

UTKAST

Råd för beskrivning av datamängder

Nationellt ramverk för öppna data

Innehållsförteckning

1. Nationellt ramverk för öppna data	3
1.1. Vägledning för att beskriva datamängder	3
1.2. Vem vänder sig vägledningen till?	3
1.3. Dokumentets struktur	3
1.4. Vilka delar av arbetet med öppna data tas upp i dokumentet?.....	3
1.5. Vägledande principer	4
2. En enkel beskrivning av publicerade datamängder	5
3. Beskrivning av metadata på webben	7
3.1. Hur är metadatan strukturerad	7
3.2. Metadatan presenterad på webben	7
4. Att beskriva datamängder	9
4.1. Datakatalogen	9
4.2. Datamängd, kategori, licens med mera	11
4.3. Distribution	14
5. Fördefinierade termer och begrepp	17
5.1. Licenser	17
5.2. Språk	18
5.3. Rumslig/geografisk (kommuner och landsting)	19
5.4. Utgivartyp	19
5.5. Teman/kategori	19
5.6. Uppdateringsfrekvens	20
5.7. Filformat	20
5.8. Begränsningar	21
5.9. Mognad	21

1. Nationellt ramverk för öppna data

Dessa råd för hur kommuner och landsting kan beskriva datamängder som gjorts tillgängliga som öppna data är utarbetade av Kommunförbundet Stockholms län, Peter Krantz AB, PwC och Sveriges Kommuner och landsting med stöd av finansiering från Vinnova. De är en del i ett nationellt ramverk för öppna data som innehåller vägledande principer, råd, mallar och verktyg som är framtagna för att erbjuda ett stöd till kommuner och landsting i deras arbete med öppna data. Övriga delar av det nationella ramverket finns publicerade på <http://www.skl.se/oppnadata>.

Råden är licensierade under en Creative Commons Erkännandelicens (CC-BY). Det innebär att materialet är fritt att använda, sprida, göra om, modifiera och bygga vidare på, även i kommersiella sammanhang så länge som uppgifter om källa och upphovsman anges enligt följande: Råd för beskrivning av datamängder utarbetade av Kommunförbundet Stockholms län, Peter Krantz AB, PwC och Sveriges Kommuner och Landsting med stöd av Vinnova.



1.1. Råd för beskrivning av datamängder

För att utvecklare av programvaror och tjänster samt andra användare på ett enkelt sätt ska kunna hitta relevant data är det viktigt att kommuner och landsting beskriver de datamängder de gör tillgängliga som öppna data på ett strukturerat och enhetligt sätt. Dessa råd syftar till att ge kommunerna och landstingen stöd i detta arbete.

Råden bygger på de rekommendationer för beskrivningar av datamängder i europeiska datakataloger som utarbetats av EU-kommissionen. Syftet med rekommendationerna är att kataloger över offentliga öppna data, så som exempelvis oppnadata.se, ska beskriva datamängder på ett likartat sätt.

1.2. Vem vänder sig råden till?

Dessa råd vänder sig till de personer i kommunen och landstinget som arbetar med informationshantering. Det kan exempelvis vara registratorer, arkivarier eller nämndadministratörer.

1.3. Dokumentets struktur

Inledningsvis beskriver dokumentet en målbild för hur de olika delarna av metadatan kring kommunens eller landstingets datamängder kan komma att se ut när de publiceras på kommunens eller landstingets egna webbplats för öppna data samt på den nationella portalen oppnadata.se.

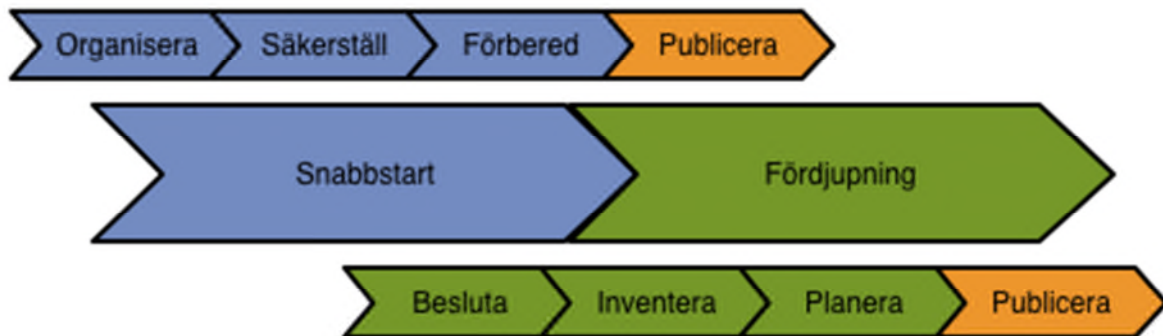
Därefter förklaras steg för steg hur kommunens eller landstingets datakatalog och de enskilda datamängderna kan beskrivas. Avslutningsvis finns ett antal listor med begrepp som ska användas för att beskriva vissa typer av metadata. Det är viktigt att begränsa sig till just dessa begrepp för att underlätta för dem som vill använda motsvarande datamängder från flera olika kommuner eller landsting.

Till råden finns en paketerad folder (zip-fil) som innehåller mallar för att beskriva metadatan.

1.4. Vilka delar av arbetet med öppna data tas upp i dokumentet?

De vägledande principerna för öppna data är uppbyggda dels kring att vissa delar har en högre prioritet än andra, dels utifrån att det ska gå att snabbt komma igång med arbetet för att sedan fördjupa det. Denna vägledning omfattar åtgärder som är aktuella när kommunen/landstinget ska fördjupa sitt arbete med öppna data.

Nedan markeras de faser som behandlas i dokumentet med orange färg. Avsnitt 2 behandlar publicering under snabbstartsskedet och de övriga avsnitten behandlar publicering under det fördjupade skedet.



1.5. Vägledande principer

De vägledande principerna för öppna data rekommenderar kommuner och landsting att publicera sina data på sin egen webbplats så att de kan nås via adressen www.kommunen.se/psidata¹. De vägledande principerna innehåller även en mall för att vilka metadata som bör beskrivas.

Den mall som finns med i de vägledande principerna beskrivs även inledningsvis i denna vägledning. Den utgör ett bra första steg för dem som snabbt vill komma igång med arbetet att publicera information som öppna data. I övrigt innebär denna vägledning en viss utökning av vilka metadata som rekommenderas. Utökningen följer av det arbete som bedrivs på europeisk nivå för att skapa högre användbarhet för de data som tillgängliggörs för vidareutnyttjande.

¹ För landsting blir adressen www.landstinget.se/psidata

2. En enkel beskrivning av publicerade datamängder

För att det ska vara enkelt att hitta de data som kommunen/landstinget tillhandahåller för vidareutnyttjande bör dessa finnas samlade så att de enkelt kan nås på ett ställe. Kommunen/landstinget ansvarar själv för åtkomst till och lagring av de data som man tillhandahåller.

De som snabbt vill komma igång med arbetet att tillgängliggöra information som öppna data bör inledningsvis ordna en förteckning med information om kommunens/landstingets arbete med öppna data samt länkar till tillgängliga datamängder. Sidan bör nås via adressen www.kommunen.se/psidata², denna adress kan antingen omdirigera till den sida där förteckningen finns eller tillhandahålla beskrivningen direkt.

En första enkel förteckning kan utformas med hjälp av tabellen nedan, som också återfinns som bilaga till de vägledande principerna för öppna data. Genom att kommunen/landstinget använder mallen nedan kan intresserade vidareutnyttjare snabbt skapa sig en förståelse för möjliga användningsområden för datamängden. Det blir även enklare för vidareutnyttjare att återanvända data från flera kommuner eller landsting samtidigt.

Värt att notera är att de vägledande principerna föreslår att beskrivningen av datamängder ska göras något tydligare än vad E-delegationen föreslår i sin vägledning. Detta beror på att kommuner och landsting har en större bredd bland sina informationsresurser än statliga myndigheter och att det därför är viktigt att klassificera vilken typ av information som en datamängd innehåller. Att dela upp beskrivningen på fler rubriker gör det också möjligt att begränsa beskrivningar i fritext för att istället använda standardiserade termer och begrepp.

Nedanstående tabeller visar hur de vägledande principerna förhåller sig till E-delegationens vägledning.

E-delegationen	Vägledande principer	Beskrivning
Namn	Namn	Namn på datamängden
Beskrivning	Utgivare	Vilken nämnd/förvaltning som datamängden kommer ifrån
	Verksamhetsområde	Vilket verksamhetsområde som datamängden hör till
	Beskrivning	En kort beskrivning i fritext av datamängden
	Adress	Adress till datamängden
Uppdatering	Publiceringsdatum	Första datum då datamängden gjordes tillgänglig
	Senaste uppdatering	Datum för senaste uppdatering av datamängden
	Uppdateringsfrekvens	Hur ofta datamängden uppdateras
Villkor	Avgift	Vad det kostar att vidareutnyttja datamängden
	Villkor	Vilka villkor som finns gällande upphovsrätt och andra begränsningar för vidareutnyttjande av datamängden

² För landsting blir adressen www.landstinget.se/psidata

Format	Format	Vilket eller vilka filformat som datamängden finns tillgänglig i
	Storlek	Hur stor filen är vid nedladdning
Begränsningar	Begränsningar	Om det finns kända framtida förändringar i tillgången eller utformningen av datamängden
	Mognad	Mognadsgrad i enlighet med femstjärnmodellen (se 5.9 nedan)
Särskild kontaktperson	Särskild kontaktperson	Kontaktperson för datamängden

I förteckningen bör både datamängder som kan laddas ner i sin helhet och datamängder som tillhandahålls genom maskinella gränssnitt för utbyte av information (API:er) beskrivas.

I den mån kommunen eller landstinget tillhandahåller API:er bör det gå att nå fullständiga beskrivningar över API:ernas funktion via förteckningen.

3. Beskrivning av metadata på webben

De kommuner och landsting som vill fördjupa sitt arbete med öppna data och försäkra sig om att deras beskrivningar av publicerade datamängder överensstämmer med den profil som tagits fram för europeiska datakataloger bör följa vägledningen nedan.

3.1. Hur är metadatan strukturerad

Den rekommendation som av EU-kommissionen har tagits fram för hur metadata ska beskrivas för europeiska datakataloger är uppdelad i olika så kallade klasser, där metadatan egenskaper beskrivs. De klasser som finns upptagna är de obligatoriska Agent, Kategori, Kategorischema, Katalog, Datamängd, Literal och Resurs; den rekommenderade Distribution samt de valbara Katalogpost, Dokument, Frekvens, Identifierare, Licensdokument, Lingvistiskt system, Lokal, Medietyp, Tidsperiod, Utgivartyp, Rättighetsklargörande, Standard, Status och vCard.

Egenskaperna är uppdelade i obligatoriska, rekommenderade och valbara egenskaper. Obligatoriska egenskaper är sådana som alltid ska publiceras. Rekommenderade egenskaper ska publiceras om de finns tillgängliga. Vad gäller valbara egenskaper är det upp till varje kommun/landsting att avgöra om de ska publiceras eller inte.

Denna vägledning behandlar de obligatoriska och rekommenderade klasserna och egenskaperna samt vissa valbara egenskaper som finns upptagna i EU:s applikationsprofil. Dessutom behandlas sådana egenskaper som finns upptagna i E-delegationens vägledning i de vägledande principerna för öppna data och i Rådet för främjande av kommunala analysers metadatabeskrivningar för datamängder i Kolada om dessa inte finns med i EU:s applikationsprofil. Utöver detta har egenskaperna också kompletterats med vissa egenskaper som identifierats som önskvärda under en workshop kring frågan som hölls med företrädare för offentlig sektor i samband med Offentliga rummet 2013. De svenska tilläggen utgör i denna vägledning valbara egenskaper även om de finns med som en rekommendation från E-delegationen eller i de vägledande principerna för öppna data.

En fullständig beskrivning av de olika klasserna, hur de kopplar till varandra samt hur de olika egenskaperna definieras finns i DCAT Application Profile for Data Portals in Europe³.

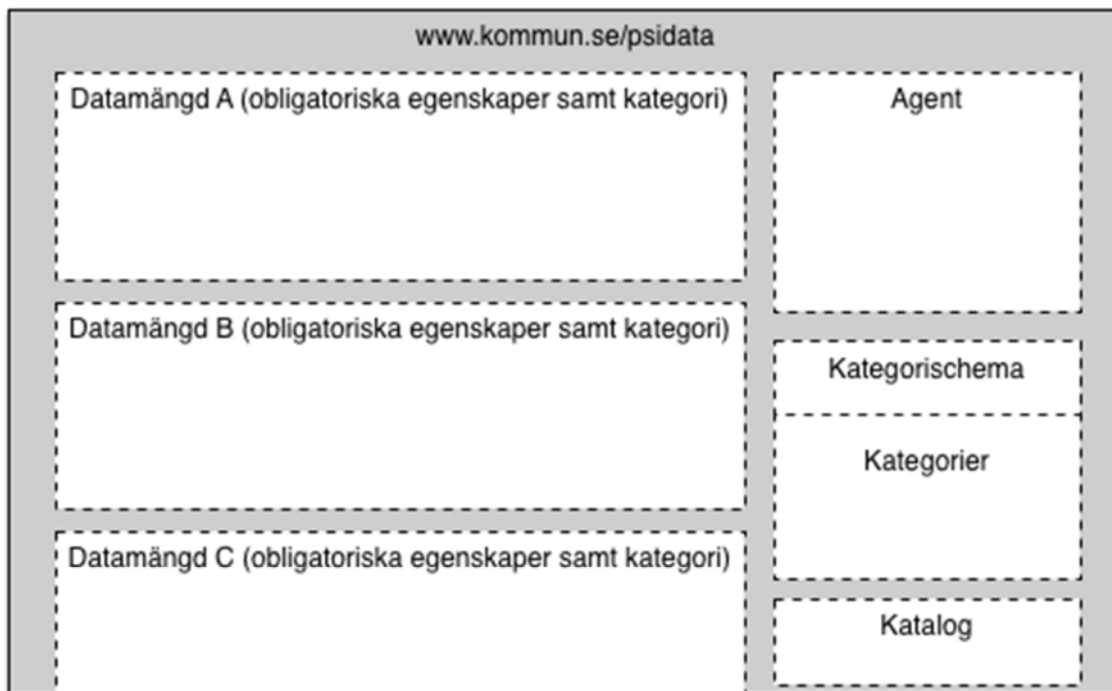
3.2. Metadatan presenterad på webben

Metadata är till för att både människor och maskiner ska kunna förstå de datamängder som publiceras bättre. I nedanstående figurer redogörs översiktligt för hur metadatan kan presenteras i läsbar form på en webbplats exempelvis den egna kommunens eller landstingets samlingsplats för öppna data på www.kommunen.se/psidata⁴.

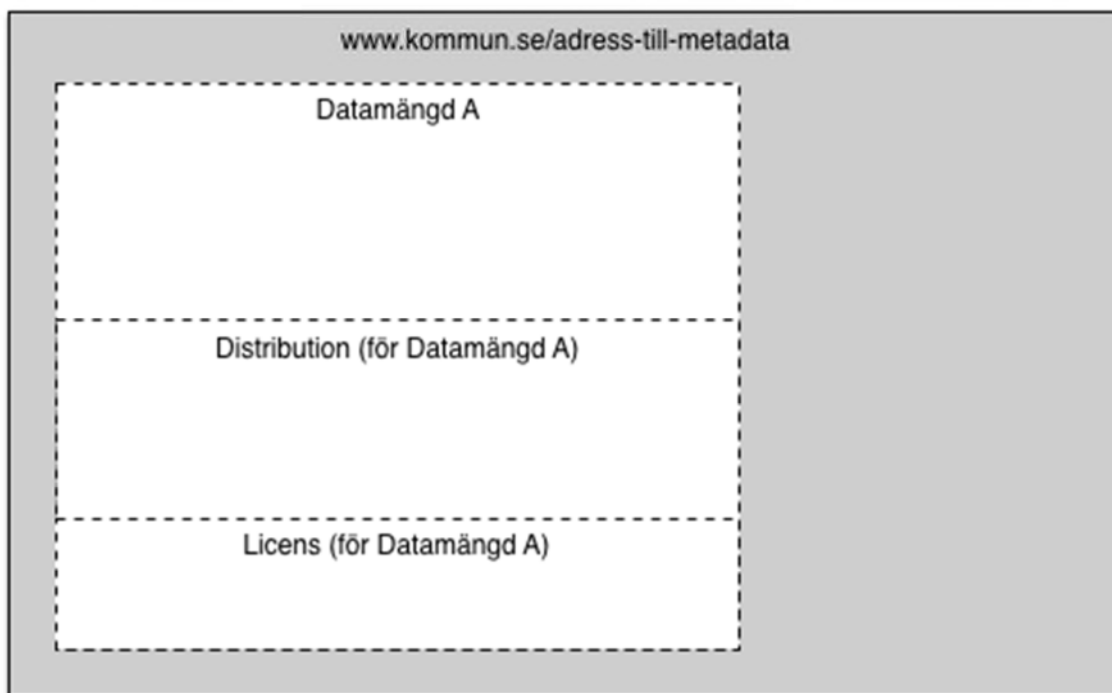
De metadata som beskrivs i enlighet med den här vägledningen kan delas upp på ett flertal webbsidor. Den första sidan är katalogens huvudsida där grundläggande information om katalogen presenteras tillsammans med information om vilken vokabulär som används för att beskriva datamängdernas kategorier, vilka kategorier som finns tillgängliga samt information om vilken organisation (kommun eller landsting) som ansvarar för katalogen. På denna sida finns även en kortfattad beskrivning av de olika tillgängliga datamängderna, exempelvis de obligatoriska egenskaperna samt egenskapen kategori.

³ https://joinup.ec.europa.eu/asset/dcat_application_profile/home

⁴ För landsting blir adressen www.landstinget.se/psidata



Övriga sidor är individuella sidor för metadatan för varje enskild datamängd. Här finns uppgifter om datamängden, dess olika distributioner samt licensvillkor för datamängden.



4. Att beskriva datamängder

Nedan förklaras hur datakatalogens och datamängdernas olika egenskaper kan beskrivas. I den bifogade zip-filen finns två exceldokument. Det ena är för att beskriva datakatalogen, det andra är för att beskriva varje enskild datamängd.

Varje datamängd ska beskrivas i ett eget dokument. Detta gör att ett paket (en zip-fil) med beskrivningar av en kommuns eller landstings datamängder kommer att innehålla ett dokument med beskrivning av datakatalogen samt ett flertal dokument med beskrivningar av individuella datamängder som exemplifieras i nedanstående illustration.



4.1. Datakatalogen

I filen Katalogbeskrivning.xlsx beskrivs uppgifter om den katalog, samlingsplats, som kommunen eller landstinget har för alla sina publicerade datamängder. Denna beskrivning ska endast göras en gång.

I den tillhörande excelmallen är vissa fält markerade med gul färg. Värdena för dessa egenskaper bör endast beskrivas med de fördefinierade termer och begrepp som finns under avsnitt 5 nedan.

4.1.1. Obligatoriska uppgifter

Dessa uppgifter ska alltid finnas med.

Egenskap	Förklaring	Bakgrund
Datamängd	Lämna detta fält tomt. Det är tänkt att användas vid framställning av maskinläsbara beskrivningar för att koppla en katalog till en visst datamängd.	Denna egenskap finns i den europeiska profilen för dataportaler i klassen Catalog med benämningen Dataset.
Beskrivning	Detta ska vara en fritext-beskrivning av datakatalogen som ger en överblick över vem som publicerar den och vad den kan förväntas innehålla.	Denna egenskap finns i den europeiska profilen för dataportaler i klassen Catalog med benämningen Description.
Utgivare	Här anges vilken organisation (kommun eller landsting) som ansvarar för katalogen.	Denna egenskap finns i den europeiska profilen för dataportaler i klassen Catalog med benämningen Publisher.
Titel	Katalogens namn.	Denna egenskap finns i den europeiska profilen för dataportaler i klassen Catalog med benämningen Title.

4.1.2. Rekommenderade uppgifter

Dessa uppgifter ska alltid finnas med om de finns tillgängliga.

Egenskap	Förklaring	Bakgrund
Hemsida	Webbadress till katalogens hemsida	Denna egenskap finns i den europeiska profilen för dataportaler i klassen Catalog med benämningen Homepage.
Språk	Beskriver vilket språk som katalogens metadata (beskrivningarna av katalogen och datamängderna) finns på. Om metadatan finns på flera språk ska alla språk anges med hjälp av de fördefinierade termer och begrepp som beskrivs under 5.2 nedan.	Denna egenskap finns i den europeiska profilen för dataportaler i klassen Catalog med benämningen.
Utgivningsdatum	Det datum då katalogen publicerades första gången.	Denna egenskap finns i den europeiska profilen för dataportaler i klassen Catalog med benämningen Release date.

Licens	Denna egenskap beskriver vilken licens som gäller för dem som vill återanvända metadatan i katalogen, alltså beskrivningarna av katalogen och de enskilda datamängderna. Här ska något av de begrepp som finns nedan under fördefinierade termer och begrepp, 5.1 Licenser, användas.	Denna egenskap finns i den europeiska profilen för dataportaler i klassen Catalog med benämningen License.
Tema	Här anges vilken vokabulär som används för att beskriva datamängdernas kategorier. I detta fall ska värdet vara EuroVoc eftersom det är den vokabulär som rekommenderas i den europeiska profilen för dataportaler.	Denna egenskap finns i den europeiska profilen för dataportaler i klassen Catalog med benämningen Themes.
Uppdateringsdatum	Detta är det senaste datum då katalogen uppdaterades.	Denna egenskap finns i den europeiska profilen för dataportaler i klassen Catalog med benämningen Update/modification date.

4.1.3. Valbara uppgifter

Dessa uppgifter finns inte med i förslaget till europeisk profil för dataportaler eller har angetts som valbara uppgifter i det förslaget.

Egenskap	Förklaring	Bakgrund
Rumslig/ geografisk	Om katalogen innehåller data som är geografiskt avgränsad kan det anges här, antingen genom att ange koordinater eller genom att använda en kontrollerad vokabulär i enlighet med 5.3 nedan.	Denna egenskap finns i den europeiska profilen för dataportaler i klassen Catalog med benämningen Spatial/geografic.

4.2. Datamängd, kategori, licens med mera

I filen Datamängd.xlsx finns två flikar: Datamängd samt Distribution. På fliken Datamängd beskrivs alla uppgifter om *en* datamängd, förutom uppgifterna om hur den distribueras. Observera att det behöver skapas en separat fil för varje datamängds metadata. I den tillhörande excelmallen är vissa fält markerade med gul färg. Värdena för dessa egenskaper bör endast beskrivas med de fördefinierade termer och begrepp som finns under avsnitt 5 nedan.

4.2.1. Obligatoriska uppgifter

Dessa uppgifter ska alltid finnas med.

Egenskap	Förklaring	Bakgrund
Titel	Detta är datamängdens namn.	Denna egenskap finns med i E-delegationens vägledning samt i Vägledande principer för öppna data och benämns där "Namn". Den finns också med i den europeiska profilen för dataportaler i klassen Dataset med benämningen Title.
Beskrivning	Detta ska vara en fritext-beskrivning av datamängden som ger en överblick över vem som publicerar den, vad den kan förväntas innehålla och hur den är strukturerad.	Denna egenskap finns med i E-delegationens vägledning samt i Vägledande principer för öppna data. Den finns också med i den europeiska profilen för dataportaler i klassen Dataset med benämningen Description.
Namn	Här anges namnet på den förvaltning eller nämnd som står bakom informationen i datasetet.	Denna egenskap finns i den europeiska profilen för dataportaler i klassen Agent med benämningen Name.
Etikett	Här anges vilket kategori som datasetet tillhör. Kategorierna väljs bland dem som listas under 5.5 nedan.	Denna egenskap finns i den europeiska profilen för dataportaler i klassen Category med benämningen Preferred label.
Titel	Här anges vilken vokabulär som används för att beskriva datamängdernas kategorier. I detta fall ska värdet vara EuroVoc eftersom det är den vokabulär som rekommenderas i utkastet till europeisk profil för dataportaler.	Denna egenskap finns i den europeiska profilen för dataportaler i klassen Category scheme med benämningen Title.

4.2.2. Rekommenderade uppgifter

Dessa uppgifter ska alltid finnas med om de finns tillgängliga.

Egenskap	Förklaring	Bakgrund
Utgivare	Här anges vilken organisation (kommun eller landsting) som ansvarar för datamängden.	Denna egenskap finns med i Vägledande principer för öppna data. Den finns också med i den europeiska profilen för dataportaler i klassen Dataset med benämningen Publisher.
Kontakt	Här anges kontaktuppgifter antingen till en person eller en funktion som kan svara på frågor kring datamängden.	Denna egenskap finns med i E-delegationens vägledning samt i Vägledande principer för öppna data. Den finns också

		med i den europeiska profilen för dataportaler i klassen Dataset med benämningen Contact point.
Distribution	Här kan man ange en länk till en sida som innehåller information om hur datamängden distribueras eller infoga informationen under 4.3 nedan.	Denna egenskap finns i den europeiska profilen för dataportaler i klassen Dataset med benämningen Dataset distribution.
Teman/kategori	Här anges vilket kategori datamängden tillhör. Kategorierna väljs bland dem som listas under 5.5 nedan.	Denna egenskap finns med i Vägledande principer för öppna data med benämningen "Verksamhetsområde". Den finns också med i den europeiska profilen för dataportaler i klassen Dataset med benämningen Theme/category.
Nyckelord	Här anges ett nyckelord som beskriver datamängden.	Denna egenskap finns i den europeiska profilen för dataportaler i klassen Dataset med benämningen Keyword/tag
Utgivartyp	Här anges vilken typ av utgivare i enlighet med 5.4 nedan som står bakom datakatalogen och de datamängder som publiceras där.	Denna egenskap finns i den europeiska profilen för dataportaler i klassen Agent och med benämningen Publisher type. Den identifierades också som en viktig egenskap i den workshop som hölls i samband med Offentliga rummet 2013.
Licens	Här anges vilka villkor som gäller för att få återanvända den aktuella datamängden. Villkoren beskrivs med begrepp i enlighet med 5.1 nedan.	Denna egenskap finns med i E-delegationens vägledning samt i Vägledande principer för öppna data. Den finns också med i den europeiska profilen för dataportaler i klassen License document med benämningen License type.

4.2.3. Valbara uppgifter

Dessa uppgifter finns inte med i förslaget till europeisk profil för dataportaler eller har angetts som valbara uppgifter i det förslaget.

Egenskap	Förklaring	Bakgrund
Uppdateringsfrekvens	Här anges hur ofta datamängden uppdateras med hjälp av de fördefinierade termerna under 5.6 nedan.	Denna egenskap finns med i E-delegationens vägledning samt i Vägledande principer för öppna data. Den finns också med i den europeiska profilen för dataportaler i klassen

		Dataset med benämningen Frequency.
Utgivningsdatum	Det datum då datamängden publicerades första gången.	Denna egenskap finns med i Vägledande principer för öppna data. Den finns också med i den europeiska profilen för dataportaler i klassen Dataset med benämningen Release date.
Avgift	Om det finns någon avgift som måste betalas för att kunna återanvända datamängden så beskrivs den avgiften här.	Denna egenskap finns med i Vägledande principer för öppna data.

4.3. Distribution

I filen Datamängd.xlsx finns två flikar: Datamängd och Distribution. På fliken Distribution beskrivs hur datamängden distribueras, alltså hur någon som vill återanvända den kan komma åt den. Det kan finnas flera distributioner av en och samma datamängd. Det betyder att det går att fylla i flera rader på denna flik; en rad för varje distribution. Observera att det behöver skapas en separat fil för varje datamängds metadata. I den tillhörande excelmallen är vissa fält markerade med gul färg. Värdena för dessa egenskaper bör endast beskrivas med de fördefinierade termer och begrepp som finns under avsnitt 5 nedan.

4.3.1. Obligatoriska uppgifter

Dessa uppgifter ska alltid finnas med.

Egenskap	Förklaring	Bakgrund
Adress	Här anges den adress, URL, som den aktuella distributionen nås via.	Denna egenskap finns med i Vägledande principer för öppna data. Den finns också med i den europeiska profilen för dataportaler i klassen Distribution med benämningen Access URL.

4.3.2. Rekommenderade uppgifter

Dessa uppgifter ska alltid finnas med om de finns tillgängliga.

Egenskap	Förklaring	Bakgrund
Beskrivning	Denna egenskap är en fritext-beskrivning av distributionen.	Denna egenskap finns i den europeiska profilen för dataportaler i klassen Distribution med benämningen Description.
Format	Denna egenskap beskriver distributionens format. I möjligaste mån bör de fördefinierade termerna under 5.7 nedan användas.	Denna egenskap finns med i E-delegationens vägledning samt i Vägledande principer för öppna data. Den finns också med i den europeiska profilen för dataportaler i klassen Distribution med benämningen Format.
Licens	Här anges vilka villkor som gäller för att få återanvända den aktuella distributionen. Villkoren beskrivs med begrepp i enlighet med 5.1 nedan.	Denna egenskap finns med i E-delegationens vägledning samt i Vägledande principer för öppna data. Den finns också med i den europeiska profilen för dataportaler i klassen Distribution med benämningen License.

4.3.3. Valbara uppgifter

Dessa uppgifter finns inte med i förslaget till europeisk profil för dataportaler eller har angetts som valbara uppgifter i det förslaget.

Egenskap	Förklaring	Bakgrund
Storlek	Här anges storleken på datamängden i bytes.	Denna egenskap finns med i Vägledande principer för öppna data. Den finns också med i den europeiska profilen för dataportaler i klassen Distribution med benämningen ByteSize.
Uppdateringsdatum	Detta är det senaste datum då distributionen uppdaterades.	Denna egenskap finns med i E-delegationens vägledning samt i Vägledande principer för öppna data. Den finns också med i den europeiska profilen för dataportaler i klassen Distribution med benämningen Update/modification date.
Begränsningar	Om det finns kända begränsningar, exempelvis gällande framtida tillgänglighet, så beskrivs de begränsningarna här med hjälp av begreppen under 5.8 nedan.	Denna egenskap finns med i E-delegationens vägledning samt i Vägledande principer för öppna data. Den motsvaras närmast av egenskapen Status i klassen Distribution i den europeiska profilen för dataportaler.
Mognad	Här anges en gradering av distributionens mognad enligt den så kallade femstjärne-modellen som beskrivs under 5.9 nedan.	Denna egenskap finns med i Vägledande principer för öppna data.

5. Fördefinierade termer och begrepp

5.1. Licenser

För att beskriva licenser bör i första hand ADMS license type vocabulary användas⁵. De licenser som specificeras där är:

Erkännande

Den som återanvänder datamängden måste uppge varifrån informationen man använder kommer ifrån. Man måste ge ett erkännande. Ett exempel på en licens som skulle kunna betecknas med denna term är Creative Commons CC-BY.

Ingen upphovsrätt / Public Domain

Detta begrepp används om datamängden inte omfattas av upphovsrätt. Detta är det vanligaste vad gäller offentliga data. Begreppet går också att använda för att beteckna Creative Commons Zero CC-0.

Viral effekt (alias Dela lika)

Den som återanvänder datamängden måste dela vidare det man framställer med samma villkor. Exempel på en licens som skulle kunna betecknas med denna term är Creative Commons Share alike CC-BY-SA även om en sådan licens bäst betecknas med att detta begrepp används tillsammans med begreppet Erkännande.

Dela lika / copyleft - inte kompatibel / interoperabel med andra copyleft licenser

Detta begrepp används i de fall datamängdens licens kräver att de som återanvänder datamängden licensierar det de framställer med samma licens även om datamängden bara bidrar med en liten del till det slutliga resultatet. Exempel på sådana licenser är GPL, AGPL och OSL.

Dela lika / copyleft på källkod eller med kompatibilitet undantaget större arbeten och interoperabilitet

Detta begrepp används i de fall datamängdens licens kräver att de som återanvänder datamängden licensierar det de framställer med samma licens. Dock görs undantag från detta om det krävs för att uppnå kompatibilitet.

1. Kravet omfattar bara källkod och inte binära filer. Exempel på sådana licenser är: LGPL, EPL, APSL 2.0, CDDL och Ms-PL.
2. Kravet omfattar både källkod och binära filer men tillåter att datamängden används som en del i ett större sammanhang om detta licensieras under vissa kontrollerade licenser. Exempel på sådana licenser är EUPL v1.1 och MPLv2.

Endast för icke-kommersiellt bruk

Detta begrepp ska användas om datamängden endast får användas för icke-kommersiellt bruk. Det motsvarar exempelvis Creative Commons Non Commercial CC-BY-NC även om en sådan licens bäst betecknas med att detta begrepp används tillsammans med begreppet Erkännande.

⁵ http://joinup.ec.europa.eu/asset/adms/asset_release/adms

Inga bearbetningar tillåtna

Om datamängden endast får användas i sitt ursprungsformat, det vill säga att det endast är tillåtet att kopiera datamängden och exponera datamängden i den form den är utan att göra några bearbetningar, så ska denna term användas. Den motsvaras av Creative Commons No Derivates CC-BY-ND även om en sådan licens bäst betecknas med att detta begrepp används tillsammans med begreppet Erkännande.

Royalties krävs

Detta begrepp används i de fall det finns en avgift av något slag som måste betalas för att återanvända datamängden.

Skyddat namn / Godkännande krävs / officiell status

Denna term betecknar att utgivaren bakom datamängden har ett skyddat namn som inte får användas utan utgivarens tillstånd. Det kan till exempel handla om att förhindra att någon förespeglar att en tjänst de framställer med hjälp av en datamängd har en officiell status. Detta villkor torde vara svårt för svenska kommuner och landsting att sätta upp för återanvändare av offentliga data.

Självkostnad

Detta begrepp används om datamängden får återanvändas mot en mindre avgift högst motsvarande självkostnad för upprätthållandet och distributionen av datamängden.

Ge tillbaka

Detta begrepp beskriver att det finns krav på att den som återanvänder datamängden ska ge tillbaka en rättighet till den som publicerat datamängden att använda och vidareutveckla resultatet av återanvändarens arbete.

Inom europeisk domvärjo

Begreppet anger att det i licensen finns ett krav på att tvister ska avgöras enligt rättspraxis från en av EU:s medlemsstater och i vissa fall även av en domstol som ligger inom EU. Exempel på detta är artikel 15 i EUPL v1.1.

Andra restriktiva villkor

Detta begrepp används för att beskriva att det finns andra restriktiva villkor än de som tagits upp ovan.

Kända patentbegränsningar

Om det utöver licensvillkor även krävs att den som vill återanvända ett material också följer vissa patentvillkor anges det med denna term.

Okänd upphovsrättsstatus

Om datamängdens upphovsrättsliga status är okänd anges det med detta begrepp.

5.2. Språk

Språk anges i enlighet med de definitioner som återfinns i den Europeiska Publikationsbyråns Metadata Registry (MDR) över auktoriteter.

Språket ska anges i klartext (linguistic version) och i enlighet med ISO 639-3. Exempelvis: svenska (swe).

En fullständig lista finns tillgänglig på <http://publications.europa.eu/mdr/authority/language/index.html>.

5.3. Rumslig/geografisk (kommuner och landsting)

Kommuner och landsting bör anges med namnet i klartext tillsammans med SCB:s kodnummer för kommunen/ landstinget. Exempelvis: Nacka kommun (0182) eller Västra Götalands län (14).

Kodnumren finns tillgängliga på http://www.scb.se/Pages/List_257281.aspx.

5.4. Utgivartyp

För att beskriva licenser bör i första hand ADMS publisher type vocabulary användas⁶. De utgivartyper som specificeras där är:

Akademi/ Vetenskaplig organisation	Ideell organisation
Företag	Privatperson
Industrikonsortium	Regional myndighet
Kommunal myndighet	Standardiseringsorgan
Nationell myndighet	Överstatligt organ
Icke-offentlig organisation	

5.5. Teman/kategori

För att beskriva vilken kategori eller vilket tema en viss datamängd tillhör används de 21 ämnesområdena i EuroVoc⁷.

Politik (04)	Sysselsättning och arbete (44)
Internationella förbindelser (08)	Transport (48)
Europeiska unionen (10)	Miljö (52)
Lag och rätt (12)	Jordbruk, skogsbruk och fiske (56)
Ekonomi (16)	Jordbrukslivsmedelsindustri (60)
Handel och affärsverksamhet (20)	Produktion, teknik och forskning (64)
Finanser (24)	Energi (66)
Sociala frågor (28)	Industri (68)
Utbildning och kommunikation (32)	Geografi (72)
Vetenskap (36)	Internationella organisationer (76)
Företag och konkurrens (40)	

Skriv ut både kategorins namn i klartext samt kategorins nummer inom parentes.

⁶ http://joinup.ec.europa.eu/asset/adms/asset_release/adms

⁷ <http://eurovoc.europa.eu>

På <http://eurovoc.europa.eu> går det att söka på olika begrepp för att få fram vilken kategori en viss datamängd ska inordnas under. En sökning på "förskola" ger exempelvis resultatet "MT 3211 Utbildningsväsen". De två första siffrorna indikerar att det är kategori 32 Utbildning och kommunikation som är den rätta i det här fallet.

Det kan vara bra att klicka sig igenom sökresultatet för att se vilka begrepp som hör hemma i en kategori för att vara säker på att datamängden hänförs till rätt kategori.

5.6. Uppdateringsfrekvens

Beskrivningen av uppdateringsfrekvens ska utgå från Dublin Core Collection Description Frequency Vocabulary⁸.

Vart tredje år	Varannan vecka
Vartannat år	Tre gånger i månaden
Årligen	Veckovis
Två gånger om året	Två gånger i veckan
Tre gånger om året	Tre gånger i veckan
Kvartalsvis	Dagligen
Varannan månad	Kontinuerligt
Månadsvis	Oregelbundet
Två gånger i månaden	

5.7. Filformat

Filformat anges så långt det är möjligt i enlighet med de definitioner som återfinns i den Europeiska Publikationsbyråns Metadata Registry (MDR) över auktoriteter.

Filformatet ska anges med det värde som finns i kolumnen "Label" vid <http://publications.europa.eu/mdr/resource/authority/file-type/html/filetypes-eng.html>. Exempelvis: Word DOCX för en fil från en nyare version av MS Word eller PDF/A-1a för en PDF-fil som följer Riksarkivets föreskrift vad gäller elektroniska dokument för långtidsbevarande.

En svaghet i Publikationsbyråns förteckning är att den saknar svensk översättning samt att vissa vanliga format för öppna data, exempelvis CSV, saknas.

Begreppet eBook bör översättas till e-bok, resultat bör översättas till resultat och Plain Text bör översättas till Text.

Vad gäller CSV-filer så bör de betecknas: Text CSV. Övriga filer som saknar beskrivning bör ges en beskrivning som är konstruerad på samma sätt som de moderna MS Officeformaten, det vill säga originalprogramvarans namn samt filändelsen exempelvis Open Document Spreadsheet ODS.

Flera distributioner av förteckningen över filformat finns på <http://publications.europa.eu/mdr/authority/file-type/index.html>.

⁸ <http://dublincore.org/groups/collections/frequency>

⁹ http://joinup.ec.europa.eu/asset/adms/asset_release/adms

5.8. Begränsningar

För beskrivningen av datamängdens begränsningar bör i första hand ADMS status vocabulary användas⁹. Vokabulären specificerar inte begränsningar fullt ut utan beskriver endast en resurs status med följande begrepp:

Färdigställd	Under utveckling
Föråldrad	Återkallad

5.9. Mognad

Beskrivningen av en datamängds mognad görs i enlighet med den så kallade femstjärnmodellen. Modellen erbjuder vidareutnyttjare av data ett enkelt sätt att bedöma kvaliteten på olika datamängder. Det är också ett bra verktyg för att skapa nyckeltal för det egna kvalitetsarbetet samt för att jämföra den egna verksamheten med andras.

Varje datamängd klassificeras individuellt och antalet stjärnor anges med antalet stjärnor samt en beskrivning av antalet stjärnor i klartext exempelvis: ★ ★ ★ (tre stjärnor).

De olika stegen i femstjärnmodellen¹⁰:

Beskrivning	Antal stjärnor
Data publicerad på webben med en öppen licens	★
Maskinläsbar, strukturerad data	★ ★
Data i ett öppet format (t.ex. CSV istället för Excel)	★ ★ ★
Använd URI:er som identifierare så att andra kan peka på dina data	★ ★ ★ ★
Data sammanlänkade med andra datamängder	★ ★ ★ ★ ★

Exempel:

En datafil med befolkningsstatistik i CSV-format med licensen CC-BY ger tre stjärnor.

En datafil med ostrukturerad information om lekparkar i en PDF-fil med licensen CC-BY ger en stjärna.

¹⁰ <http://5stardata.info>