(Some) PDF Pitfalls

Mueller Report

Three-part series from the PDF Association

1. A Technical and Cultural Assessment of the Mueller Report PDF 2019-04-19

2. Even with OCR, the Mueller Report PDF isn't fully searchable 2019-04-20

3. DoJ reposts the Mueller Report! 2019-04-23

Terms

- born digital: text or data initially created in digital format

- redaction: removing context from a document

- optical character recognition (OCR): software that extracts texts from images

- section 508 of the Rehabilitation Act
  - Fed. IT must be accessible to the disabled

Links

- more extensive analysis, worth a read

Exercises

1. Does the metadata in the newer PDF we downloaded indicate that it is PDF/A compliant?

2. Do some of the searchability issues in the earlier PDFs still exist in the one we downloaded?