Bilag 2 – punkt 2.1. 9 og 2.1.10

Metadatafelter:

- Overskrift
- Underrubrik
- Byline
- Brødtekst
- Billedtekst
- Kilde
- Udgave/Region
- Sektionsnummer
- Sektionsnavn [Hvis de findes i avisen]
- Sidetal
- Sidenavn [Hvis det findes i avisen]
- Angivelse af de(n) tilknyttede pdf-side(r)

```
<?xml version="1.0" encoding="utf-8"?>
<pdfinfo>
  <filename></filename>
  <publishing>
    <source></source>
    <edition></edition>
    <publishDate></publishDate>
  </publishing>
  <positional>
    <sectionName></sectionName>
    <sectionNumber></sectionNumber>
    <pageName></pageName>
    <pageNumber></pageNumber>
  </positional>
  <content>
    <text>
      <raw></raw>
    </text>
  </content>
  <articles>
    <article>
      <filename></filename>
    </article>
  </articles>
</pdfinfo>
```

Forklaring af felter:

- <pdfinfo> er xml root
- <filename> er pdf filens navn. <filename>20160501-jyllandsposten-udgave1-page001.pdf</filename>
- <publishing> indeholder info der relaterer til publiseringen af pdf
- <source> er kilden. Eks. <source>Jyllandsposten</source>

Bilag 2 – punkt 2.1. 9 og 2.1.10

- <edition> tilsvarer udgave. Eks:<edition>Udgave1</edition>
- <publishDate> hvornår var pdf'n publiseret af udbyder. Ikke hvornår den er modtaget af IFM. Eks <publishDate>20160501</publishDate>
- <positional> omhandler positionel metadata
- <sectionName> er sektionens navn hvor pdf'en er fundet i avisen. Eks. <sectionName>Kultur</sectionName>
- <sectionNumber> Den side som sektionen optræder på i avisen. Eks. <sectionNumber>2</sectionNumber>
- <pageName> Navnet på den side som pdf'en er taget fra. Eks. <pageName>Kultur</pageName>
- <pageNumber> sidenummeret i avisen hvorfra pdf'en er taget. Eks. <pageNumber>2</pageNumber>
- <text> her findes informationer om tekst. Er optional og hvis vi ikke kan levere fuld sidetekst bør hele denne node udelades.
- <raw> den rå sidetekst. Uformateret og uden delimitationer mellem områder i pdf. Dette er simpelthen den tekst der er på siden til ren indexering. Er kun relevant såfremt denne tekst kan frembringes.
- <articles> Informationer om artikler (Optional)
- <article> en artikel som xml text der er leveret som en del af denne pdf.
- <filename></filename> indeholder filnavnet på den leverede artikel xml.