

| Kontaktperson RISE | Datum | Beteckning | Sida |
|---|------------|-------------------------------|-------|
| Joakim Jakobsson joakim.jakobsson@ri.se 072-452 66 32 | 2020-12-09 | Remissyttrande SOU 2020:55 | 1 (5) |

Remissyttrande RISE Öppna data-utredningens huvudbetänkande Innovation genom information, SOU 2020:55

RISE är ett statligt ägt forskningsinstitut. Våra 2 800 medarbetare driver och stöder alla typer av innovationsprocesser och erbjuder expertis och ett 120-tal test- och demonstrationsmiljöer för framtidssäkra teknologier, produkter och tjänster. Inom en av RISE fem divisioner, divisionen Digitala system, arbetar drygt 450 medarbetare med elektronik, informations-och kommunikationsteknik samt mjukvaruutveckling, mobilitet, systemanalys och interaktionsdesign, cybersäkerhet och artificiell intelligens. RISE bedriver en rad forskningsprojekt i samverkan med offentliga aktörer kring öppna och delade data.

Det är RISE bedömning att vi som organisation enbart omfattas av utredningens förslag om ett krav på att tillgängliggöra forskningsdata som helt eller delvis har tagits fram med offentlig finansiering och gjorts tillgänglig genom ett institutionellt eller ämnesbaserat register.

I beredningen av detta ärende har Joakim Jakobsson, Susanne K Hansson, Björn Flintberg, Anders Berntson, Kristina Knaving, Peter Ljungstrand, Claus Popp Larsen och Daniel Rudmark deltagit.

Inledande kommentar

RISE tillstyrker många av utredningens förslag och delar utredningens antagande att det för samhället som helhet finns stora värden som kan frigöras genom att i högre grad tillgängliggöra data. En reflektion från vår sida är bristen på analys i utredningen kring de utmaningar och medföljande kostnader som förslagen kommer att innebära för många delar av offentlig verksamhet, exempelvis kommuner med begränsad erfarenhet av att tillgängliggöra data där vi bedömer att det kommer krävas både stora resurser ekonomiskt som personalmässigt för att nå målen med lagstiftningen.

Att digitaliseringen tillämpas fullt ut i offentlig sektor är en nödvändighet för en kvalitativ och effektiv service till medborgarna. Digitalisering kan ge effektivare processer, kvalitets- och integritetssäkrad datahantering, bättre beslutsunderlag genom datainsamling och leda till visionen om ett datadrivet samhälle. För att nå dit krävs att vi adresserar de grundläggande tekniska och organisatoriska mekanismerna för att kunna samla in och tillgängliggöra data på ett kontrollerat sätt.

RISE Research Institutes of Sweden AB

| Postadress | Besöksadress | Telefon / Telefax | E-post / Internet | Org.nummer |
|------------------------------------|----------------------------------|-------------------------------|-------------------------|-------------|
| Lindholmen 8077 402 78 GÖTEBORG | Lindholmen 7A 417 56 GÖTEBORG | 010-516 50 00 033-13 55 02 | info@ri.se www.ri.se | 556464-6874 |

| Datum | Beteckning | Sida |
|------------|-------------------------------|-------|
| 2020-12-09 | Remissyttrande SOU 2020:55 | 2 (5) |

Utöver detta vill vi notera att värdet på insamlade data är mycket beroende av tydliga standarder runt insamling (data har samlats in med samma metodik och därmed är jämförbart med annat, liknande data), struktur, format och metadata. Utan tydligt stöd i den frågan riskerar vi en splittrad respons, särskilt från små offentliga aktörer. Vi välkomnar därför förslaget om teknisk reglering.

Det finns potentiellt stora socioekonomiska värden kopplade till att statliga myndigheter tillgängliggör öppna data, och man anar att utredningen främst fokuserar på dessa aspekter. En statlig myndighet har ett begränsat tematisk fokus och därmed ett begränsat antal olika typer av datamängder. En kommun har i jämförelse ett stort antal förvaltningar och bolag med många olika typer av datamängder. Man kan därför generellt förvänta att uppgiften att tillhandahålla öppna data kommer att bli mycket mer omfattande och betungande för en kommun än för en statlig myndighet.

I kommuner finns stora mängder av data som är kopplade till olika verksamhetssystem. Dessa data är oftast svårt tillgängliga trots att de kan skapa värden på andra ställen än för det ursprungliga ändamålet. Kommuner behöver alltså precis som statliga myndigheter skapa mekanismer för att kunna tillgängliggöra data. Incitamentet för kommuner att jobba mer datadrivet är dock inte främst att frigöra socioekonomiska värden för andra aktörer. Det är däremot att med hjälp av data effektivisera sin egen verksamhet samt kunna ta bättre och mer underbyggda beslut. Och den typen av data är ofta känslig och därmed olämplig att publicera som öppna data, exempelvis om den beskriver en infrastruktur eller innehåller personuppgifter. Det är tekniskt och organisatoriskt betydligt mer komplext att tillgängliggöra begränsade data (alltså icke-öppna data) jämfört med öppna data. Men när en kommun har utvecklat förmågan att kunna hantera och tillgängliggöra begränsade data på ett säkert och kontrollerat sätt, då är det enkelt att kunna tillgängliggöra de datamängder som lämpar sig som öppna data. Man kan alltså säga att storskaligt tillgängliggörande öppna data i kommun kommer att vara en biprodukt av att man har lärt sig hantera begränsade data.

Med detta sagt kommer det finnas ett flertal ekonomiska och demokratiska värden kopplade till att kommuner börjar publicera mera öppna data, men det finns en risk för att det i kommunerna fattas ekonomiskt incitament för storskaligt tillgängliggörande av öppna data på det sättet som utredningen och lagförslaget föreställer sig.

I utredningen nämns informationssäkerhet flera gånger i förbindelse med speciellt skyddsvärda verksamheter. I **17.5.4 Informationssäkerhet** står följande: ”Informationssäkerhet är en viktig del av den digitala förvaltningen och innebär en strävan efter att skydda information så att den alltid är tillgänglig när den behövs (tillgänglighet), att det går att lita på att den är korrekt och inte manipulerad eller förstörd (riktighet) samt att endast behöriga personer får ta del av informationen (konfidentialitet)”. Vi anser dock att organisationer måste jobba strukturerat och genomgripande med informationssäkerhet, alltså inte bara inom speciellt skyddsvärda verksamheter. För *alla* typer av data som tillgängliggörs (oavsett om det är som öppna eller begränsade data) måste informationsägaren genomföra en informationsklassning som just inkluderar tillgänglighet, riktighet och konfidentialitet. Att kunna genomföra en informationsklassning är en helt central men generellt kraftigt underskattad del av den organisatoriska förmågan att kunna hantera och tillgängliggöra data. Först efter en konfidentialitetsanalys kan man säkerställa att man faktiskt får göra data tillgängliga som öppna data, och om man inte kan garantera tillgänglighet och riktighet av sina data kommer värdet av data att förminska och data kommer dessutom att användas mycket mindre. Det

| Datum | Beteckning | Sida |
|------------|-------------------------------|-------|
| 2020-12-09 | Remissyttrande SOU 2020:55 | 3 (5) |

räcker alltså inte att kräva att data kan tillgängliggöras i maskinläsbara format om man inte kan lita på datas tillgänglighet och riktighet. Vi föreslår att detta förtydligas i lagtexten.

Kommentarer till enstaka förslag i utredningens huvudbetänkande

5.3 Risker med ökad tillgång till information

RISE tillstyrker de generella resonemang i stycket, men påpekar att bedömningen huruvida viss data i sammanslagning med annan data sammantaget kan utgöra skäl för säkerhetsskyddsklassificering svårligen kan ses som en rutinuppgift på exempelvis kommunal eller regional nivå, varpå det kan bli svårt att göra sådan bedömning. Utifrån en försiktighetsprincip riskerar därför mer data än lagstiftaren avsett att inte tillgängliggöras.

8.3.1 Tillgängliggörande av information

RISE tillstyrker utredningens förslag.

Kommentar: Definitionen av *tillgängliggörande* stämmer överens med bland forskningsorganisationer etablerat vokabulär för öppna forskningsdata.

8.3.2 Tillhandahållande av information

RISE tillstyrker utredningens förslag.

Kommentar: RISE avser att i forskningsdataprocessen tillhandahålla data, som omfattas av krav på publicering, enligt FAIR.

8.3.3 Lagen ger inte rätt att få tillgång till information

Kommentar: RISE tillstyrker utredningens förslag.

Kommentar: Detta innebär att RISE forskningsdata *ska* tillgängliggöras endast på begäran av forskningsfinansiärer, andra parter såsom tidskriftsförlag samt ev. grundat på framtida intern policy. Inget juridiskt "tvång" är lagstadgat vilket ger RISE rörelsefrihet i samverkan med t.ex näringsliv.

8.4 Begäran om vidareutnyttjande

RISE tillstyrker utredningens förslag.

Kommentar: RISE omfattas inte av förslaget text om vidareutnyttjande, enligt definition på s. 371 i utredningen.

9.1.9 Öppna data

RISE tillstyrker utredningens förslag.

Kommentar: Kommentarer ang. CC-BY-licens, som omnämns på s. 156: Det finns en viss skillnad mellan offentliga data och forskningsdata. Offentliga data är troligen mer självklara i sitt ursprung, medan forskningsdata kan vara svårare att spåra. CC-BY licensen är mycket viktig för den vetenskapliga transparensen då den garanterar uppbyggnad av en spårbarhet i var forskningsdata kommer ifrån och hur den bearbetats. CC-BY licenser är etablerad praxis vid publicering av forskningsdata.

9.1.10 Forskningsdata

RISE tillstyrker utredningens förslag.

Kommentar: Definitionen stämmer överens med rådande praxis. Se också kommentar under 10.5.1 Datorprogram.

10.4.1 Information som innehas av myndigheter och offentliga företag

| Datum | Beteckning | Sida |
|------------|-------------------------------|-------|
| 2020-12-09 | Remissyttrande SOU 2020:55 | 4 (5) |

RISE tillstyrker utredningens förslag.

Kommentar: RISE omfattas inte inom denna definition och lagen skulle bara beröra den forskningsdata RISE tillgängliggör pga exempelvis krav från finansier.

10.4.2 Forskningsdata

RISE tillstyrker utredningens förslag.

Kommentar: I och med att lagen inte innehåller några incitament att tillgängliggöra forskningsdata ser vi ingen påverkan på RISE. Eventuella konflikter kan uppstå i samverkansavtal om en forskningsfinansier har krav på tillgängliggörande av data medan en projektpartner till RISE vill hålla data konfidentiell. Men detta är inget som påverkas av lagförslaget.

10.5.1 Datorprogram

Kommentar: Termen *datorprogram* behöver definieras. För forskningsdata brukar programkod/script som använts för analys räknas in som forskningsdata då de krävs för att validera resultat. Det kan då bli missvisande att påstå att lagen inte ska omfatta datorprogram, om termen också omfattar denna typ av programkod.

11.4.3 Befintliga format och språkversioner

RISE tillstyrker utredningens förslag

Kommentar: Förslaget förordar användande av öppna standarder för att öka möjligheterna till vidareförädling och bearbetning. Standarder inom öppna data kan tillämpas på tre nivåer: format, metadata och semantisk nivå, och utredningen adresserar två av dessa, format och metadata. Den *semantiska* nivån (som dikterar hur en entitet ska beskrivas) är i många sammanhang en nyckel till skalbarhet och innovation vad gäller återanvändning av öppna data, och har varit en framgångsfaktor inom ex. kollektivtrafik och geodata. Detta är därför en publiceringsstrategi som RISE förespråkar i ex. projekt NSÖD¹, ODIN² och CaaP³. Eftersom data vid denna typ av publicering uttrycks i en semantiskt förutbestämd form (snarare än i sin ursprungliga form), kan en sådan publiceringsstrategi komma i konflikt med ”4 kap Krav på format 1 §: Information ska tillhandahållas i befintliga format och språkversioner.” Vi ser därför att nuvarande skrivning kan ha en hämmande effekt på semantiskt standardiserad publicering av data då sådan publicering inte nödvändigtvis uppfyller lagkraven. Genom tillägg i utredningen som medger att information som publiceras i enlighet med semantiska krav också uppfyller lagkraven anser vi att själva syftet med lagen (dvs innovation genom information) uppfylls till en högre grad än vad som nu riskerar bli fallet.

11.4.4 Maskinläsbarhet och andra särskilda krav på format

RISE tillstyrker utredningen förslag.

Kommentar: RISE arbetar för att kunna hantera och publicera den forskningsdata som skulle omfattas av lagen enligt FAIR-principerna vilket uppfyller lagens krav. Då FAIR-principerna (FAIR innebär bland annat att forskningsdata ska vara Findable (sökbara), Accessible (tillgängliga), Interoperable (interoperabla) och Reusable (återanvändbara)) är väl kända bland organisationer som innehar forskningsdata undrar vi varför utredningen valt att inte hänvisa till dessa principer här?

I näst sista stycket står det följande: ”Utgångspunkten är, i likhet med vad som anförts i avsnitt 11.3.3, att kärnverksamheten inte ska påverkas. Det kan inte heller krävas att

¹ <https://www.ri.se/sv/vad-vi-gor/projekt/nationell-skalning-oppna-data>

² <https://nordicopenmobilitydata.eu/>

³ <https://cityasaplatform.se/>

kostnadskrävande teknisk utrustning anskaffas eller att personal anställs för att kunna efterleva de nu aktuella formatkraven.” Ett offentligt företags kärnverksamhet är kopplad till företagets uppdrag. Om ett offentligt företag – eller vilken som helst annan organisation – börjar jobba mer datadrivet kommer kärnverksamheten att påverkas. Även om det är samma uppdrag som ska lösas kommer de olika aktiviteter och processer som former kärnverksamheten att förändras. Beroende på exakt man definierar ”kärnverksamhet” kommer den alltså på något sätt att påverkas av att man ska tillhandahålla information på maskinläsbart format. Om en organisation verkligen ska ta till sig alla digitaliseringens möjligheter och jobba datadrivet helt i grunden (vilket typiskt inte är fallet i offentlig verksamhet idag) kommer detta att kräva och leda till att man anskaffar nya typer av teknisk utrustning och verktyg, och man måste anställa personal med nya av kompetenser samtidigt med att existerande kompetenser blir irrelevanta p.g.a. effektiviseringar. Det är besvärligt att hantera när det händer, men det är ingenting konstigt i detta; tekniksprång har flera gånger tidigare i historien lett till omvälvande förändringar. Vi anser att det är defensivt på gränsen till missvisande att påpeka att kärnverksamheten och organisationen inte kommer att behöva förändras om man börjar tillgängliggöra data. Tvärtom kommer det att krävas stora kompetens- och resursförändringar framöver om offentlig sektor ska jobba datadrivet på riktigt.

11.4.5 Dynamiska data

RISE tillstyrker utredningens förslag

Kommentar: Vi stödjer utredningens förslag men önskar, i tillägg, när det är möjligt och lämpligt, att dynamiska data ska tillhandahållas i format som är öppna, maskinläsbara, tillgängliga och sökbara. Detta ökar signifikant möjligheterna till återanvändning och värdeskapande.

11.4.6 Särskilda värdefulla datamängder

RISE tillstyrker utredningens förslag

Kommentar: Vi stödjer utredningens förslag men önskar, i tillägg, när det är möjligt och lämpligt, att särskilda värdefulla datamängder tillhandahålls i format som är öppna, tillgängliga och sökbara. Detta ökar signifikant möjligheterna till återanvändning och värdeskapande

14.6.3 Forskningsdata som omfattas av den nya lagen ska tillhandahållas avgiftsfritt

RISE tillstyrker utredningens förslag.

Kommentar: För RISE innebär utredningens förslag inte några krav på förändrat arbetssätt. Vi tillhandahåller forskningsdata som kommer omfattas av den nya lagen avgiftsfritt idag.

18.2.1 Utvärdering av lagens finansiella konsekvenser

Kommentar: Det är vår erfarenhet att den digitala mognaden hos offentliga aktörer skiljer sig markant, och vi bedömer att vissa aktörer, som mindre kommuner, kommer att drabbas av markant ökade kostnader (alternativt undvika frågan).

Att säkerställa att tillhandahållandet av information inte äventyrar integritet och informationssäkerhet (17.5), att en utvärdering av risker sker korrekt och samtidigt motverka att information inte lämnas ut på grund av osäkerhet runt riskerna kräver kompetens och vägledning. Särskilt problematiska att bedöma kan vara dynamiska data samt data som lätt kan samköras med annat data (informationssäkerhetsklassning av enskilda ingående dataset ger inte nödvändigtvis en fingervisning om vilken informationssäkerhetsklassning det resulterande datasetet har).