

Standarder for tegnsett

Tegnsett ved utveksling av informasjon mellom offentlige virksomheter og med innbyggere/næringsliv

Ved **all utveksling av informasjon** mellom statlige virksomheter og med innbyggere/næringsliv skal tegnsettstandarden ISO/IEC 10646 representert ved UTF-8 benyttes. Inntil videre aksepteres det en begrenset støtte til tegn som finnes i ISO 8859-1 supplert med ytterligere 6 nordsamiske tegn i store og små representasjoner. Statlig sektor skal tilfredsstille kravet fra 1. januar 2012.

Det er obligatorisk for alle statlige virksomheter som skal gjøre større **omlegginger, gjennom nyetablering eller videreutvikling**, å implementere full støtte for ISO/IEC 10646. Kravet skal gjennomføres fra ikrafttredelsesdatoen for denne katalogen. Dersom prosjektet innebærer utvikling av fagsystem bør det stilles krav om dette i kravsspesifikasjonen.

Hensikten bak standarden er å fremme ensartet samhandling i offentlig sektor. Videre vil standarden bidra til støtte for alle nødvendige tegn. Dette er særlig aktuelt i forhold til samiske og europeiske tegn som ofte ikke støttes i eksisterende løsninger.

Dersom fagsystemet skal utveksle informasjon med andre systemer utenfor virksomheten som ikke har støtte for å motta informasjonen (eller avgi informasjon) i UTF-8 bør systemeierne informeres om at de må sørge for konvertering fra UTF-8 til det tegnsett fagsystemene støtter.

Kravet om UTF-8 gjelder ikke ved utveksling internt i virksomheten.

Nærmere om forvaltningsstandarden

Tredje kulepunkt angir at når en offentlig virksomhet foretar større omlegginger eller lager nytt fagsystem, skal man implementere tegnsettstandarden ISO/IEC 10646. Dette ivaretas ved å stille dette kravet som en del av kravsspesifikasjonen. Det kan være vanskelig å vurdere hva som ligger i "større omlegginger" eller "nyetablering". Det mest sentrale moment er at krav til tegnsett må utgjøre en relativt uvesentlig del av den totale kostnad i prosjektet. Hensikten bak bestemmelsen er at offentlige virksomheter gradvis skal gå over til felles tegnsettstandard.

Dette betyr man ikke må gjennomføre endring av tegnsett i eksisterende fagsystemer uten at det samtidig gjennomføres en større oppgradering av systemet.

Her ligger det en langsiktig tenking, som innebærer at man ikke vil oppnå en overgang til full støtte for ISO/IEC 10646 før om mange år.

Utveksling

Ved utveksling mellom virksomheter skal man benytte ISO/IEC 10646 representert ved UTF-8. Det er særlig dette området som vil generere kostnader frem til overgangstidspunktet i

2012. Denne løsning er en midlertidig ordning, inntil fagsystemene har full støtte for ISO/IEC 10646.

Utvexling av informasjon innen virksomheten/etaten er ikke omfattet av forslaget.

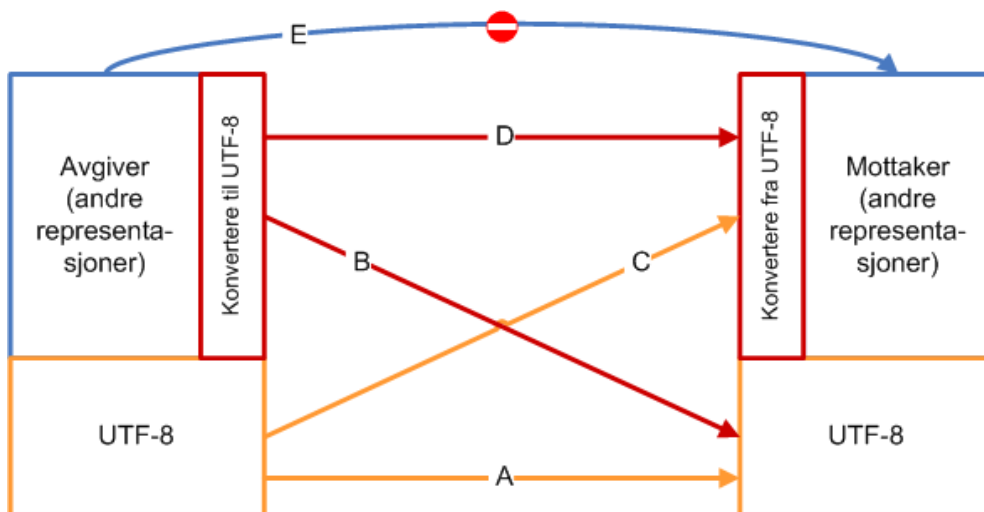
For å lette overgangen er det ikke stilt krav om full støtte for ISO/IEC 10646 representert ved UTF-8: ” Dette gjelder ikke hele ISO/IEC 10646 representert med UTF-8, men det begrensede tegnsettet som også er representert i ISO8859-1, samt ytterligere 6 nordsamiske tegn i store og små representasjoner.”

Dette er angitt som et minimumskrav, og gir ingen fullgod løsning før de interne fagsystemene støtter ISO/IEC 10646.

Det er flere forhold som ligger til grunn for forslaget:

- En felles tegnsettstandard for utveksling vil forenkle samhandling. Når avgiver eller mottaker endrer fagsystem vil ikke dette innebære endringer i det tegnsettformat som utveksles.
- Det er den som ikke støtter ISO/IEC 10646 internt i sine systemer som vil ha ansvaret for å sørge for at informasjon blir teknisk korrekt presentert for omverden som UTF-8, dog begrenset til de tegn en kan håndtere internt i sine systemer.
- For å få en rimelig entydighet i forvaltningen er det også foreslått hvilke erstatningstegn som skal brukes for de spesielle samiske tegnene som ikke er definert i ISO 8859-1, når de forekommer i kommunikasjonen.

Følgende diagram forklarer hvordan utveksling er tenkt gjennomført:

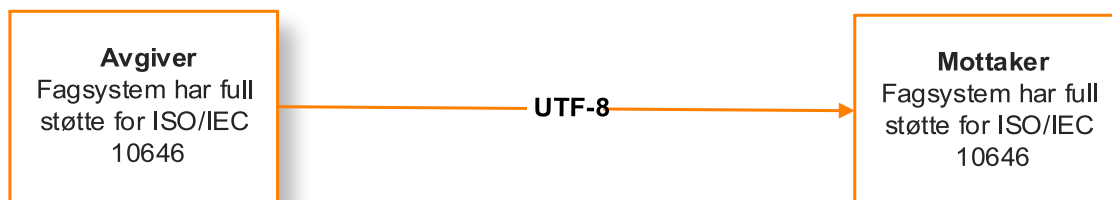


Figur 1: Utveksling

Til venstre er avgiver av informasjon – til høyre er mottaker. Diagrammet kommer også til anvendelse hvor det skjer gjensidig utveksling av informasjon.

Ved å ta utgangspunkt i figur 1 overfor kan man tenke seg forskjellige scenarier:

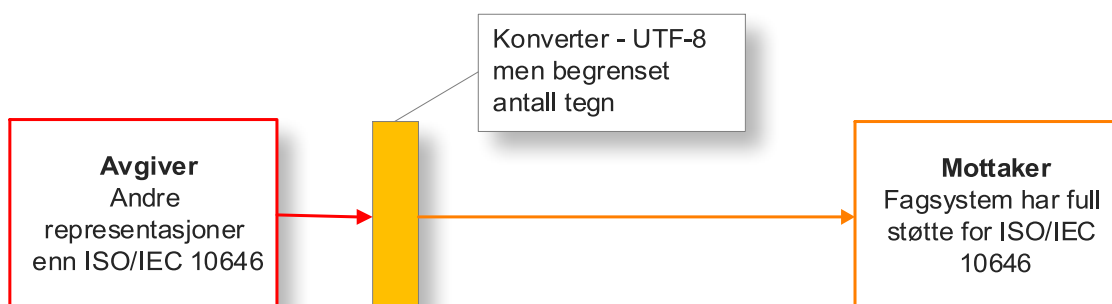
A: Hvor fagsystemene hos avgiver og mottaker har full støtte for ISO/IEC 10646



Figur 2: Utveksling - full støtte for ISO/IEC 10646 hos både avgiver og mottaker

I dette tilfellet har man full samhandling mellom avgiver og mottaker i tråd med tegnsettstandarden. Det er ikke behov for tiltak.

B: Hvor avgiver ikke har full støtte for ISO/IEC 10646 og mottaker har full støtte for ISO/IEC 10646



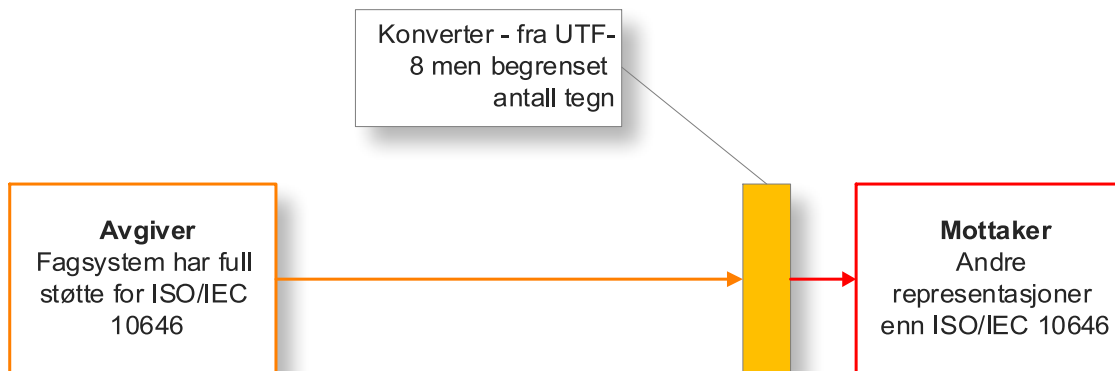
Figur 3: Utveksling - Avgiver støtter ikke ISO/IEC 10646, mottaker støtter ISO/IEC 10646

I dette tilfellet må avgiver sette inn en konverter før avgivelse til mottaker. Kravet er det begrensede tegnsettet som også er representert i ISO8859-1, samt ytterligere 6 nordsamiske tegn i store og små representasjoner. Disse er:

Konsekvens av dette forslaget:

Avgivers fagsystem støtter ikke alle samiske tegn. Ved konvertering vil tegnene være erstatninger – se tabell for erstatning. Mottaker vil kun motta det erstattede tegnet.

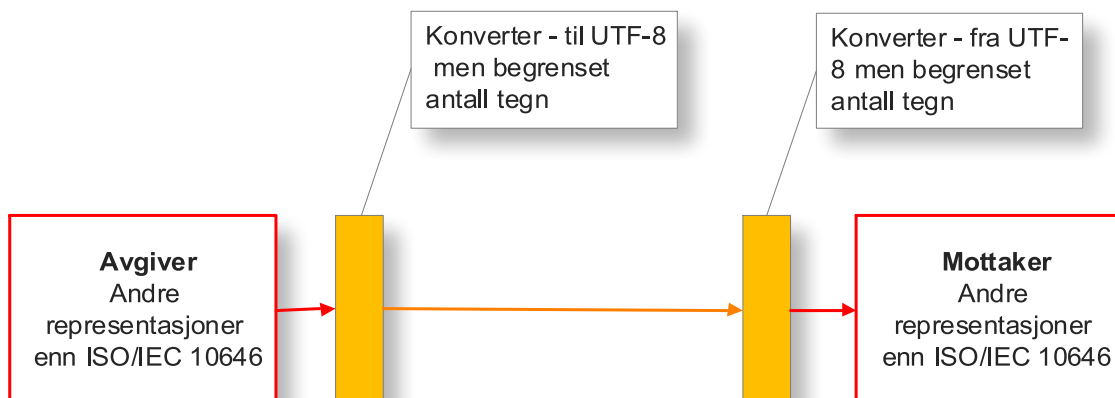
C: Hvor avgiver har full støtte for ISO/IEC 10646 og mottaker ikke har full støtte for ISO/IEC 10646



Figur 4: - Avgiver støtter ISO/IEC 10646, mottaker støtter ikke ISO/IEC 10646

I dette tilfellet må mottaker foreta en konvertering til tegnsett i tråd med fagsystemets. Her må det skje en forenkling av tegnene, jf tabell 1, med erstatning av samiske tegn. Inntil mottaker får oppgradert sitt fagsystem, vil samiske tegn ugjendrivelig erstattes (slik at informasjonen går tapt).

D: Hvor både avgiver og mottaker mangler full støtte for ISO/IEC 10646



Figur 5: - Avgiver og mottaker støtter ikke ISO/IEC 10646

For de tilfeller hvor verken avgiver eller mottaker har ISO/IEC 10646 representert i sine fagsystemer skal det skje en konvertering til UTF-8 ved utveksling med annen virksomhet. Mottaker må da foreta nødvendig konvertering for å få korrekt kodesett i sitt fagsystem. I de tilfeller hvor samiske tegn ikke er representert hos avgiver, vil de heller ikke bli representert hos mottaker. Hensynet bak dette er å sikre at dersom avgiver eller mottaker oppgraderer eller får nytt fagsystem vil overgangen være enklere i forhold til de man samhandler med.

Alternativ D skal benyttes istedenfor alternativ E i figur 1 ovenfor.

Konverteringsløsning

Så langt som mulig bør konverteringsløsningen plasseres utenfor fagsystemene, knyttet til tekststrømmen. De er mange forskjellige løsninger og små programmer som kan løse dette og den enkelte virksomhet må velge løsning ut fra konkrete behov.

Web skjema

Web skjema benyttes på nettsider, men i denne sammenheng er det sentrale at man samhandler med bakenforliggende fagsystemer, og behandles som utveksling (utveksling mellom den som legger inn informasjonen og den som mottar informasjonen). Selv om bruk av tegnsett på nettsider er anbefalt, behandles dette som utveksling.

Gjennomføring

Det er lagt til grunn at den enkelte virksomhet må gjennomføre et prosjekt. Prosjektet innebærer kartlegging hvor man identifiserer kommunikasjonspunkter med andre virksomheter/næringsliv/innbyggere, og identifiserer hvor det er behov for å gjennomføre tiltak. Ved utveksling må konverteringsløsning implementeres og testes.

Det er foreslått satt ned en samordningsgruppe, som i samarbeid med virksomhetene bistår med erfaringer, løsninger og samordning. Gjenbruk av erfaring og løsninger vil stå sentralt i dette arbeidet.

Erstatningstabell for samiske tegn

Ved konvertering fra UTF-8, hvor samiske tegn er korrekt representert, skal følgende tabell benyttes når det ikke er støtte for tegnene i egne fagsystemer:

Tabell 1: Erstatningstabell for samiske tegn

Samisk tegn	Kode	Erstatningstegn
Č	U+010C	C
č	U+010D	c
Ď	U+0110	D
ď	U+0111	d
Ň	U+014A	N
ň	U+014B	n
Š	U+0160	S
š	U+0161	s
Ț	U+0166	T
ț	U+0167	t
Ž	U+017D	Z
ž	U+017E	z