

Presisering av tidligere krav om UTF-8 deklarerer av XML sendt til Pasientens prøvesvar.

Innhold

1. Hvorfor krav om UTF-8 deklarerer nå?	1
2. Hvordan deklarerer UTF-8 format?	1
3. Hvordan sjekke om en XML-melding faktisk er UTF-8?	2
4. Hva skjer hvis vi ikke har med UTF-8 deklarerer i XML-meldingen?	3
5. Vi sender med UTF-8 deklarerer, men får fortsatt negativ apprec	3
a. Ugyldig UTF-8 deklarerer i XML sendt til Pasientens prøvesvar	3
6. Vi sender uten UTF-8 deklarerer, men får ikke Negativ AppRec	4
a. XML-meldingen vises uten UTF-8 i en nettleser	4
b. Indikasjon på UTF-8 format i ukryptert ebXML-konvolutt	5
c. UTF-8 deklarerer legges på av mellomvare/msh etc.	6

1. Hvorfor krav om UTF-8 deklarerer nå?

Årsaken er at måten nordiske tegn håndteres på i de forskjellige løsningene, har så stor grad av variasjon at spesielt tegnene æ, ø og å ofte ble vist helt feil i tidlig utprøving til kvalitetssikring, selv om avsender sa at innholdet var av korrekt format UTF-8.

Visningen i både KJ Portal og Helsenorge, gjorde at PPS allerede i ca 2022, stilte krav til faktisk deklarerer av UTF-8 i fagmeldingen, med støtte hos Digdir.no med krav om innføring av tegnsett i 2013; *Det er **obligatorisk** å nytte teiknsettstandarden ISO/IEC 10646 representert ved UTF-8 ved all utveksling av informasjon mellom offentlege verksemder og med innbyggjarar og næringsliv.*

I mellomtiden bygde Meldingstjener hos NHN inn en «støttefunksjon», som la inn denne UTF-8 deklarerer i selve XML-fagmeldingen, før Svarrapporten ble sendt til PPS.

Støttefunksjonen «ble senere glemt», men ble fjernet som et avvik/bug fra Meldingstjener 4.9.2025 etter omlegging av løsningsdesign, slik at innholdet som Pasientens prøvesvar nå validerer, er slik det er sendt fra hver lab/rad-system/Meldingstjener.

2. Hvordan deklarerer UTF-8 format?

Kravet gjelder for deklarerer i selve XML-meldingen, hvor dette skal legges inn som en deklarerer først i XML-meldingen.

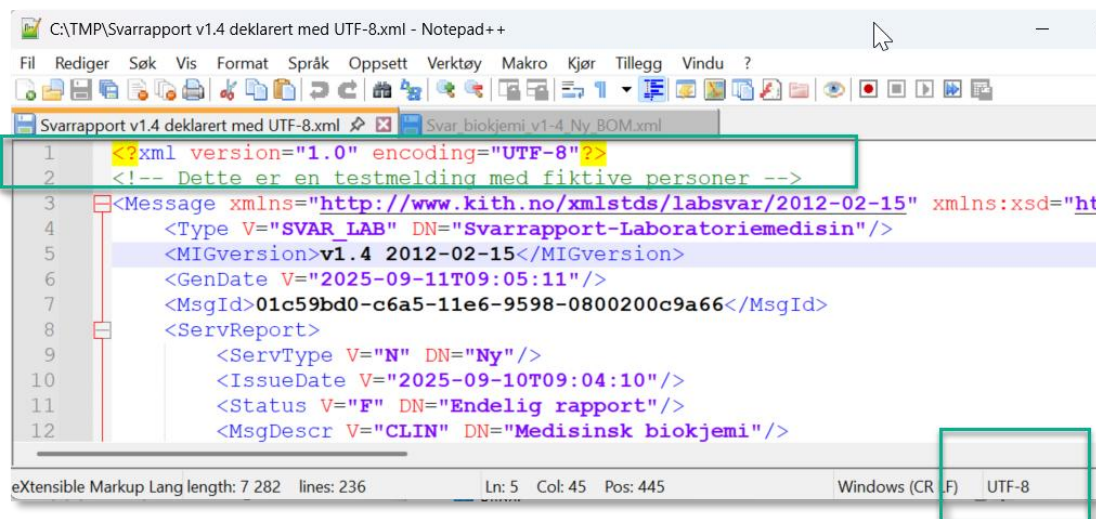
<?xml version="1.0" encoding="UTF-8"?>

Om selve deklareringslinjen legges på ved generering av XML fra fagsystemet, eller av en mellomvare før utsending, er ikke like viktig, men linjen med UTF-8 deklareringslinjen må være på plass, og innholdet må faktisk være av format UTF-8.

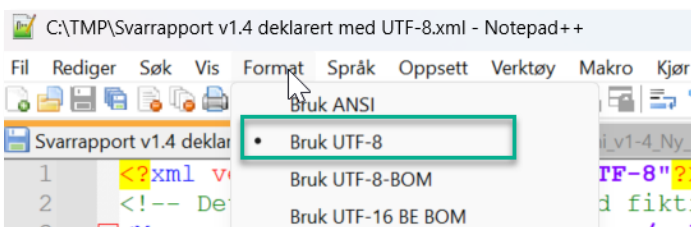
3. Hvordan sjekke om en XML-melding faktisk er UTF-8?

Utfordringer med å sjekke om en XML-melding faktisk er av riktig format, er at vi da henter xml-meldingen ut av et system, hvor vi kopierer, kanskje sender på mail, flytter mellom tekst-editorer etc, som alle på en eller annen måte kan påvirke eller endre opprinnelig format.

Men den enkleste måten er å åpne selve XML-meldingen i f.eks. Notepad ++, og se at innholdet starter med **<?xml version="1.0" encoding="UTF-8"?>**



Når XML-meldingen er åpnet i Notepad ++, ser vi også på nederste linje i editoren, at formatet er UTF-8. Dette ser man også ved å åpne menyen Format – og «Bruk UTF-8» er forhåndsvalgt.



4. Hva skjer hvis vi ikke har med UTF-8 deklarereringen i XML-meldingen?

Hvis det ikke er inkludert deklarerering av UTF-8 i selve meldingen, vil denne feile med en negativ AppRec, og ikke noe av innholdet blir lagret i Pasientens prøvesvar.

```
<Status V="2" DN="Avvist" />
<Error V="X99" S="2.16.578.1.12.4.1.1.8221" DN="Annen feil"
OT="XML encoding is not UTF-8 (found encoding: '')." />
<OriginalMsgId>
  <MsgType V="SVAR_LAB" DN="Svarrapport-Laboratoriemedisin" />
  <TestData>2025-09-11T07:30:17</TestData>
```

5. Vi sender med UTF-8 deklarerering, men får fortsatt negativ AppRec

NB: Dette punktet er til vurdering hos Pasientens prøvesvar om å kunne tillate, og vil avklares i løpet av september 2025.

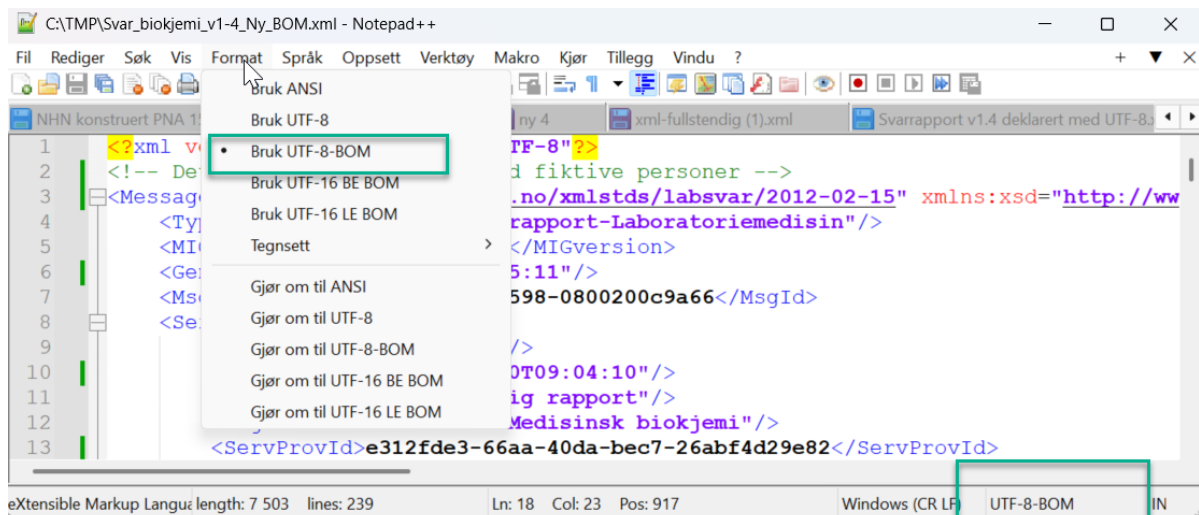
Det er viktig at innholdet faktisk ER utf-8, og ikke varianter av hvordan formidle at innholdet er av format UTF-8. Dette kan være tilfelle hvis lab/rad-systemet benytter en eldre metode for å fortelle at innholdet er av UTF-8 format, og dette ved hjelp av f.eks. Byte Order Mark, BOM. I slike tilfeller legges det på flere binære tegn helt først i XML-meldingen, som ikke er synlige i en normal tekst-editor, men som for Pasientens prøvesvar oppfattes som noe annet enn en UTF-8 deklarerering.

a. Ugyldig UTF-8 deklarerering i XML sendt til Pasientens prøvesvar

Ved å åpne XML-meldingen i f.eks. Notepad ++, ser meldingen tilsynelatende ut til å være korrekt deklartert som UTF-8.

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <!-- Dette er en testmelding med fiktive personer -->
3  <Message xmlns="http://www.kith.no/xmlstds/labsvar/2012-02-15" ;
4  <Type V="SVAR_LAB" DN="Svarrapport-Laboratoriemedisin"/>
```

Men når dere åpner Format, ser dere at dette ikke er UTF-8, men UTF-8-BOM, som inneholder flere binære spesialtegn helt først i meldingen, som medfører at innlesing feiler hos pasientens prøvesvar.



Tilsvarende ser vi i nederste høyre hjørne, at formatet er beskrevet som UTF-8-BOM.

Hvis dere opplever slike situasjoner, ta kontakt med deres leverandør, og få de til å enten slutte å bruke UTF-8-BOM, hvis dette er mulig, eller endre formatet til ren UTF-8 før sending til Pasientens prøvesvar.

Om det er LAB/RAD-systemet eller mellomvare/meldingstjener etc som utfører dette, må avgjøres hos den enkelte virksomhet i hver enkelt situasjon.

6. Vi sender uten UTF-8 deklarerer, men får ikke Negativ AppRec

Hvis dere opplever å sende inn svarrapporter hvor dere ser at det IKKE ligger en deklarerer i starten av filen **<?xml version="1.0" encoding="UTF-8"?>**, men allikevel ikke blir avvist, krever dette mere undersøkelse.

Alternative årsaker kan nevnes, uten at noen av nødvendigvis er årsaken. I og med at dette skjer i Prod, har vi veldig begrenset mulighet til å se hva som faktisk ligger i selve fagmeldingen.

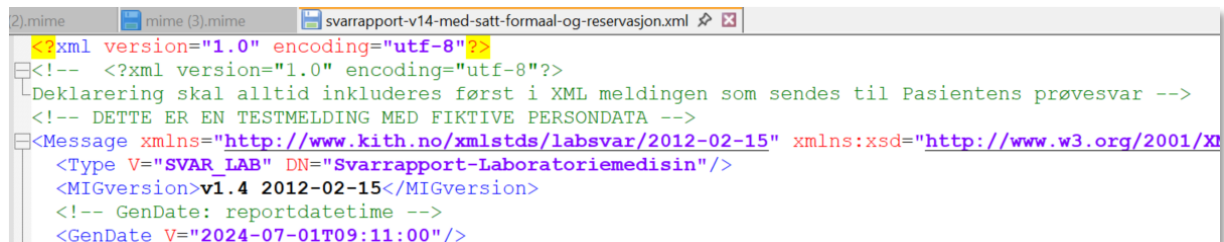
a. XML-meldingen vises uten UTF-8 i en nettleser

Hvis XML-meldingen vises i en browser, er det ikke alle nettlesere som viser tegn som ligger i deklarereringen, som starter med **<?**

Eksempel på dette er på NHN [Utviklerportal](#), hvor det er lenke til hvordan formål og reservasjon legges inn i en Svarrapport XML, hvor vi har måtte lagt inn deklareringslinjen som en tekst. Ved nedlasting og visning i f.eks. Notepad ++, vises selve deklarereringen som normalt.

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<!-- <?xml version="1.0" encoding="utf-8"?>
Deklarering skal alltid inkluderes først i XML meldingen som sendes til Pasientens prøvesvar -->
<!-- DETTE ER EN TESTMELDING MED FIKTIVE PERSONDATA -->
<Message xmlns="http://www.kith.no/xmlstds/labsvar/2012-02-15" xmlns:xsd="http://www.w3.org/2001/XMLSchema"
xsi:schemaLocation="http://www.kith.no/xmlstds/labsvar/2012-02-15 svar-v1.4.xsd">
  <type V="SVAR_LAB" DN="Svarrapport-Laboratoriemedisin"/>
  <MIGversion>v1.4 2012-02-15</MIGversion>
  <!-- GenDate: reportdatetime -->
```



I slike tilfeller vil UTF-8 deklarerings være med ved sending, og bli validert korrekt.

b. Indikasjon på UTF-8 format i ukryptert ebXML-konvolutt

Selve XML-fagmeldingen er pakket inn i en ebXML-konvolutt når denne sendes, kryptert til mottaker. Men selve konvolutten er ukryptert, (selv om den ofte oppleves noe kryptisk...) og inneholder en MIME-del som *kan* gi en indikasjon på hvilket format meldingen er sendt med.

Men igjen, dette er kun en indikasjon, og ingen forklaring på hvorfor man kan oppleve at en XML uten UTF-8 deklarerer faktisk valideres korrekt av Pasientens prøvesvar.

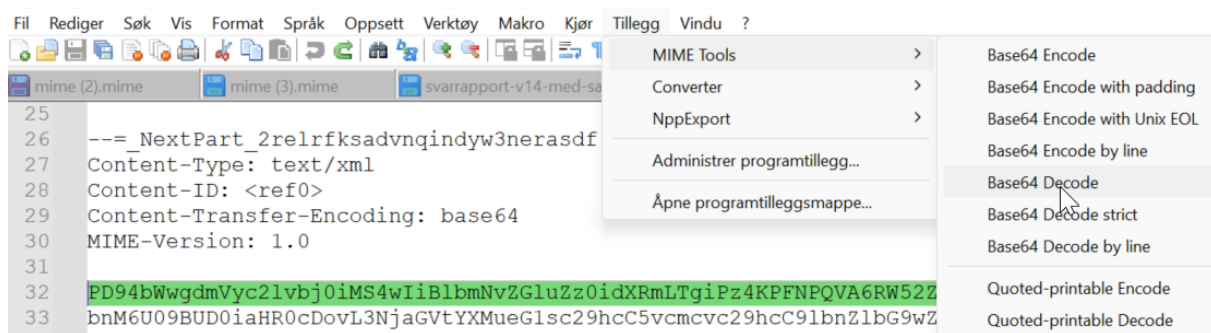
```

26 ---_NextPart_2relrfksadvngindyw3nerasdf
27 Content-Type: text/xml
28 Content-ID: <ref0>
29 Content-Transfer-Encoding: base64
30 MIME-Version: 1.0
31
32 PD94bWwgdmVyc2lvbj0iMS4wIiBlbmNvZGluZz0idXRmLTgiPz4KPFNPQVA6RW52ZWxvcGUgeGls
33 bnM6U09BUD0iaHR0cDovL3NjaGVtYXMuG1sc29hcC5vcmcvc29hcC91bnZlbG9wZS8iIHhtbG5z
34 OmViPSJodHRwOi8vd3d3Lm9hc2l2LW9wZW4ub3JnL2NvbWlpdHRlZXMvZWJ4bWwtbXNnL3NjaGVt
35 YS9tc2ctaGVhZGVyLTFjfmC54c2QiIHhtbG5zOnhsaW5rPSJodHRwOi8vd3d3LnczLm9yZy8xOTk5

```

Ved å åpne selve .mime-meldingen i f.eks. Notepad ++, KAN det stå allerede her i klartekst hvilket format konvoluttens innhold har, men igjen, den sier ikke noe om fagmeldingens format.

Ved å merke første linje og benytte Tillegg i Notepad ++, MIME Tools og Base64 Decode, kan jeg få opp noe mere informasjon om konvolutten, men fortsatt ikke om innholdet i den krypterte fagmeldingen.



```
<?xml version="1.0" encoding="utf-8"?>
<SOAP:Envelope xml
bnM6U09BUD0iaHR0cDovL3NjaGVtYXMueGlsc29hcC5vcmcvc29hcC9lbnZlbG9wZS8iIHhtbG5z
```

Her vises nå informasjon i klartekst at konvolutten som fraktet fagmeldingen, hadde en deklarerings med UTF-8, som KAN være en indikasjon på at fagmeldingen også hadde samme deklarerings.

c. UTF-8 deklarerings legges på av mellomvare/msh etc.

Hvis XML-meldingen hentes ut fra LAB/RAD-systemet, og XML-meldingen beviselig ikke inneholder UTF-8 deklarerings, kan det allikevel være at andre deler av meldingsformidlingen inkluderer dette i selve XML-rapporten, før sending til Pasientens prøvesvar.

I slike tilfeller vil man ikke vite dette sikkert, før man kan hente ut XML som faktisk sendes, og da fortrinnsvis med tilsvarende melding som sendes til test, hvor vi faktisk kan se innholdet i selve fagmeldingen. Dette forutsetter så klart at meldingsløsningene oppfører seg likt i både prod og i test.