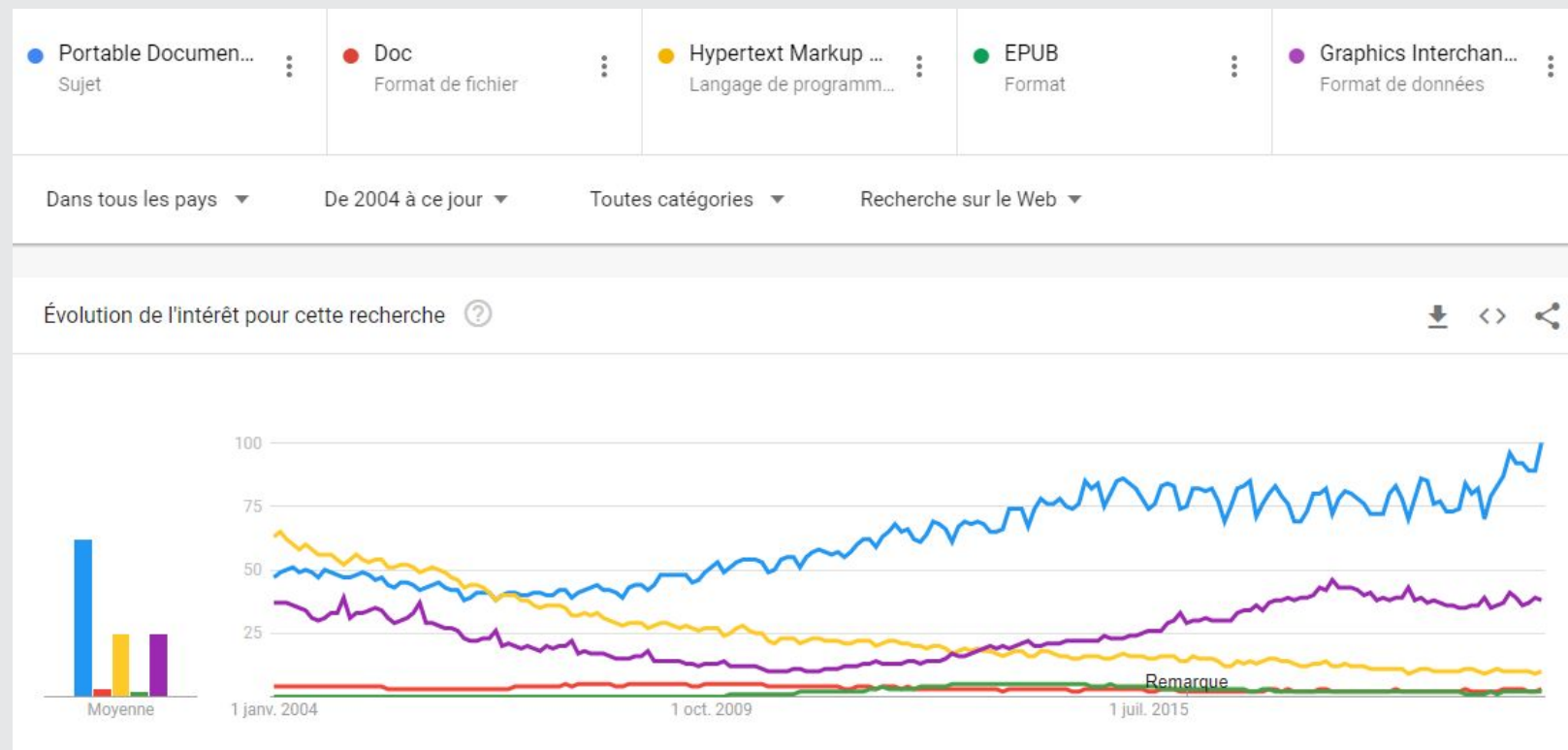


Elodie TELLIER, DG

membre du CA de la PDF Association, organisatrice du 1er PDF Day France à Toulouse en 2019



Introduction



"...and would you like your meal in a PDF file as well, sir?"

<http://www.martybucella.com/>

en pratique :



- interopérabilité
- omniprésent
- perçu comme difficilement modifiable (vs. formats office)



Les bénéfices des normes ISO :

- Format ouvert
- Bien documenté
- Gage de confiance
- Suivi et évolution dans le temps



Le 19/10 OctoberPDFest

<https://www.pdfa.org/presentation/the-iso-standards-process/>

2001 : **PDF/X** - eXchange

2005 : **PDF/A** - Archive

2008 : **PDF/E** - Engineering

2008 : **PDF 1.7**

2010 : **PDF/VT** - Variable data &
Transactional printing

2012 : **PDF/UA** - Universal Access

2017 : **PDF 2.0**

Nouveau

- 2020 : PDF/R - Raster image transport and storage

A venir

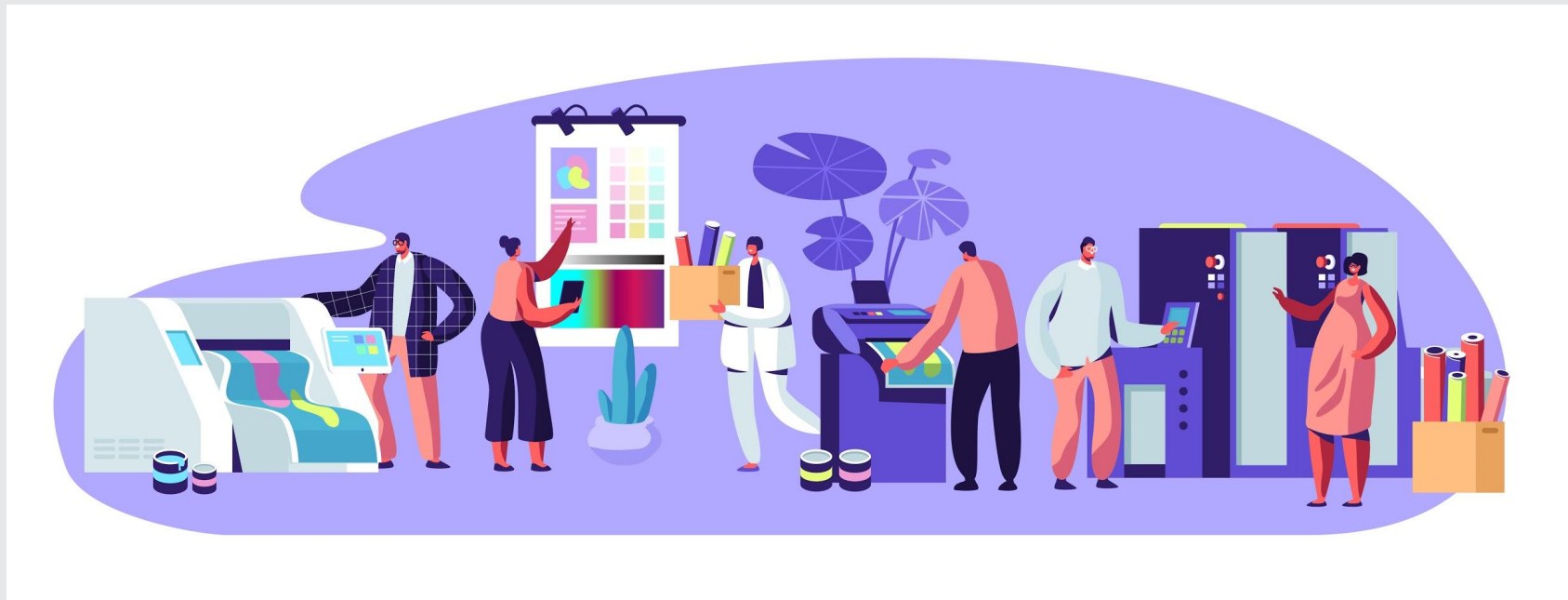
- nombreuses mises à jour : PDF/2.0 (dated revision), PDF/A-4, PDF/X-6, PDF/VT-3

PDF/X : ISO 15930

« Echange de données digitales pour la phase prépresse utilisant le PDF »

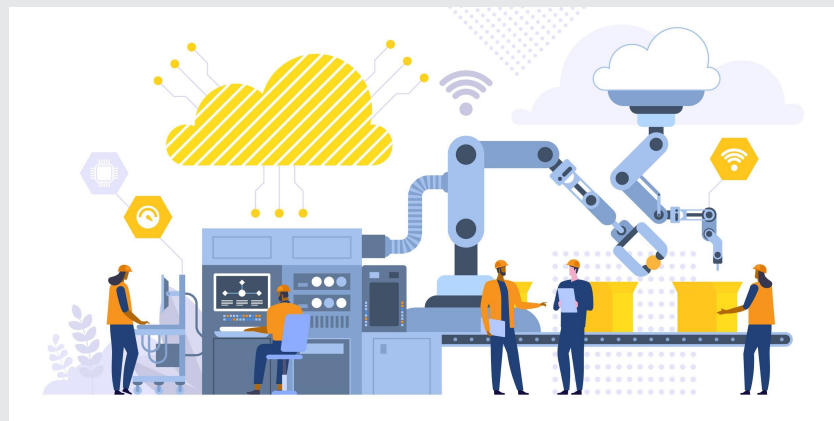
PDF/VT : ISO 16612

« PDF pour les données variables et l'impression transactionnelle »



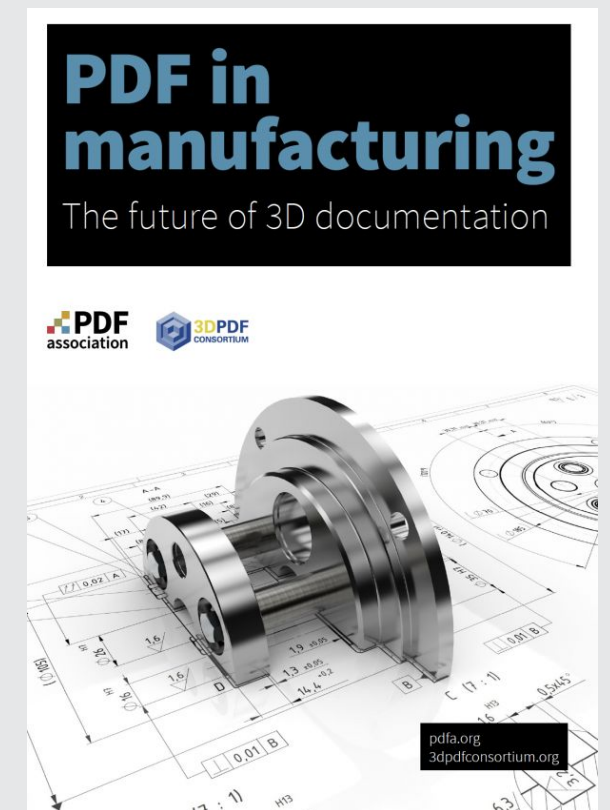
L'ingénierie : PDF/E

- documentation technique
- support des animations et de la 3D
- norme assez peu suivie



La 3D et le PDF :

- depuis 1.6 (2004) version de base pour le PDF/E
- PDF 2.0
- PDF/A-4 (remplace PDF/E)



Un outil indispensable pour :

- extraire les informations de **tout type de document** (tous formats / scannés)
- l'**identification des données sensibles** pour les supprimer
- le travail **collaboratif**
- l'**automatisation**
- l'**édition**

Des options nécessaires :

- recherche dans les mises en page complexes (orientation, graphiques, tableaux...)
- recherche contextualisée
- respect de la casse /mot entier / partie de mot
- l'OCR c'est aussi : MICR, OMR, MRZ, ADR, MRC

- le PDF, premier format à proposer les formulaires (1.2), et utilisé par l'IRS dès 1995
- puis le HTML est arrivé...
- groupe de travail pour améliorer le support des formulaires (responsive)
- format XFA exclu de la version 2.0 (annexe K)

Acme Ltd.

Invoice n° 28

Article	Qty	Price
XXXXXXXX	1	100
YYYYYYY	1	200

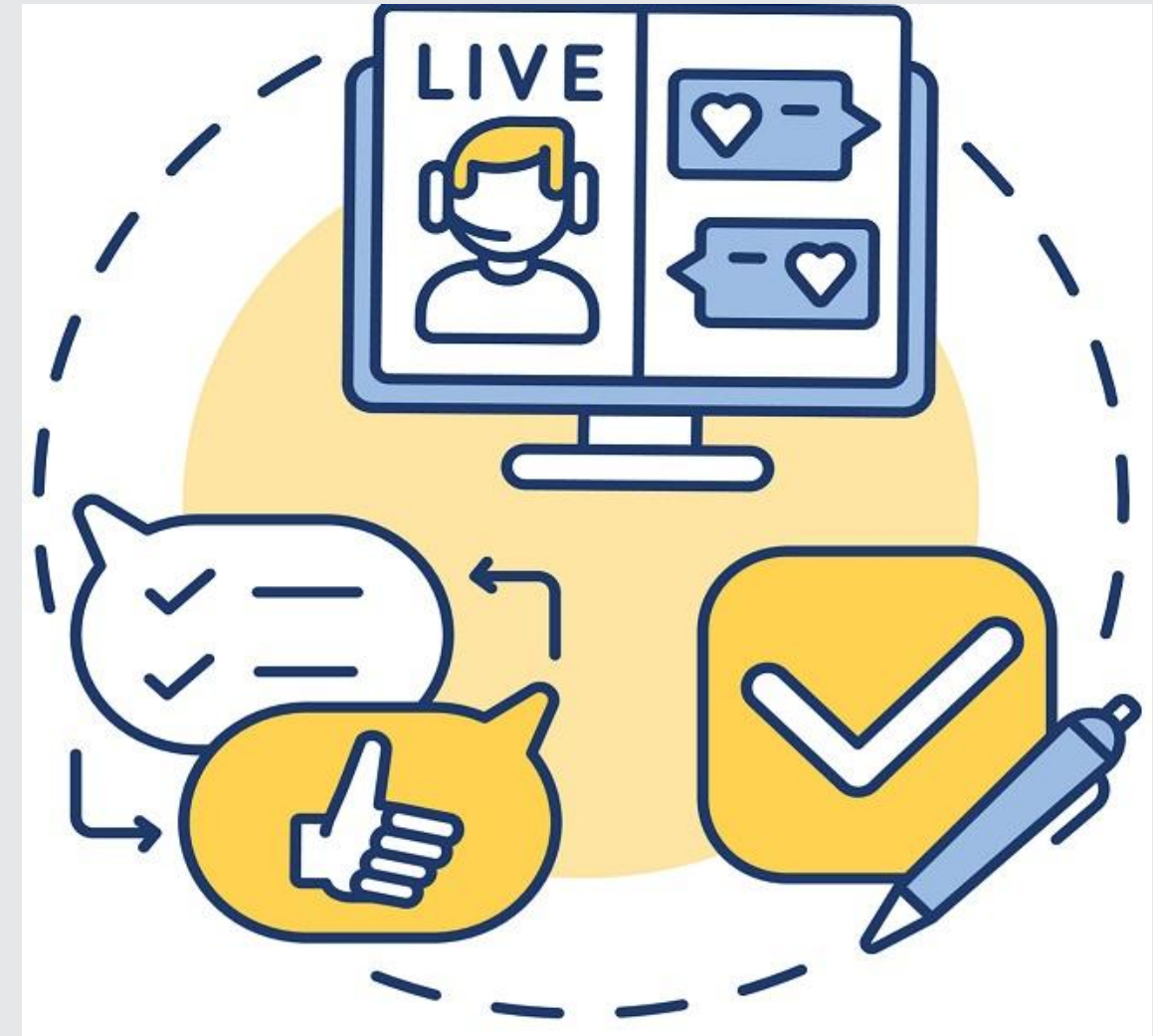
Subtotal

VAT % 20% ▼

VAT

Total

- modèle standardisé pour des annotations entièrement **personnalisables** (12.5)
- formats FDF et XFDF facilitent la communication avec un **serveur**
- collaboration dans le **Cloud** facilitée
- support **multimedia** (13.7)

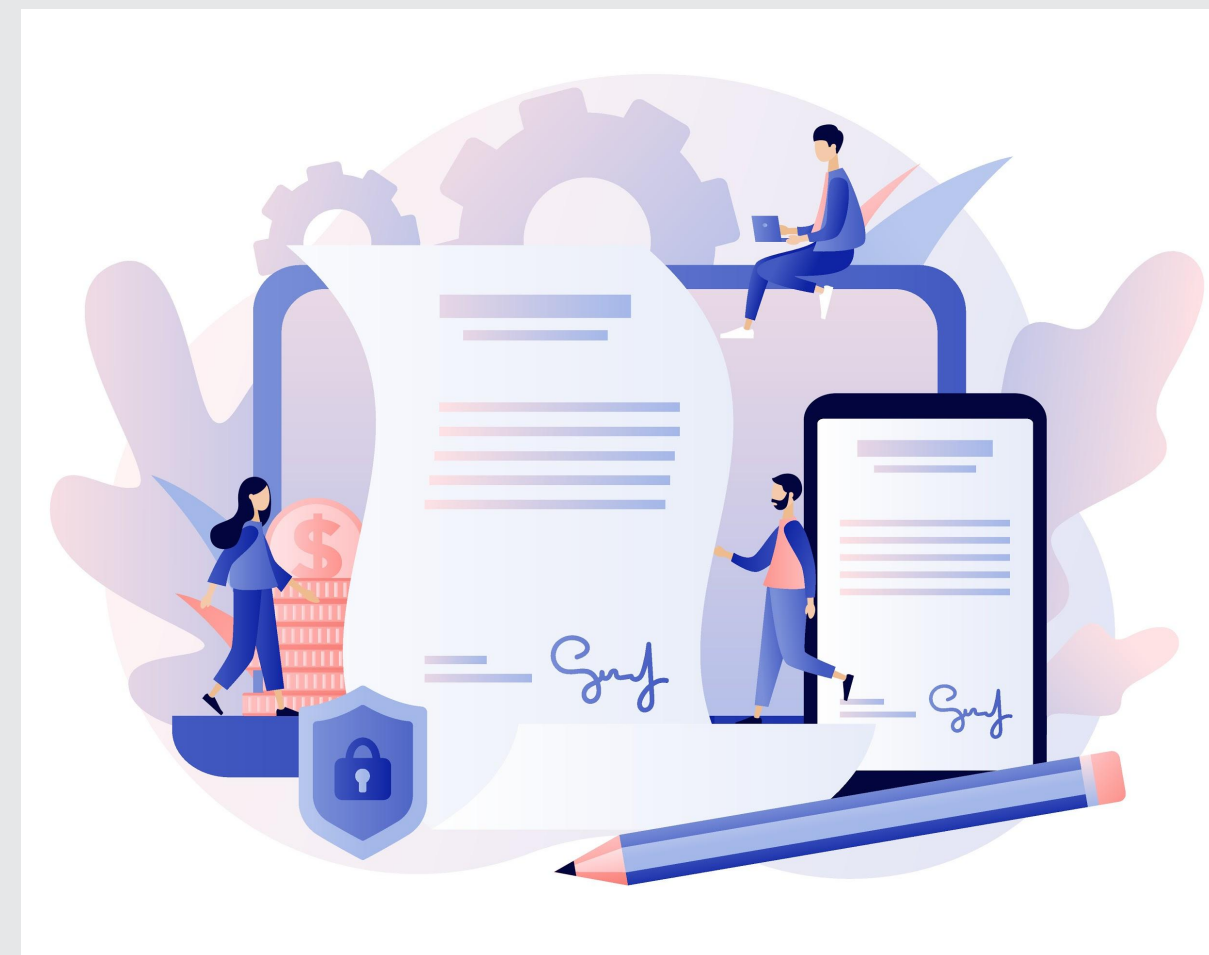


gestion des versions avec
compatibilité des versions PDF

7.5.6 Incremental updates

[WILL REMOVE THIS SLIDE IF THE
WEBINAR IS TOO LONG]

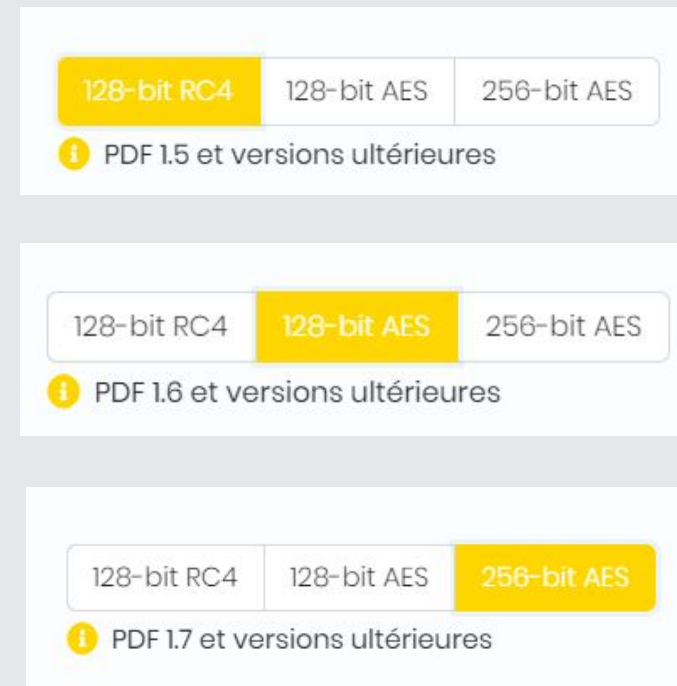
- 3 **types** de signatures électroniques autorisés par l'eIDAS : simple, avancée et qualifiée.
- utilisation du **certificat** numérique pour la signature avancée
- **sécurité** du processus de signature électronique : cryptographie asymétrique (clé publique / privée) et algorithme de hachage SHA-256
- **horodatage** (section 12.8.5 PDF 2.0)



Les différentes options

- niveau d'encryption
- mots de passe lecture / admin
- gestion des droits et permissions pour l'impression, la modification, la copie d'éléments et l'accès au texte pour les lecteurs d'écrans destinés aux malvoyants

Les différents niveaux d'encryption pour le PDF



The image displays three panels illustrating encryption options for different PDF versions. Each panel shows three buttons: '128-bit RC4', '128-bit AES', and '256-bit AES'. The selected option is highlighted in yellow.

- PDF 1.5 et versions ultérieures:** '128-bit RC4' is selected.
- PDF 1.6 et versions ultérieures:** '128-bit AES' is selected.
- PDF 1.7 et versions ultérieures:** '256-bit AES' is selected.

Suppression **définitive** des données :

- “PDF redaction”/ biffer
- outil indispensable (lois)
- à ne pas confondre avec l’édition ou les annotations
- processus manuel ou automatique
- à associer à un “assainissement” du document

p. 44

U.S. Department of Justice
Attorney Work Product // May Contain Material Protected Under Fed. R. Crim. P. 6(e)

III. RUSSIAN HACKING AND DUMPING OPERATIONS

Beginning in March 2016, units of the Russian Federation’s Main Intelligence Directorate of the General Staff (GRU) hacked the computers and email accounts of organizations, employees, and volunteers supporting the Clinton Campaign, including the email account of campaign chairman John Podesta. Starting in April 2016, the GRU hacked into the computer networks of the Democratic Congressional Campaign Committee (DCCC) and the Democratic National Committee (DNC). The GRU targeted hundreds of email accounts used by Clinton Campaign employees, advisors, and volunteers. In total, the GRU stole hundreds of thousands of documents from the compromised email accounts and networks.¹⁰⁹ The GRU later released stolen Clinton Campaign and DNC documents through online personas, “DCLeaks” and “Guccifer 2.0,” and later through the organization WikiLeaks. The release of the documents was designed and timed to interfere with the 2016 U.S. presidential election and undermine the Clinton Campaign.

The Trump Campaign showed interest in the WikiLeaks releases and, in the summer and fall of 2016, **Harm to Ongoing Matter**

HOM WikiLeaks’s first Clinton-related release **HOM**. After **HOM** stayed in contact **HOM** about WikiLeaks’s activities. The investigation was unable to resolve **Harm to Ongoing Matter** WikiLeaks’s release of the stolen Podesta emails on October 7, 2016, the same day a video from years earlier was published of Trump using graphic language about women.

A. GRU Hacking Directed at the Clinton Campaign

1. GRU Units Target the Clinton Campaign

Two military units of the GRU carried out the computer intrusions into the Clinton Campaign, DNC, and DCCC: Military Units 26165 and 74455.¹¹⁰ Military Unit 26165 is a GRU cyber unit dedicated to targeting military, political, governmental, and non-governmental organizations outside of Russia, including in the United States.¹¹¹ The unit was sub-divided into departments with different specialties. One department, for example, developed specialized malicious software (“malware”), while another department conducted large-scale spearphishing campaigns.¹¹² **Investigative Technique** a bitcoin mining operation to

¹⁰⁹ As discussed in Section V below, our Office charged 12 GRU officers for crimes arising from the hacking of these computers, principally with conspiring to commit computer intrusions, in violation of 18 U.S.C. §§1030 and 371. See Volume I, Section V.B, *infra*; Indictment, *United States v. Netyksho*, No. 1:18-cr-215 (D.D.C. July 13, 2018), Doc. 1 (“*Netyksho* Indictment”).

¹¹⁰ *Netyksho* Indictment ¶ 1.

¹¹¹ Separate from this Office’s indictment of GRU officers, in October 2018 a grand jury sitting in the Western District of Pennsylvania returned an indictment charging certain members of Unit 26165 with hacking the U.S. Anti-Doping Agency, the World Anti-Doping Agency, and other international sport associations. *United States v. Aleksei Sergeyevich Morenets*, No. 18-263 (W.D. Pa.).

¹¹² A spearphishing email is designed to appear as though it originates from a trusted source, and solicits information to enable the sender to gain access to an account or network, or causes the recipient to

<https://www.pdfa.org/presentation/remove-sensitive-and-personal-data-from-your-pdf-documents-with-redaction/>

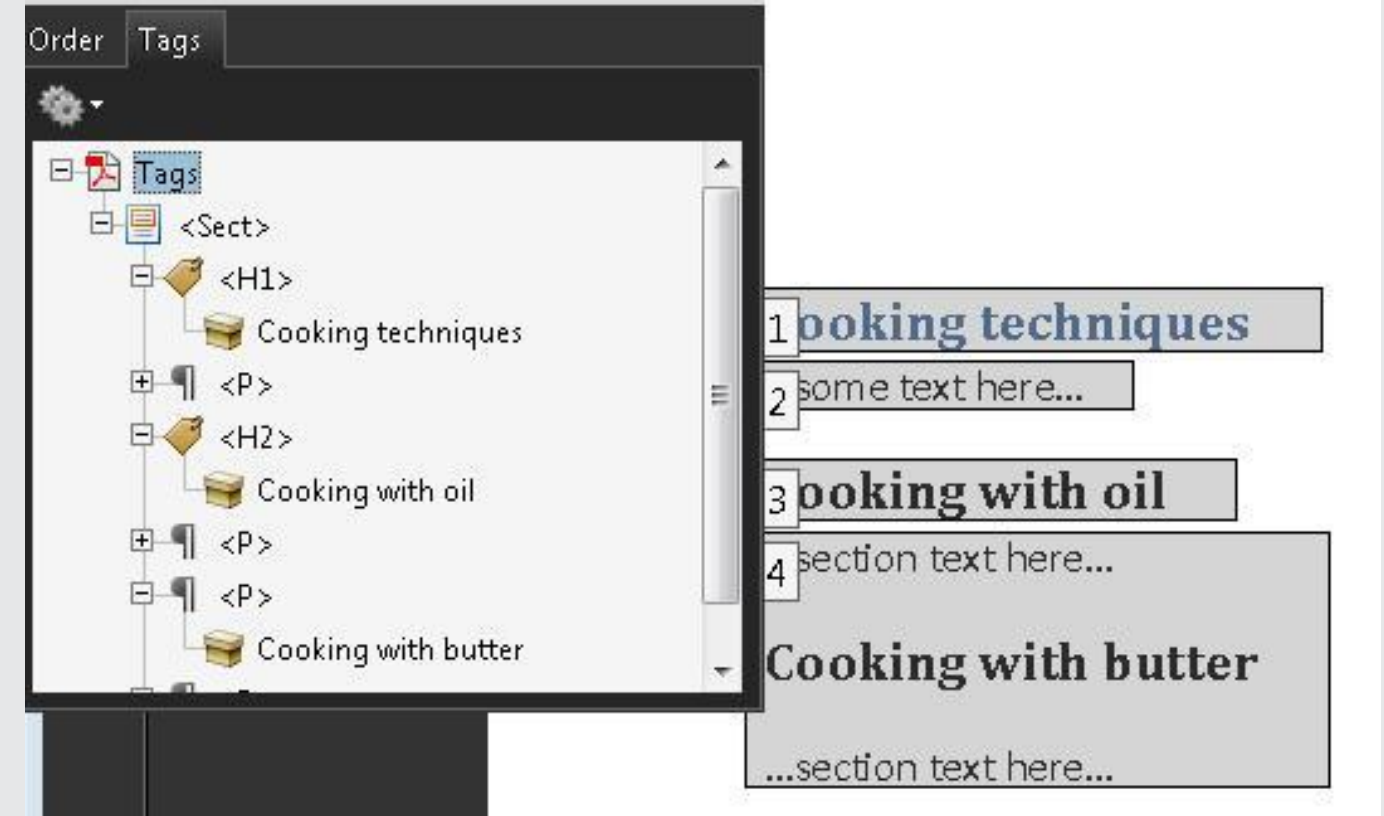
ISO 14289, 2012, 2014

- **spécifications** techniques pour la création, la visualisation et l'utilisation de fichiers PDF accessibles (pour les éditeurs de logiciels PDF et de technologies d'assistance)
- éléments de **validation** pour la conformité PDF/UA

Pour un accès universel

- nécessité **sociale** pour les personnes en situation de handicap
- nécessité **légale** dans de nombreux pays
- compatible avec l'**archivage** à long terme (PDF/A)

- importance des balises (tags) pour des PDF structurés Tagged PDF Best Practice Guide: Syntax
- réutilisation du contenu / reflow PDF Reuse TWG / Liquid Mode pour Acrobat Mobile
- processus de “redaction” facilité Why the need to redact implies using PDF



<https://www.w3.org/TR/WCAG20-TECHS/pdf#PDF9>

ISO 19005 : archivage à long terme

- à associer avec **NF Z 42-013** : 1999, 2009 et **ISO 14641-1** : 2012 (norme internationale)
- format critique pour le secteur public, la finance, la santé et l'ingénierie (contraintes légales)
- **PDF/A-4** à venir, basé sur PDF 2.0

Fonctionnalités clé :

- recherche de texte
- accessibilité
- support de la signature électronique
- compression
- peut être validé



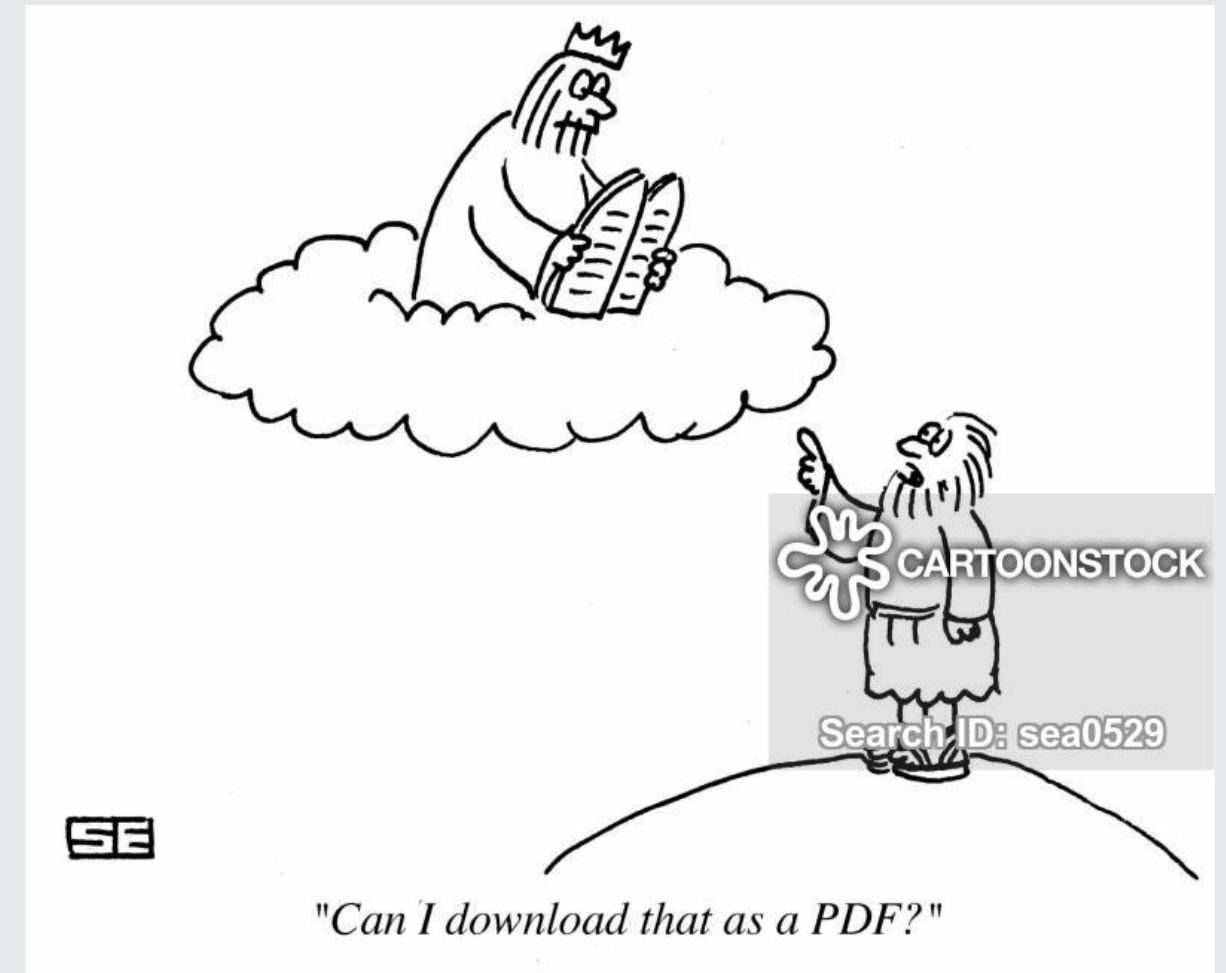
Le mouvement **GreenIT** apporte des recommandations pour :
les logiciels / le matériel / la gestion de l'énergie

Le PDF est un format qui peut être **facilement hyper-compressé**



Le PDF est bien plus polyvalent que ce qu'on pense :

- un format en perpétuelle **évolution**
- **s'adapte** : reflow/reuse pour les smartphones
- un format développé pour **durer**



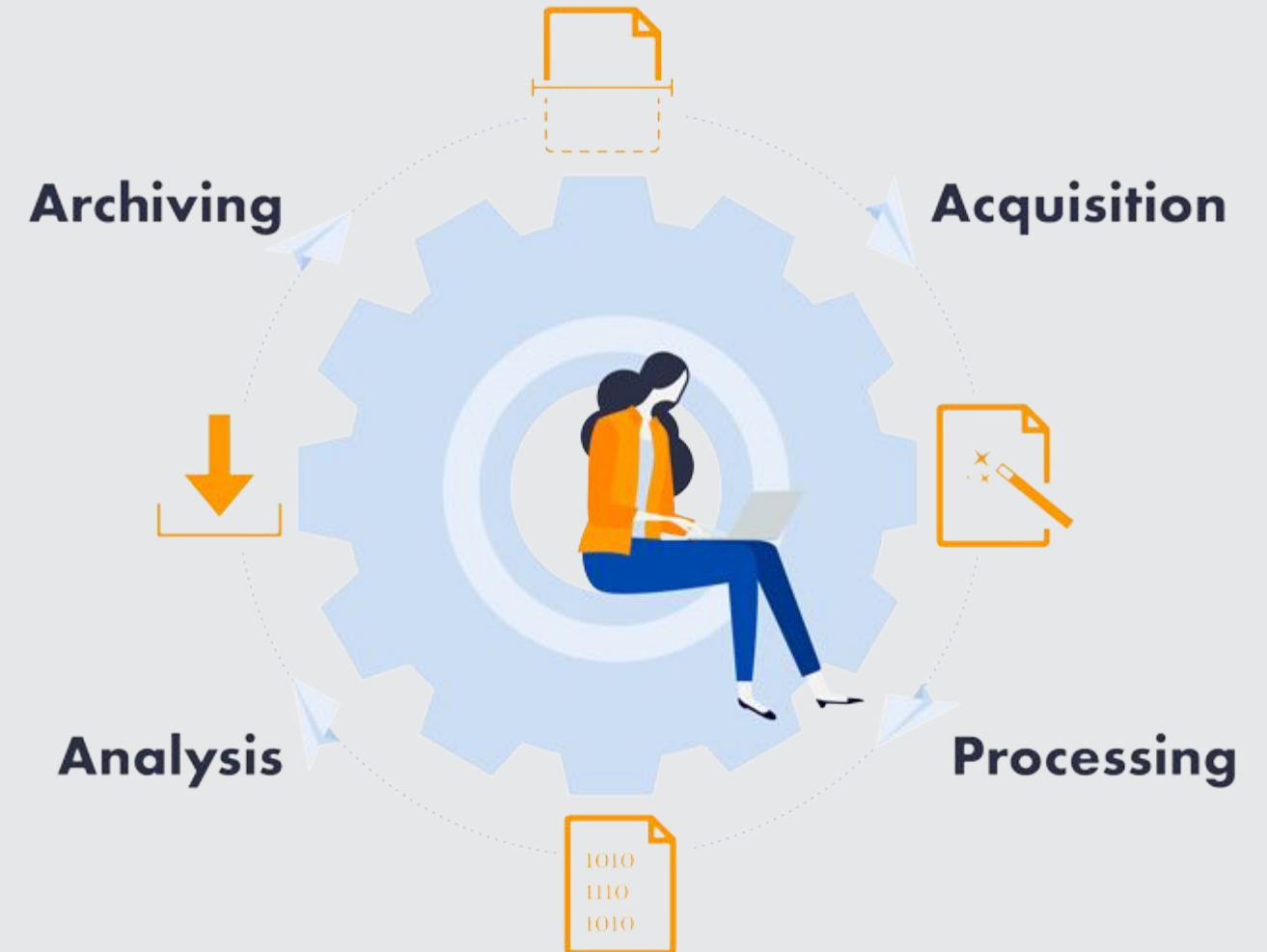
<https://www.staneales.com/>

GdPicture.NET

+ de 3 000 fonctionnalités

PDF, OCR, imagerie, formats, codes
barres

<https://www.gdpicture.com/>



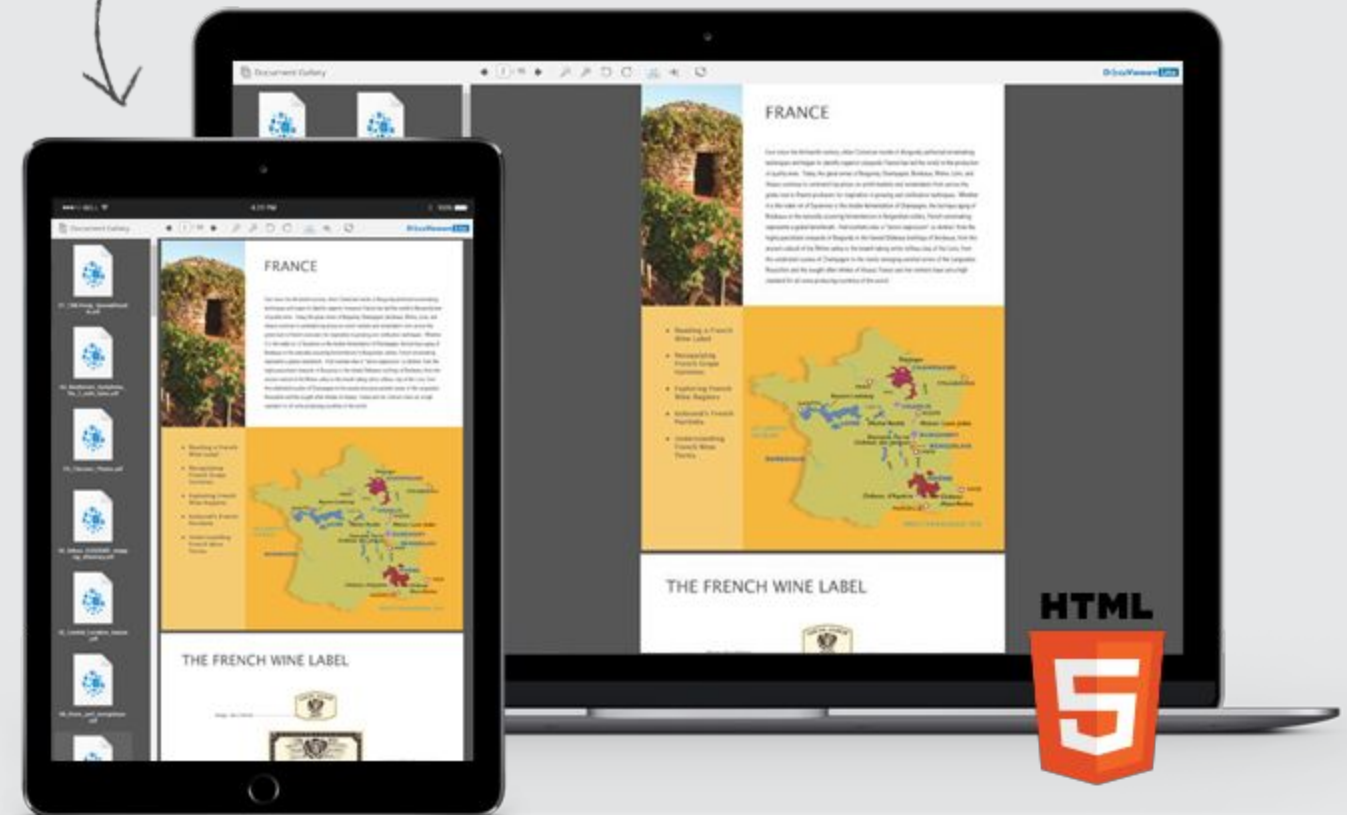
Nos solutions PDF : Web SDK



Capturez, visualisez, modifiez,
annotez, biffez, convertissez, traitez,
compressez, gérez et signez tous vos
documents **en ligne**

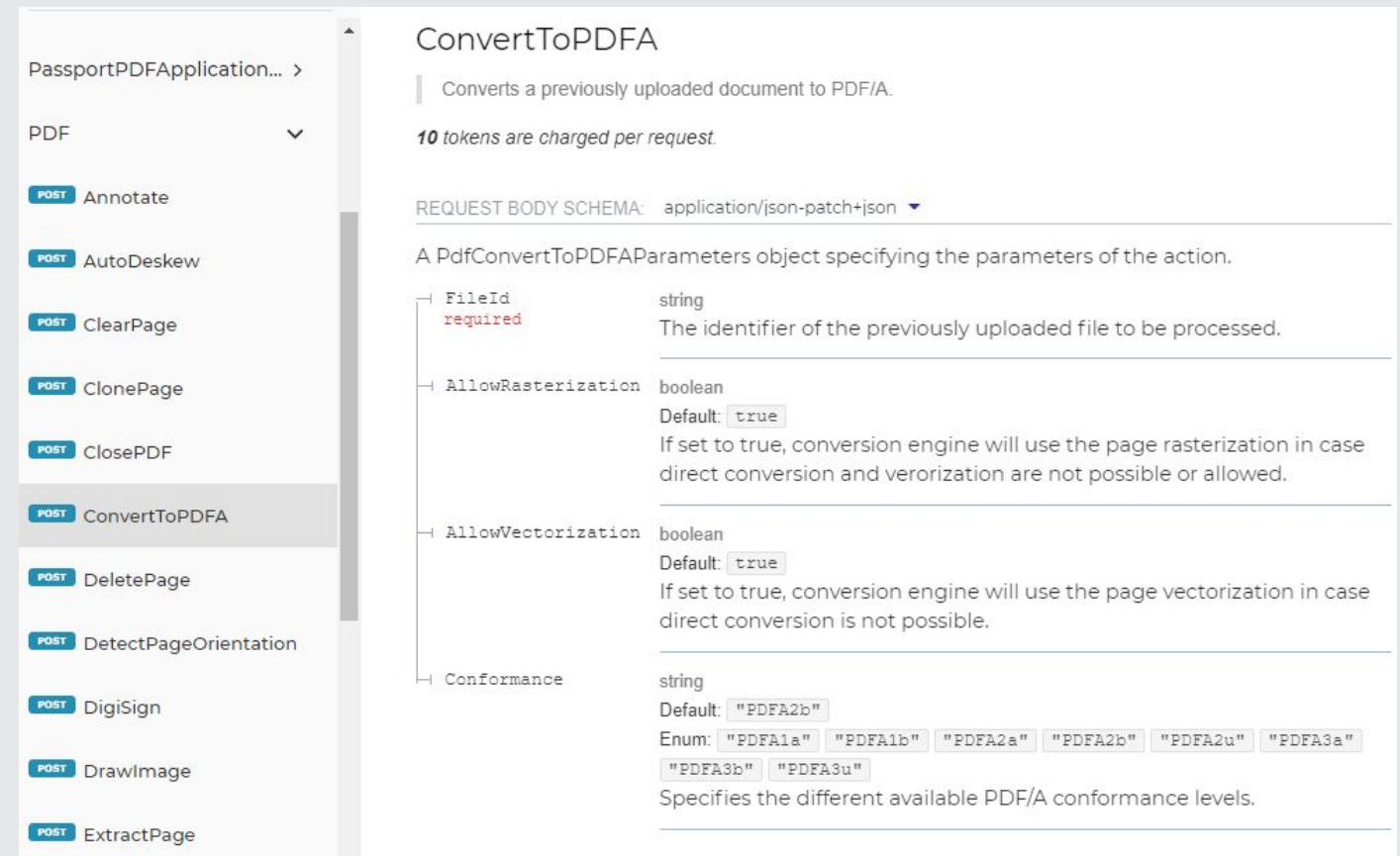
<https://www.docuviewware.com/>

100+ supported formats





- des dizaines de microservices pour gérer les PDF et les documents électroniques
- tutoriels complets pour le développement multiplateforme
- outils de développement open-source



The screenshot displays the PassportPDF API REST interface. On the left, a sidebar lists various endpoints under the 'PDF' category, with 'ConvertToPDFA' selected. The main panel shows the details for the 'ConvertToPDFA' endpoint, including a description, token usage, request body schema, and a table of parameters.

ConvertToPDFA
Converts a previously uploaded document to PDF/A.
10 tokens are charged per request.

REQUEST BODY SCHEMA: application/json-patch+json

A PdfConvertToPDFAParameters object specifying the parameters of the action.

Parameter	Type	Description
FileId	string	The identifier of the previously uploaded file to be processed.
AllowRasterization	boolean	Default: true If set to true, conversion engine will use the page rasterization in case direct conversion and vectorization are not possible or allowed.
AllowVectorization	boolean	Default: true If set to true, conversion engine will use the page vectorization in case direct conversion is not possible.
Conformance	string	Default: "PDFA2b" Enum: "PDFA1a", "PDFA1b", "PDFA2a", "PDFA2b", "PDFA2u", "PDFA3a", "PDFA3b", "PDFA3u" Specifies the different available PDF/A conformance levels.

<https://passportpdf.com/fr/developers/api-reference>



Microservices

- outil **gratuit** en ligne pour les PDF et les documents électroniques
- sécurisation des processus
- des fonctionnalités pour les professionnels



<https://avepdf.com/fr>

Nos solutions PDF : logiciel de numérisation



Logiciel de numérisation et de gestion de documents électroniques



Fonctionnalités de numérisation, traitement d'image et impression avancées



Support de l'OCR pour plus de 60 langues



Technologies d'hyper-compression performantes



Conversion de documents pour l'archivage à long terme



Gestion collaborative avec les annotations



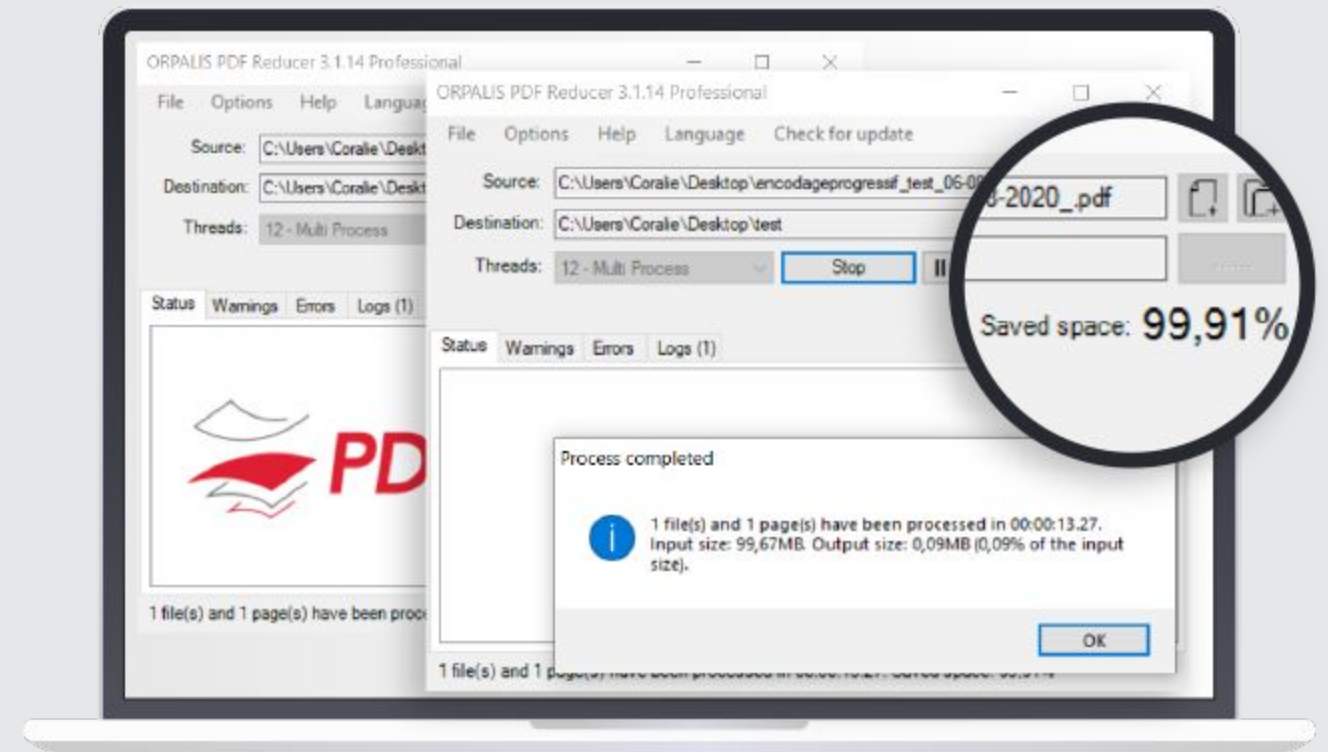
Signature électronique - **Nouveau**

<https://paperscan.orpalis.com/fr/>

ORPALIS **PDF Reducer**

Outil de productivité pour
l'**hyper-compression**

- cloud
- on-premises
- desktop
- serveur



<https://pdfreducer.orpalis.com/>

ORPALIS **PDF OCR** : outil de productivité pour l'**OCR**
cloud / on-premises / desktop / serveur

<https://pdfocr.orpalis.com/>



Site de la PDF Association <https://www.pdfa.org/>
PDF in manufacturing (3D PDF) <https://www.pdfa.org/wp-content/uploads/2020/05/PDF-in-Manufacturing.pdf>
Tagged PDF <https://www.pdfa.org/wp-content/uploads/2019/06/TaggedPDFBestPracticeGuideSyntax.pdf>
Norme PDF 32000 https://www.adobe.com/content/dam/acom/en/devnet/pdf/pdfs/PDF32000_2008.pdf
Norme PDF 2.0 <https://www.iso.org/obp/ui/#iso:std:iso:32000:-2:ed-1:v1:en>
Norme PDF/A <https://www.iso.org/obp/ui/#iso:std:iso:19005:-3:ed-1:v1:en>
Validation PDF/A <https://verapdf.org/home/>
Norme PDF/UA <https://www.iso.org/obp/ui/#iso:std:iso:14289:-1:ed-2:v1:en>
Norme PDF/X <https://www.iso.org/obp/ui/#iso:std:iso:15930:-7:ed-2:v1:en>
Norme PDF/VT <https://www.iso.org/obp/ui/#iso:std:iso:16612:-2:ed-1:v1:en>
Norme PDF/E (préférer PDF/A-4) <https://www.iso.org/obp/ui/#iso:std:iso:24517:-1:ed-1:v1:en>
Norme PDF/R <https://www.iso.org/obp/ui/#iso:std:iso:23504:-1:ed-1:v1:en>
GreenIT <https://www.greenit.fr/>
Hyper-compression https://www.gdpicture.com/wp-content/uploads/pdf_optimization_orpalis.pdf



QUESTIONS

Restons en contact: e.tellier@orpalis.com / l.carrere@orpalis.com

<https://www.orpalis.com/fr/>