

**EXAMENSARBETE** Using data contracts to fulfill requirements on data in an intra-organizational data ecosystem

**STUDENTER** Erik Svensson, Max Torsby

**HANDLEDARE** Per Runeson (LTH), Anton Frost (Axis Communications)

**EXAMINATOR** Elizabeth Bjarnason (LTH)

# Datakontrakt, nya trenden inom datahantering?

POPULÄRVETENSKAPLIG SAMMANFATTNING **Erik Svensson, Max Torsby**

När ett företag växer så tenderar det att bildas en klyfta mellan dataproducenter, som ansvarar för insamling av olika typer av data, och datakonsumenter, som har rollen att extrahera värde från data för företaget. Detta arbete har utforskat hur datakontrakt kan användas för att minska denna klyfta.

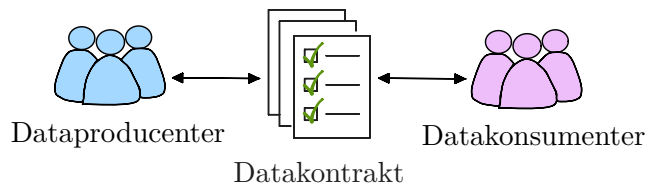
När datakonsumenter använder dataset, det kan till exempel handla om data från försäljning som används vid affärsanalys, uppkommer naturligt många frågor. Att hitta svar på dessa är både svårt och tidskrävande. Ofta måste man vända sig till dataproducenten för att få svar på frågorna. För att lösa detta har datakontrakt framkommit som ett nytt verktyg.

Ett datakontrakt är ett textdokument som tillhör ett dataset. Det är en samling metadata, alltså data som beskriver data, för ett dataset. Det kan till exempel vara kontaktuppgifter till dataproducenten, hur man får använda datan, hur tillförlitlig datan är och en beskrivning av strukturen på datasetet. Genom att samla metadata i ett datakontrakt så blir informationen lättillgänglig för datakonsumenterna. På grund av att datakontrakt är ett nytt verktyg är det brist på forskning om dem och deras användningsområden.

Organisationer har ofta stora mängder dataset vilket leder till utmaningar. För att kunna få en överblick över dessa kan en datakatalog användas. Det är ett verktyg som likt en produktkatalog samlar och visar upp dataset.

Vårt arbete har utförts på ett företag som hanterar många dataset och där klyftan mellan datakonsumenter och dataproducenter har

märkts av. Företaget har tidigare försökt använda datakataloger för att samla sina dataset. De stötte då på problem på grund av brist på strukturerad metadata över sina dataset.



Genom att implementera datakontrakt på företaget har vi dels undersökt om datakontrakt kan fungera som en standard för strukturering av metadata på företaget samt vilken metadata som är relevant att ha med i ett datakontrakt. Vår implementation bygger på en tidigare kravspecifikation för företagets datahantering.

Vi konstaterar att datakontrakt är värdefulla och uppfyller många av de krav företaget har på sin datahantering. Vi bedömer också att datakontrakt kan användas tillsammans med datakataloger och att de tillsammans har större slagkraft än var för sig. Avslutningsvis konstaterar vi att datakontrakt och datakataloger båda kan bidra till att minska klyftan mellan dataproducenter och datakonsumenter.