

Rapport

# Dataprodukt och dataproduktspecifikation

ES2024-11





## Innehållsförteckning

1.	Inledning.....	4
1.1	Målgrupp .....	4
1.2	Bakgrund .....	4
1.3	Syfte.....	4
2.	Data som produkt.....	5
2.1	Datadriven verksamhet .....	5
2.2	Dataprodukt.....	5
2.2.1	Uppbyggnaden av en dataprodukt .....	6
2.3	Dataproduktspecifikation (DPS).....	8
2.3.1	Innehåll i en dataproduktspecifikation .....	9
2.3.2	Exempel på dataproduktspecifikation .....	12
2.4	Dokumentering av begrepp och information .....	13
2.4.1	Begreppsmodellering.....	13
2.4.2	Informationsmodellering.....	14
2.5	Begreppsmodell över dataprodukt och dataproduktspecifikation.....	14



# 1. Inledning

## 1.1 Målgrupp

Denna beskrivning riktar sig i första hand till arkitekter inom en organisation som avser att utveckla sin datahantering för att effektivare kunna katalogisera samt dela data inom och utanför den egna organisationen.

## 1.2 Bakgrund

Hanteringen av data är fragmenterad mellan myndigheterna i Sverige. Ökade krav på tillgång till data och på själva data driver oss myndigheter mot en mer harmoniserad syn på hur vårt organisationsdata ska hanteras och även kunna utbytas mellan organisationerna. I ljuset av detta så har eSam tagit fram en referensarkitektur för hantering av data för att stödja en mer gemensam datahantering för organisationsverige, i referensarkitekturen så är dataprodukt och dataproduktspecifikation en central komponenterna.

## 1.3 Syfte

Syftet är att ge stöd och vägledning avseende vad konceptet med data som produkt innebär, detta för att få en gemensam bild över beskrivningen av information. Det framkommer även hur dataprodukter kan dokumenteras samt hur kartläggning av begrepp och information utförs. Ambitionen med denna beskrivning är att ge ett generellt perspektiv, skapa förståelse samt guidning av användande inom en organisation.



## 2. Data som produkt

### 2.1 Datadriven verksamhet

Inom eSam gjordes det 2023 ett arbete ([länk](#)) för att förklara grunderna för en datadriven organisation. De konstaterade att en datadriven verksamhet:

- har en systematik för att leta lösningar på problem och utmaningar genom att **utforska kvalitetssäkrad data** för att förstå möjliga scenarion och skaffa sig insikter.
- använder insikterna för att utveckla, utvärdera, och implementera nya lösningar, leveranser, produkter, arbetsätt, m.m.
- utvecklar och driver sin verksamhet genom att systematiskt och innovativt använda **kvalitetssäkrad data**.

Fortsatt beskrivs att en **korrekt och effektiv informations- och datahantering** är ett grundfundament för en analytisk förmåga och kultur.

Konceptet med en datadriven verksamhet fortsätter med att realisera verksamhetsuppföljning och analyser genom diverse tekniker, såsom datalager och data warehouse.

(<https://www.esamverka.se/intranat/grupper/filer.html?file=18.63a599d187e0b4119866a5>)

### 2.2 Dataprodukt

#### **Definition**

Datamängd eller serie av datamängder som erbjuds en konsument

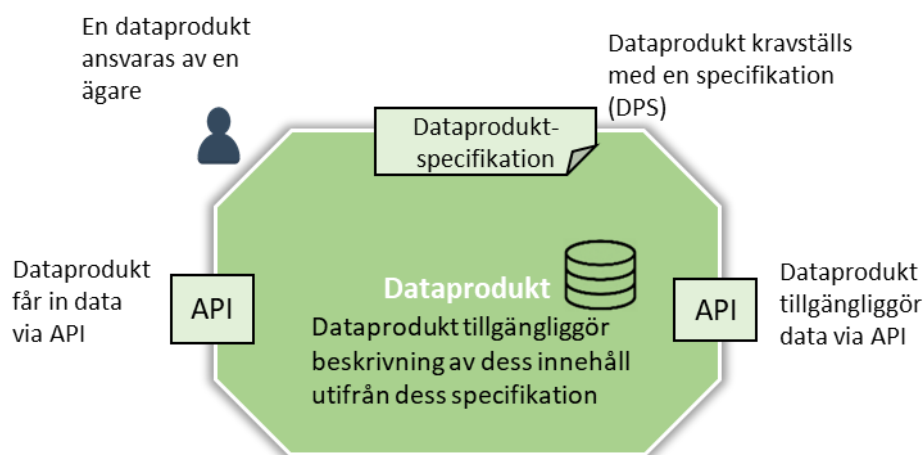
Detta arbete tar fokus på att dokumentera den data som senare ska analyseras. En ansats för att förtydliga detta är via konceptet *data som produkt*. Via ett produkttänk skapas en förståelse för vad datat innehåller och på bästa sätt kan förvaltas och användas av konsumenterna.



Att strukturera data som produkt, eller som en dataproduct, grundar sig i konceptet Data Mesh. (Denna rapport har tagit inspiration av Data Mesh, men inte följt dess koncept rakt av.) Det centrala med att beskriva data som dataproduct är att skapa en väl definierad, och för konsumenten, lättanvänd delning av data. Datat ska ses självständig entitet och behandlas oberoende av annan data. Dock kan flera dataproducter kombineras för att i sin tur skapa nya dataproducter.

Enkelt förklarar så skapar dataproducter en högre flexibilitet och ett högre värde då flera konsumenter kan använda samma mängd data, detta inom olika användningsområden.

## 2.2.1 Uppbyggnaden av en dataproduct



### 2.2.1.1 Principer för en dataproduct

- En Dataproduct har *en* ägare  
För att säkerställa förvaltning av dataproducten ska en person med kunskap inom informationsmängden, dess krav och tillgänglighet finnas, som säkerställer dess kvalitet över hela dess livscykel.
- En Dataproduct styrs utifrån *en* specifikation  
Beskrivningen av vad dataproducten innehåller ska framkomma via en specifikation. I specifikationen ska intressenter läsa sig till syfte, mål och innehåll. Se exempel på hur en dataproductspecifikation kan byggas upp i kapitlet *Dataproductspecifikation (DPS)*.
- En Dataproduct tillgängliggörs via standardiserade gränssnitt  
Dataproducten ska integrera via standardiserade gränssnitt via API:er, som



möjliggör ett utbyte av data mellan aktörer. Dataprodukten möter då upp en API-specifikation som beskriver hur den tekniska realiseringen ska se ut.

- En Dataprodukt tillhör ett specifikt område och sammanhang  
Användandet av dataprodukten ska förtydligas utifrån dess användningsområde (en så kallad domän) i syfte att förstå inom vilket avgränsning som den hanteras inom. Det ska även framgå dess ursprungliga syfte där det går att härleda till hur den ska hanteras. Observera att detta inte motsäger sig att dataprodukten används inom andra områden än de initialt var menade till.

### 2.2.1.2 Krav på en dataprodukts uppbyggnad

- Välbeskriven  
Det ska tydligt framkomma vad innehållet och syftet med dataprodukten är, detta så att konsumenten kan förstå hur den är uppbyggd och hur den är menad att användas. Med en god beskrivning ökar även möjligheten att dataprodukten nyttjas av flera konsumenter i flera olika kontexter.
- Tillförlitlig  
Dataprodukten ska ha en känd kvalitet i linje med beslutade kvalitetsegenskaper. Den ska även vara spårbar utifrån dess ursprungliga källa. Tillförlitligheten ger konsumenten en trygghet i att data återspeglar verkligheten och därmed minskar risken för felaktig hantering och analys av data.
- Sökbar  
Det ska vara lätt för konsumenter att hitta en dataprodukt, detta för att skapa ett så stort värde som möjligt genom användning. Användandet av metadata katalog för sökbarhet, eller centraliserad datakatalog för hantering av data, kan medföra en förbättrad sökbarhet och nyttjande.
- Semantisk och teknisk interoperabel  
Det ska finnas möjlighet att dataprodukt samverkar, inom och utanför en avgränsad domän. För att skapa möjligheten till datautbyte ska utpekade standarder och kodverk användas till största utsträckning.  
För att det ska gå att realisera samverkan rent tekniska ska kända tekniska gränssnitt användas, detta utifrån beslutade format. En permanent och unik identifiering ska finnas för att försäkra om åtkomst genom hela datalivscykeln.
- Tydligt värde  
En dataprodukt ska vara strukturerad så att den kan vara meningsfull i sig, och



ha möjlighet att agera självständigt. Samtidigt ska den vara ändamålsenlig som ger ett tydligt värde mot konsumenten.

## 2.3 Dataproduktspecifikation (DPS)

### Definition

Specifikation av en dataprodukt tillsammans med ytterligare information som möjliggör att densamma kan skapas, distribueras och användas av tredje part

En dataproduktspecifikation (DPS) är paketeringen av den dataprodukt som avses. Även om en standardisering av paketeringen eftersträvas, kan behovet av detaljeringsgrad variera utifrån verksamhetens behov och datamängdens användningsområde.

DPS blir den dokumentation över de krav som ska uppfyllas för att nå den datakvalitet som efterfrågas. DPS ligger som underlag till realiseringar i form av API-specifikationer och en indikator över om dataprodukten uppnår uppsatta krav.

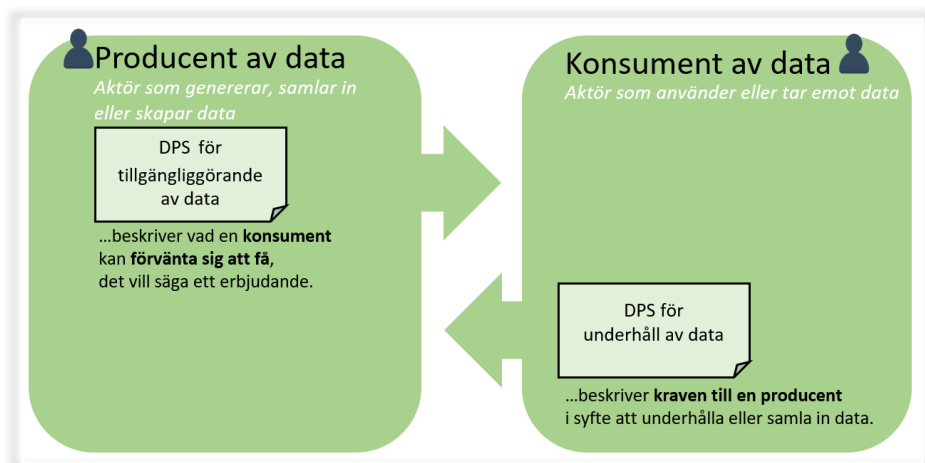
DPS är en typ av specifikation för att förtydliga överenskomet innehåll vid informationsutbyte mellan producent och konsument. **När det gäller att tillgängliggöra data via Sveriges Dataportal ska de riktlinjer som Digg tagit fram följas.** Mer information finns i följande länk: <https://docs.dataportal.se/specifications/>

DPS kan användas utifrån två perspektiv:

1. utifrån ett konsumentperspektiv där en beskrivning finns över vad som en konsument kan förvänta sig få för data
2. utifrån ett producentperspektiv där en kravställning tas fram på den data som efterfrågas



Oavsett om användningen sker utifrån perspektiv 1 eller 2 kan innehållet inom en DPS användas.



### Relation mellan Dataproduktspecifikation och API-specifikation

Både DPS och API-specifikation är dokumentation som ger anvisning till de kraven som ska uppfyllas för en specifik datamängd. Skillnaden ligger i att en DPS beskriver innehållet i datamängden och dess syfte samt användningsområde, detta för att skapa en förståelse för producent eller konsument. En API-specifikation beskriver dess tekniska implementation om hur datamängden ska tillhandahållas, detta för att möjliggöra implementationer för utvecklare. DPS kan ligga till grund för hur en API-specifikation sätts upp. Observera att användandet av dessa två specifikationer kan variera mellan organisation och kontext.

### Relation mellan Dataproduktspecifikation och Förvaltningsgemensamma specifikationer (FGS)

Till skillnad från en Dataproduktspecifikation, har FGS syfte att standardisera informationen vid överföring specifikt vid arkivering. En FGS beskriver tekniskt en generell överföring av en viss typ av information. Läs mer om detta i följande länk: <https://riksarkivet.se/intro-fgs>

#### 2.3.1 Innehåll i en dataproduktspecifikation

Kapitlet ger inspiration till hur en DPS kan byggas upp. Dock är syftet och användningsområde alltid styrande för innehållet.





### 2.3.1.1 Bakgrundsinformation om dataproduktspecifikationen

Beskrivning av dokumentet och inte själva datamängden. Information som kan inkluderas är: specifikationens namn, språk, kontakt, e-post, telefon, länk till specifikation, skyddsbehov etc.

### 2.3.1.2 Syfte och identifiering av dataprodukt

- Dataproduktens namn / alternativt namn  
Ange dataproduktens namn och ett eventuellt alternativt namn existerar
- Syfte  
Beskrivning av syftet med dataprodukten, det vill säga vad den är tänkt att användas till; varför existerar den?
- Användningsfall  
Syftet med datamängden kan även beskrivas genom användningsfall. Detta kan ge ett förtydligande i hur dataprodukten förhåller sig till sitt värdeflöde.
- Ämnesområde / domän  
Ämnesområdet beskriver på en övergripande nivå vilken information som finns i dataprodukten. Beskriver inom vilken kontext som dataprodukten berör.
- Avgränsning för användning  
Här anges eventuella begränsningar, juridiska aspekter, licenser och dylikt som gäller för dataprodukten.
- Informationsägandeskap  
Beskrivning av det ägandeskap som ansvarar för dataproduktens innehåll.

### 2.3.1.3 Datainnehåll och struktur

Här beskrivs datamängdens innehåll tillsammans med begreppsmodell, informationsmodell och objekttypskatalog (de senare finns alltid som bilaga).

- Datamängdens innehåll  
Här skriver du in en detaljerad och utförlig beskrivning av innehållet i datamängden. Detta är till för de som inte har förmågan eller viljan att läsa informationsmodellen och/eller objekttypskatalogen.  
Texten kan utformas på olika sätt, och det går bra att hänvisa till objekttypskatalogen för att hitta detaljer. Det är möjligt att utgå från informationsobjekten i informationsmodellen, men oftast är det bättre att



beskriva utifrån ett verksamhetsperspektiv (till exempel utifrån begreppsmodellen).

- Begrepp  
Verksamhetens begrepp beskriven i en begreppsmodell (se *kapitel Begreppsmodellering*)
- Information  
Kravställningen på information beskriven i en informationsutbytesmodell (se *kapitel Informationsmodellering*)
- Objekttypskatalog  
Detaljerad förteckning av de informationsobjekt som beskrivs i informationsutbytesmodellen.
- Sekretessnivå  
Redogörelse för hur informationen är informationsklassad och får spridas inom eller utanför organisationen

#### 2.3.1.4 Kvalitet på data

Här anges vilken kvalitet eller vilka kvalitetskrav som datamängden ska uppfyllas.

Önskvärt är även att ange hur kravet ska mätas (eftersom olika sätt att mäta kan ge olika resultat) samt vilka brister som kan accepteras.

Här finns skillnader mellan om det är en **DPS för underhåll av data**, där beskrivning av krav och kvalitetsmått framförs, och **DPS för tillgängliggörande** av data där en beskrivning av den datakvalitet som en konsument kan förvänta sig.

#### 2.3.1.5 Datainsamling och bearbetning

Med **datainsamling/bearbetning** avses de metoder/behandlingar som görs för att skapa digital information. I normala fall är tillägg av nya objekt i en datamängd en datainsamling/bearbetning, men även metoden för att uppdatera objekt kan räknas som datainsamling/bearbetning.

- Datainsamling / bearbetning  
Här beskrivs hur och när informationen samlades in, och eventuella bearbetningar som har gjorts.



### 2.3.1.6 Underhåll av data

Med **underhåll av data** avses en beskrivning av hur det som finns i datamängden hålls uppdaterat. Exempel på vad som kan beskrivas är hur felaktigheter uppdateras eller hur kommunikation sker om data ändras.

- **Beskrivning**  
En beskrivning av hur data underhålls.
- **Underhållsfrekvens**  
En beskrivning av hur ofta datamängden uppdateras. Ex kontinuerligt eller vid ett tidsintervall.

### 2.3.1.7 Leverans av data

Här beskrivs hur dataprodukten tillgängliggörs. Här kan relationen till en eventuell API-specifikation tydliggöras.

## 2.3.2 Exempel på dataproduktspecifikation

**Om dokumentet**  
**Specifikationens namn:** Dataproduktspecifikation för gatlampor  
**Språk:** Svenska  
**Kontakt:** ...  
**Länk till denna version av dokumentet:** <http://www.organisationA.tld/dps/gatlampor-v1.1/>  
**Länk till den senaste versionen av detta dokument:** <http://www.organisationA.tld/dps/gatlampor-v1.1/>  
**Skyddsbehov:** Inget skyddsbehov

**Syfte och identifiering av dataprodukten**  
**Dataproduktens namn:** Gatlampor  
**Alternativt namn:**  
**Syfte med dataprodukten:**  
Syftet med att tillgängliggöra informationen om gatlampor är att göra det möjligt för tredje part att utveckla applikationer som allmänheten kan använda för att rapportera trasiga gatlampor. Det här innebär att gatlampornas position inte är särskilt exakt, utan syftar snarare till att den som äger gatlampan ska komma tillräckligt nära för att se vilken som inte lyser. Det är därmed olämpligt att använda positionen som underlag för byggande, men skulle däremot kunna fungera i sammanhanget eftersök vid påkörning av viit.  
**Ämnesområde:** Gatlampor  
**Begränsning av användning:** Inga begränsningar finns för användningen. Licensformen är CC0.

**Datainnehåll och struktur**  
**Datamängdens innehåll**  
Datamängden innehåller följande information:

- Gatlampans ungefärliga position
- Vem som äger gatlampan
- Huruvida lampan lyser eller inte
- Om lampan ska vara släckt eller ej (för att spara energi)

**Informationsutbytesmodell:**  
Informationsutbytesmodellen består endast utav en objekttyp, gatlampan.

Gatlampan
position: geometry agare: string trasig: boolean nedslockt: boolean



Objekttypskatalog

Gatlampan

**Beskrivning:** Lampa utefter väg eller stig.

**Attribut:**

Attributnamn	Datotyp	Beskrivning
position	Geometri	Gatlampans läge angiven i SWEREF 99 TM.
ägare	Text	Organisation som äger gatlampan och ansvarar för dess underhåll.
trasig	boolean	Flagga som talar om huruvida lampan har rapporterats som trasig.
nedsläckt	boolean	Flagga som talar om huruvida lampan är nedsläckt av ägaren, till exempel för att spara energi eller annan orsak.

**Kvalitet på data**

Den här dataproducten ska uppfylla följande krav:

- Positionen på gatlampan ska ha en lägesosäkerhet på högst 50 meter
- Minst 90% av gatlamporna ska finnas

**Datainsamling och bearbetning**

Informationen kommer från ägare och de som underhåller gatlampor, till exempel kommuner. Informationen kan även komma från allmänheten, och ingen kontroll av dessa uppgifters korrekthet görs.

**Underhåll av data**

Informationen uppdateras kontinuerligt av såväl ägare av gatlamporna som allmänheten.

**Leverans**

...

## 2.4 Dokumentering av begrepp och information

Inom ramen för en dataproductspecifikation finns behov av att åskådliggöra centrala begrepp samt hur information är strukturerad.

Detta är relevant både vad det gäller både för *tillgängliggörande av data* där det finns en konsumerande part och *underhåll av data* där det finns en producerande part.

### 2.4.1 Begreppsmodellering

En begreppsmodell beskriver viktiga företeelser som det krävs samsyn kring för att kunna resonera om. Begreppsmodellen använder verksamhetens språkbruk och definierar en samling begrepp samt deras inbördes relationer utifrån ett visst sammanhang. Såväl begreppen som relationerna förklaras med beskrivande text (definitioner). Begreppsmodellen

- utgör en abstraktion, dvs. en förenklad och mindre detaljerad bild, av informationsmodellen.
- är helt oberoende av om/hur dess ingående begrepp hanteras rent tekniskt via IT-stöd.

#### 2.4.1.1 Begreppsmodellering inom ramen för DPS

Begreppsmodellens syfte är att för dataproducten definiera begrepp och dess relationer till andra begrepp, både inom den egna och möjligtvis andra dataproducter. Begreppsmodellerna har även ett syfte att skapa en gemensam terminologi.



För riktlinjer kring hur begreppsmodellering ska utföras hänvisas till framtaget material av Digg:

[Ramverk för nationella grunddata inom den offentliga förvaltningen – bilaga informationsarkitektur.](#)

## 2.4.2 Informationsmodellering

En informationsmodell definierar en samling informationsobjekt, deras egenskaper samt inbördes relationer utifrån ett visst sammanhang. Såväl informationsobjekten som attributen och relationerna förklaras med beskrivande text. Informationsmodellen är spårbar till och utgör en detaljering av de begrepp i en begreppsmodell för vilka information på olika sätt hanteras inom dataprodukten.

### 2.4.2.1 Informationsmodellering inom ramen för DPS

De informationsmodeller som används inom en dataprodukt är i de flesta fall en informationsutbytesmodell, som är en typ av informationsmodell. En informationsutbytesmodell är anpassad för utbyte av information. Syftet med en informationsutbytesmodell är att visa den information som är tillgänglig via en dataprodukt. Informationsutbytesmodellen kompletterar därmed den dokumentation som finns kring dataprodukten.

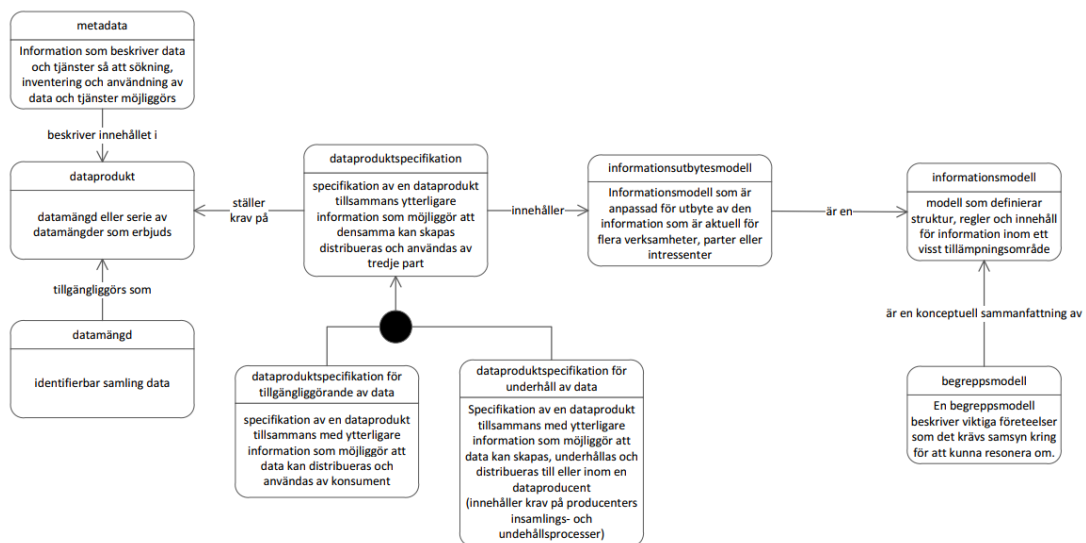
En informationsutbytesmodell kan vara spårbar till motsvarande begreppsmodell.

För riktlinjer kring hur informationsmodellering ska utföras hänvisas framtaget material av Digg:

[Ramverk för nationella grunddata inom den offentliga förvaltningen – bilaga informationsarkitektur.](#)

## 2.5 Begreppsmodell över dataprodukt och dataproduktspecifikation

Nedan illustreras en begreppsmodell som beskriver de ingående modellernas inbördes relationer.



eSam är ett medlemsdrivet program för samverkan mellan myndigheter för att underlätta och påskynda digitaliseringen inom det offentliga. eSam bildades 2015 som en frivillig fortsättning på E-delegationen. En viktig uppgift för eSam är att ta fram stöd och vägledningar som ger förutsättningar för att öka den digitala samverkan inom offentlig förvaltning.

Alla stöddokument finns på [esamverka.se](https://esamverka.se)

I eSam ingår Arbetsförmedlingen, Arbetsmiljöverket, Bolagsverket, Boverket, Centrala Studiestödsnämnden, Domstolsverket, E-hälsomyndigheten, Ekonomistyrningsverket, Finansinspektionen, Folkhälsomyndigheten, Försäkringskassan, Havs- och vattenmyndigheten, Inspektionen för vård och omsorg, Jordbruksverket, Kemikalieinspektionen, Kriminalvården, Kronofogdemyndigheten, Kustbevakningen, Lantmäteriet, Länsstyrelserna, Migrationsverket, Naturvårdsverket, Patent- och Registreringsverket, Pensionsmyndigheten, Riksarkivet, Rättsmedicinalverket, Sida, Skatteverket, Skolverket, Statens institutionsstyrelse, Statens servicecenter, Statens tjänstepensionsverk, Statistiska centralbyrån, Tillväxtverket, Trafikverket, Transportstyrelsen, Tullverket och Universitets- och högskolerådet (okt 2023)

