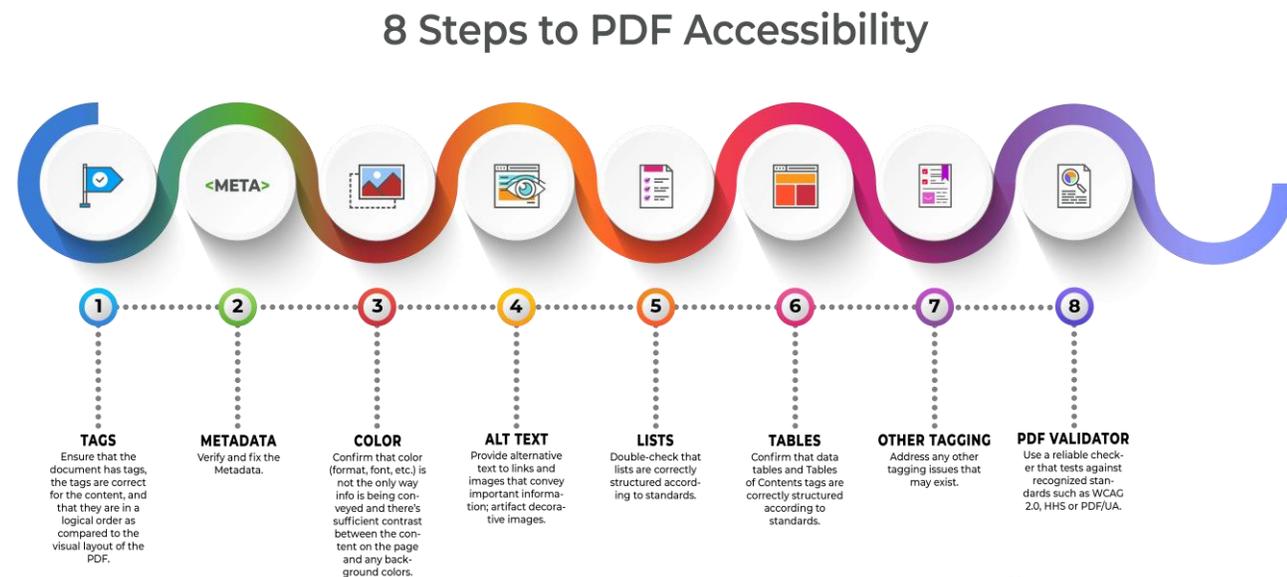


# Automatically improving accessibility of pdf files

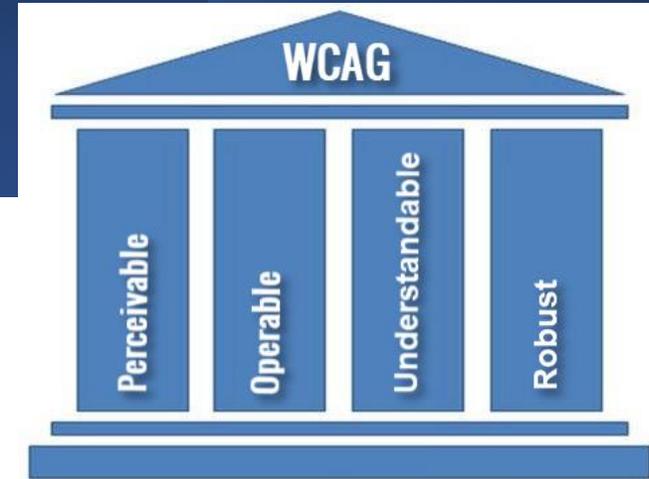


Source 1: commonlook

Student: Gensi Eugenia

Supervisor: Dr. Maarten Marx

# Accessibility



Source 2: Equidox

- EU measure to improve accessibility
- Accessibility: WCAG 2.1
  - Websites, documents, electronic media
  - 4 pillars
- Dutch law: Woo
- Woo documents do not comply

## § 3. Understandable

Information and the operation of the user interface must be understandable.

### § Guideline 3.1 Readable

Make text content readable and understandable.

[Understanding Readable](#)  
[How to Meet Readable](#)

### § Success Criterion 3.1.1 Language of Page

(Level A)

The default human language of each Web page can be programmatically determined.

[Understanding Language of Page](#)  
[How to Meet Language of Page](#)

# PDF objects

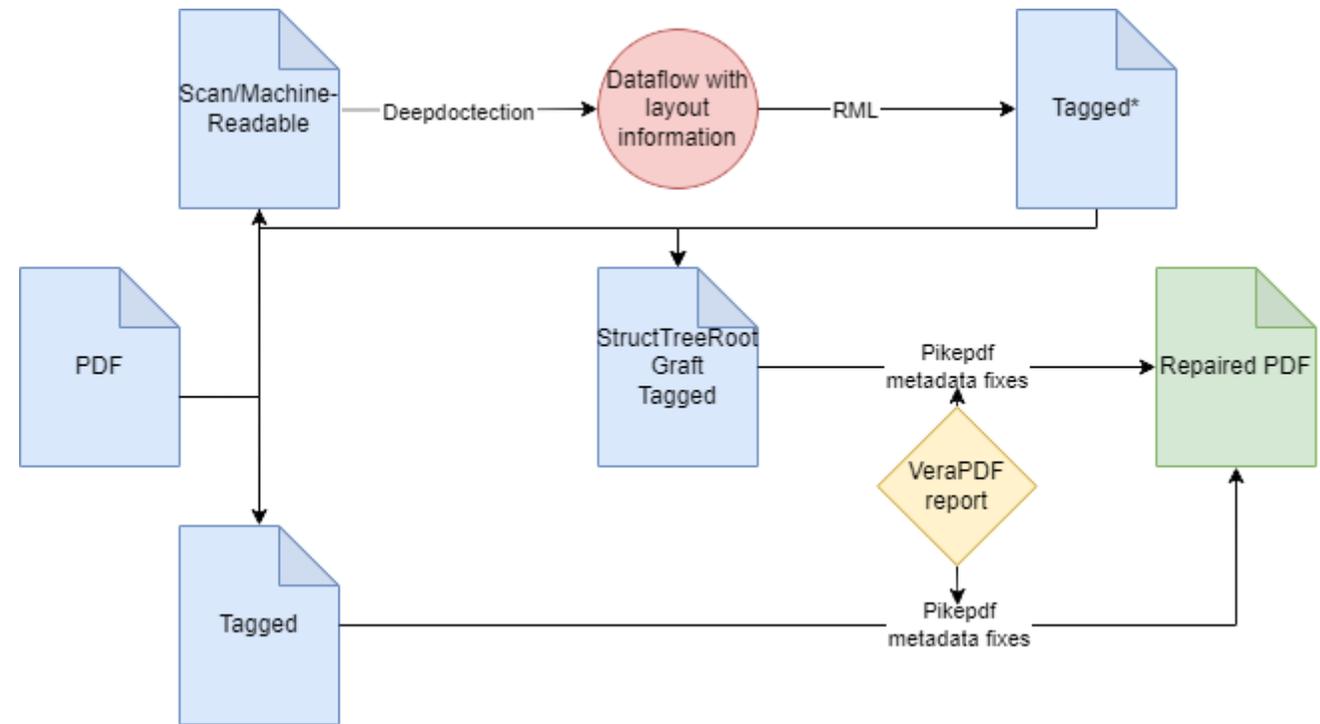
- Multiple versions
- Version in metadata
- Substandards such as PDF/UA
- Modular
- Metadata(stream)
- Content stream

```
})><pikepdf.Dictionary(Type="/Page")({  
  "/Annots": [ {  
    "/A": {  
      "/S": "/URI",  
      "/URI": "http://taggedpdf.com/"  
    },  
    "/BS": {  
      "/S": "/S",  
      "/Type": "/Border",  
      "/W": 0  
    },  
    "/Border": [ 0, 0, 0 ],  
    "/Contents": "Tagged PDF | Accessibility is the right thing to do!",  
    "/H": "/I",  
    "/Rect": [ Decimal('385.743'), Decimal('255.787'), Decimal('470.116')  
    "/StructParent": 1,  
    "/Subtype": "/Link",  
    "/Type": "/Annot"  
  }, {  
    "/Span << /Lang (nl-NL) /MCID 0 >> BDC  
    q  
    0.000008871 0 595.32 841.92 re  
    W*  
    n  
    BT  
    /F1 11.04 Tf  
    1 0 0 1 72.024 759.48 Tm  
    /GS10 gs  
    0 g  
    /GS11 gs  
    0 G  
    [ <002c> 3 <011e016f016f017d> 7 <01c1> 8 <017d> -5 <018c016f011a> ] TJ  
    ET  
    Q  
    EMC  
    /Span << /Lang (nl-NL) /MCID 1 >> BDC  
    q  
    0.000008871 0 595.32 841.92 re  
    W*  
    n  
    BT  
    /F2 11.04 Tf  
    1 0 0 1 120.86 759.48 Tm  
    0 g  
    0 G  
    [ ( ) ] TJ  
    ET  
    Q  
    EMC
```



# Approach

- Collect data
- Initial error analysis
- Tagging and metadata fix heuristic
- Second error analysis
- Correctness check using SSIM



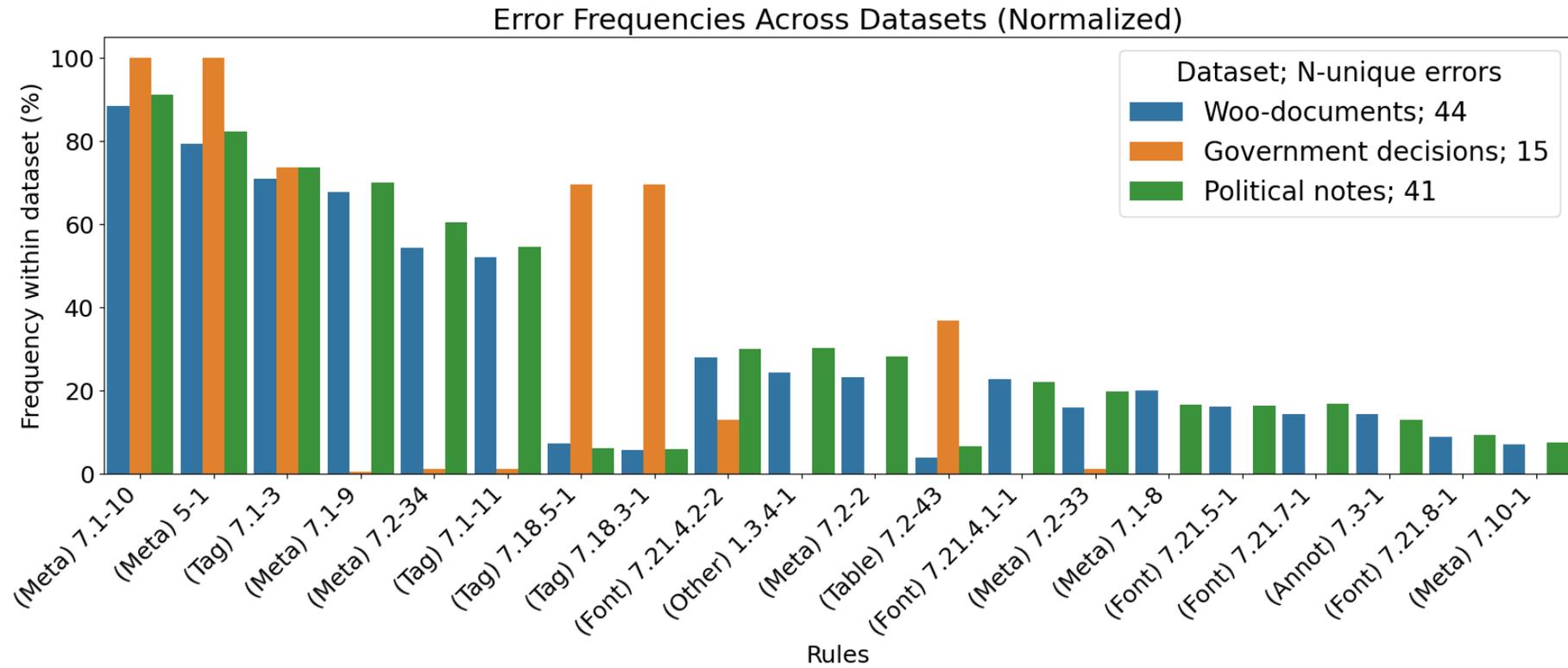
# Measurement

- VeraPDF
- Determine WCAG-2.1 compliance
- Also PDF/UA compliance
- Strictness varies
- Error gatekeeping

<a href="#">Specification: ISO 14289-1:2014, Clause: 7.18.5, Test number: 1</a>	
Links shall be tagged according to ISO 32000-1:2008, 14.8.4.4.2, Link Element.	<b>Failed</b>
2 occurrences	<a href="#">Hide</a>
PDLinkAnnot	
structParentType == 'Link'    structParentType == 'Reference'    isOutsideCropBox == true    (F & 2) == 2	
root/document[0]/pages[1](6 0 obj PDPPage)/annots[0](173 0 obj PDLinkAnnot)	
A Link annotation is nested within BasicLink tag instead of Link	
root/document[0]/pages[1](6 0 obj PDPPage)/annots[1](174 0 obj PDLinkAnnot)	
A Link annotation is nested within BasicLink tag instead of Link	

# Dataset

- From Rijksoverheid
- ~3500 documents between 06-2022 and 12-2024
  - Requests
  - Decisions
  - Inventarizations
- Two supplementary datasets



# Results

- SSIM of 1
- No table, font fixes
- All metadata fixes
- Some tag fixes

Passed rules increase	Woo- documents	Government decisions	Political notes
Mean	3.8	5.1	3.5
Minimum	1.0	4.0	0.0
Maximum	6.0	8.0	6.0





line  
Afzender

line  
Afdeling Juridische Zaken

text  
Minister van Justitie en Veiligheid  
Directie Wetgeving en Juridische Zaken

line  
Postbus 20301  
line  
2500 EH Den Haag

line  
VERZONDEN 30 JAN. 2023

line  
Datum  
line  
25 januari 2023  
line  
Uits kenmerk  
line  
Uw kenmerk

line  
Behandeld door  
line  
Lyo  
line  
Behandeld door  
line  
Kopie aan  
line  
Kopie aan  
line  
Bijlage(n)  
line  
Aantal bijlagen (getal)

line  
Onderwerp  
text  
Reactie Wijziging van het Wetboek van Strafrecht in verband met de verhoging van het strafmaximum voor deelneming aan een terroristische organisatie

line  
Geachte

text  
Met deze brief wil ik u namens de korpschef laten weten dat de wijziging van het Wetboek van Strafrecht in verband met de verhoging van het strafmaximum voor deelneming aan een terroristische organisatie de politie geen aanleiding biedt tot het maken van opmerkingen.

text  
Met vriendelijke groet,

text  
De korpschef van politie,  
namens deze

text  
plv. Hoofd Juridische Zaken

line  
text  
Staf Korpsleiding | Afdeling Juridische Zaken |  
Nieuwe Uitleg 1, 2514 BP Den Haag | Postbus 17107, 2502 CC Den Haag

postnr

Afdeling Juridische Zaken

Minister van Justitie en Veiligheid Directie Wetgeving en Juridische Zaken

VERZONDEN 30 JAN. 2023

Postbus 20301  
2500 EH Den Haag

Datum  
25 januari 2023

Behandeld door  
Tjg

Uits kenmerk  
Uw kenmerk

Kopie aan  
Bijlage(n)

Ty Kope aan  
Typ Aantal bijlagen (getal)

Onderwerp

Reactie Wijziging van het Wetboek van Strafrecht in verband met de verhoging van het strafmaximum voor deelneming aan een terroristische organisatie

Geachte brief wil ik u namens de korpschef laten weten dat de wijziging van het Wetboek van Strafrecht in verband met de verhoging van het strafmaximum voor deelneming aan een terroristische organisatie de politie geen aanleiding biedt tot het maken van opmerkingen.

Met vriendelijke groet,  
De korpschef van politie,  
namens deze

plv. Hoofd Juridische Zaken

Staf Korpsleiding | Afdeling Juridische Zaken | Nieuwe Uitleg 1, 2514 BP  
Den Haag | Postbus 17107, 2502 CC Den Haag

- Only text
- No differentiation with titles, causes small tekst
- Tables, lists, figures
- Manual inspection

# Summary

- Paid software available
- Metadata manageable
  - Not all changes are meaningful
- Tagging: Difficult but important
- Unexplored:
  - Content Stream
  - Screen readers
- Best practices:

We have recommendations that will make the PDF easier for people with disabilities to read.

 Investigate Accessibility

[Learn more about creating accessible PDFs](#)

Thank you!

# Any questions?

Sources:

1. [Volume PDF accessibility checklist – commonlook](#)
2. [The 4 pillars of Web Content Accessibility Guidelines \(WCAG\) – EQUIDOX](#)
3. [An introduction to tagged PDF files - Overleaf](#)