Anvisning för metadata och dokumentation för öppna data

1 Inledning 2

2 Metadata för datamängd 2

2.1 Vi ska fylla i följande metadata i Entryscape 2

2.1.1 Titel 2

2.1.2 Beskrivning 3

2.1.3 Tillhandahållande organisation 4

2.1.4 Kontakt 4

2.1.5 Nyckelord 4

2.1.6 Tema/kategori 5

2.2 Vi bör fylla i följande med metadata 5

2.2.1 Utgivningsdatum 5

2.2.2 Språk 5

2.2.3 Ingångssida 5

2.2.4 Uppfyller 5

2.2.5 Geografiskt område 5

2.2.6 Tidsperiod start 5

2.2.7 Tidsperiod slut 5

2.2.8 Uppdateringsfrekvens 5

2.2.9 Källa 6

2.2.10 Dokumentation 6

2.2.11 Ursprung 6

3 Metadata för distribution 6

3.1 Vi ska fylla i följande med metadata om möjligt: 6

3.1.1 Fil eller länk 6

3.1.2 Åtkomst via webbadress 7

3.1.3 Beskrivning 7

3.1.4 Vanliga mediatyper 7

3.1.5 Creative commons 7

3.2 Vi bör fylla i följande med metadata om möjligt: 7

3.2.1 Språk 7

3.2.2 Utgivningsdatum 7

3.2.3 Ändringsdatum 7

3.2.4 Ägare av upphovsrätten 7

3.2.5 Dokumentation 7

4 Dokumentation och stöd 8

4.1 Beskriv svagheter, brister och problem med informationen 8

4.2 Exempel på där vi själva använder informationen 8

4.3 Definitioner 8

4.4 Format 8

4.5 Ge kodexempel på hur informationen kan användas 8

4.6 Exempel på data 10

4.7 Skapa visualisering 10

5 Beslut 11

# Inledning

Öppna data utan goda metadata är omöjlig att använda. Öppna data med definitioner av datamängden är bättre men kan ändå vara svår att använda. En stor utmaning för öppna data är brister i metadata.

Målet är att metadata om vår information ska vara publicerad på webben så att alla kan se vilken information vi har. Denna guide ska användas både för öppna data och information som ännu inte är öppna data och som får begäras ut som allmän handling.

Det är viktigt att tänka på att metadata i så hög utsträckning som möjligt ska gå att förstå även för de som inte har kunskap om informationsmängden tidigare. Det är ofta en bra idé att blanda in en kommunikatör i arbetet med metadata för att säkerställa god kvalitet i metadata.

I Entryscape finns tre nivåer av metadata, obligatorisk, rekommenderad och frivillig. Denna indelning kommer från standarden DCAT-AP som Entryscape bygger på. I detta dokument pekar vi ut vilka fält Kronofogden ska sträva efter att fylla i och hur de bör fyllas i.

Det behövs ytterligare dokumentation och stöd av informationen utöver den metadata som publiceras i Entryscape för att öppna data ska bli möjlig och enkel att använda. Denna typ av beskrivning av öppna data detaljeras sist i detta dokument.

# Metadata för datamängd

För datamängder ska vi, förutom de obligatoriska fälten, också fylla i alla rekommenderade fält och bör sträva efter att även använda ett antal frivilliga fält. Vi ska även fylla i de obligatoriska fälten på engelska.

## Vi ska fylla i följande metadata i Entryscape

### Titel

Detta ska vara namnet på informationsmängden. Namngivning av öppna data sker ofta utifrån vad den kallas internt i organisationen och det innebär att de som vill använda den ofta inte förstår vad den innehåller och detta ska undvikas. En bra namngivning gör det också lättare att hitta informationen genom till exempel Google. (Många webbplatser får idag 40-60% av sina besökare från Google så det är väldigt viktigt.) Ett exempel på dålig namngivning är “Felrapportering” eller “Statistik”, metadata behöver vara mer specifik än så.

Varje titel ska innehålla en avgränsning, ofta är denna “från Kronofogden”.

Titel ska även finnas på engelska

### Beskrivning

En översiktlig beskrivning i fritext av informationen är viktig och ofta något av det första de som vill använda öppna data läser. Se till att beskriva informationsmängden med alternativa namn som de som letar efter den kanske använder. Till exempel bör “Leverantörsreskontra” också kallas “Inkomna leverantörsfakturor” samt “Fakturor som skickats till oss” eller något liknande. (Använd gärna Wikipedia och Google för att se vad informationsmängden kallas där.)

Beskrivning ska även finnas på engelska

För de informationsmängder som publiceras i Entryscape men som inte är öppna data (om det bara finns metadata) bör vi ange om man i normalfallet kan få informationen elektroniskt på begäran. Rekommendationen från vidsreutnyttjande.se är att vi ska lämna ut information elektroniskt på begäran om det är möjligt. Det är både billigare och enklare för oss, den som tar emot informationen samt är skonsamt för miljön.

Nedan följer några ytterligare perspektiv som bör beaktas när beskrivningen skapas.

Tänk på att beskrivning syns i listan med alla datamängder och inte bör bli för lång. Se följande bild med informationsmängder från skatteverket.se/psidata:



### Tillhandahållande organisation

Välj ”Kronofogden”.

### Kontakt

Det måste finnas någon att kontakta om man har frågor kring informationen. Välj ”Kronofogdemyndigheten”.

### Nyckelord

Det är lämpligt att skriva in nyckelord för datamängder. Försök hitta nyckelord som kan tänkas användas av externa användare som försöker hitta informationen. En datamängd om Förskolor bör till exempel ha Dagis som nyckelord.

Det är ofta lämpligt att ha ”Kronofogdemyndigheten” som nyckelord.

Nyckelorden ska även finnas på engelska

### Tema/kategori

”Regeringen och den offentliga sektorn” är nästan alltid lämplig men kan kompletteras med fler kategorier (tryck på plustecknet). ”Befolkning och samhälle” samt ”Ekonomi och finans” är även de ofta bra val.

## Vi bör fylla i följande med metadata

### Utgivningsdatum

När vi började tillhandahålla datamängden som öppna data.

### Språk

Tala om vilket språk informationen är på. Även om informationsmängden består av siffror finns oftast rubriker och metadata på svenska.

### Ingångssida

Den här egenskapen hänvisar till en webbsida som ger tillgång till datamängden eller dess distributioner och/eller ytterligare information.

### Uppfyller

Om det finns någon specifikation för informationen som publiceras så ska den refereras till här. Om det finns flera olika referenser så klicka på plustecknet för att få fram en rad till och lägg in flera referenser. Det kan röra sig om IOS-standarder eller liknande.

### Geografiskt område

Använd funktionen med förstoringsglas för att välja i lista. För Kronofogden är omfattningen oftast “Sweden”.

### Tidsperiod start

Om det är möjligt ska vi ange från när innehållet i informationen gäller. Ibland är det bara årsvis och ibland kan vi ange ned på timmar. Välj detaljeringsnivå i rullgardinsmenyn.

### Tidsperiod slut

Ange även detta om det är möjligt. Detta är viktigast om det är information som slutat publiceras och inte längre uppdateras.

### Uppdateringsfrekvens

Om det finns en regelbunden uppdatering av informationsmängden ska detta anges.

### Källa

Om informationen är en delmängd av en annan informationsmängd som också publiceras bör vi hänvisa till den. Det kan hända om vi har en stor informationsmängd publicerad och sedan också publicerar en mindre och bearbetad version av samma information. Det är då viktigt för användaren att förstå detta.

### Dokumentation

Se “Dokumentation och stöd” längre ned.

### Ursprung

En beskrivning av hur informationen samlats in och för vilket syfte. För den som ska använda information är det viktigt att veta för vilket syfte informationen samlats in då det kan hjälpa till att avgöra om den är lämpligt för att användas till det personen önskar. Länka gärna till mer information om det finns publicerat på annat ställe, det är ofta lämpligt att ange detta i dokumentationen. Ibland kan samma information publiceras flera gånger som olika datamängder/distributioner och det är då extra viktigt att beskriva vad som skiljer de olika publiceringarna åt.

# Metadata för distribution

En datamängd kan ha flera distributioner. Man kan till exempel publicera alla inkomna fakturor både som Excel, csv och xml och det blir då tre distributioner för den datamängden.

När information som är datumbaserad uppdateras är det ofta bra att INTE skapa en ny distribution utan att istället uppdatera en existerande distribution med ny information. Man uppdaterar då uppdateringsdatum för den aktuella distributionen. En fördel med detta är att det är enklare för användare som vill ha all data, de behöver inte ladda ner alla filer månadsvis.

För distributioner ska vi fylla i alla obligatoriska och rekommenderade metadata och beskrivningen ska även den finnas på engelska.

## Vi ska fylla i följande med metadata om möjligt:

### Fil eller länk

Det första valet är om vi ska ladda upp filen till Entryscape eller om informationen redan finns publicerad någon annanstans. Tänk på att om du redan har en fil (excel eller csv t.ex.) publicerad någon annanstans på internet så är det en länk till den filen som ska publiceras i Entryscape. Dubbellagra inte information om det inte finns en anledning. Vill vi lagra i Entryscape istället bör vi överväga att sluta publicera på andra ställen och istället länka till Entryscape.

### Åtkomst via webbadress

Länka till öppna data så den kan hämtas eller var den kan begäras ut om den ännu inte publicerats som öppna data. Samma information kan finnas i flera olika format (Excel, Csv och Odf t.ex.) och då ska vi skapa en distribution per format.

### Beskrivning

Ska innehålla namn på formatet och en beskrivning av eventuella standarder som informationen följer.

Beskrivning ska även finnas på engelska

### Vanliga mediatyper

Vi ska sträva efter att mediatyper (format som xml eller csv) ska vara så öppna och fria som möjligt och utgår ifrån målgruppernas behov. ”Övriga mediatyper” bör om möjligt undvikas.

### Creative commons

Använd CC0 i första hand (oftast lämplig för data och strukturerad information) och CC-BY i andra hand (ofta lämplig om det finns en ideell upphovsrätt som till exempel en fotografs ideella rätt). Vi kan komplettera med andra licenser (använd då ”Övriga” för detta) om det är nödvändigt men det bör i normalfallet undvikas.

## Vi bör fylla i följande med metadata om möjligt:

### Språk

Välj ”Svenska” om det finns text på svenska i informationsmängden.

### Utgivningsdatum

Det datum informationen först publicerades som öppna data.

### Ändringsdatum

Det datum informationen senast ändrades. Detta fält bör användas som informationen uppdateras manuellt. Vid automatisk publicering via api eller annan automatisk uppdatering är det ofta rimligt att inte använda detta fält.

### Ägare av upphovsrätten

Oftast den egna organisationen.

### Dokumentation

Länka till eventuell distributionsspecifik dokumentation.

# Dokumentation och stöd

## Beskriv svagheter, brister och problem med informationen

Ibland kan detta passa ihop med informationen om syftet för insamlingen. Om vi har kunskap om att informationen inte är möjlig eller lämplig att använda till vissa ändamål men det är lätt att tro att det går för en utomstående ska detta skrivas tydligt. Vi kan till exempel ha fått kunskap om detta genom att det blivit fel tidigare eller genom de frågor som vidareutnyttjare ställer till oss.

## Exempel på där vi själva använder informationen

Beskriv hur vi använder informationen och om det är möjligt visa på exempel där vi använt den. Det kan vara på en webbsida för staden eller i en tryckt publikation (om den finns som pdf bör vi länka till den).

## Definitioner

Beskriv vad öppna data innehåller och vad definitionen av respektive fält är. Tänk på att du ofta har mer förförståelse än målgrupperna så var tydlig och utförlig i beskrivningarna. Denna beskrivning är absolut nödvändig om öppna data ska gå att använda.

## Format

Beskriv de format som informationen publiceras med och vilka för- och nackdelar de har. Länka gärna till mer information.

## Ge kodexempel på hur informationen kan användas

Om informationen delas ut i ett format och på ett sätt som lämpar sig för att programmera mot den direkt (som ett api eller en nedladdningsbar statisk json-fil till exempel) är det ofta lämpligt att ge exempelkod för hur den kan användas. För vilka programmeringsspråk vi bör skapa kodexempel kan variera beroende på formatet på data och målgruppernas behov men generella rekommendationer är javaskript, C# och Java.

Här är en bild på hur det ser ut för Örebro kommuns öppna kartdata:



## Exempel på data

I vissa fall är det relevant att ge tillgång till en exempelfil för data som är enkel att titta på. Det kan till exempel vara relevant om informationsmängden är väldigt stor eller om den är komplicerad och vi vill visa på vissa specifika svårigheter som användarna kan behöva vara medvetna om eller om det krävs registrering för att komma åt öppna data.

Ett exempel på hur det kan se ut från Örebro kommuns öppna kartdata:



## Skapa visualisering

För vissa data och för vissa målgrupper kan det vara nödvändigt att skapa enkla visualiseringar av data. Detta underlättar för målgrupperna att avgöra om informationen innehåller det de är ute efter. [Bilden ger ett exempel på visualisering av öppna geodata från Örebro kommun](http://data.karta.orebro.se/).



# Beslut

Denna anvisning har beslutats av Utvecklingsdirektör Arwid Dahlberg efter föredragning av Huvudarkitekt VI, Patrik Fälth.

Arwid Dahlberg

Patrik Fälth