

Digitiserat svenskt tryck

Bedömning av nuläget inom satsningen Det Digitaliserade Svenska Trycket (DST)

Anders Söderbäck, Kungliga biblioteket

Digitiserat svenskt tryck



Kungliga biblioteket

www.kb.se

Diarienummer: KB 2024-538

Digitiserat svenskt tryck

Bedömning av nuläget inom satsningen Det Digitaliserade Svenska Trycket (DST)

Anders Söderbäck, Kungliga biblioteket

Innehållsförteckning

Sammanfattning	5
Uppdraget	5
Avgränsningar	5
DST-biblioteken	6
Göteborgs universitetsbibliotek	7
Lunds universitetsbibliotek	7
Stockholms universitetsbibliotek	7
Umeå universitetsbibliotek	8
Uppsala UB	8
Kungliga biblioteket	9
Sammanställning	9
Analys	10
Samlingarna vid DST-biblioteken och Libris deldatabas DST	10
Övriga frågeställningar	10
Slutsats	11

Sammanfattning

Denna rapport beskriver nuläget för de digitiserade samlingarna vid de sex bibliotek som ingår i DST-samarbetet; Kungliga biblioteket (KB) samt universitetsbiblioteken i Göteborg (GUB), Lund (LUB), Stockholm (SUB), Umeå (UmUB) och Uppsala (UUB). Avsaknaden av en exakt definition av ”det svenska trycket” liksom skiftande praxis vid de deltagande biblioteken omöjliggör en exakt beskrivning, men bör ge en tillräckligt bra överblick för att fatta beslut om det fortsatta arbetet. **Rapporten bedömer att KB, framför allt tack vare den pågående digitiseringen av svenska dagstidningar (ca 400 TB), står för den absolut största volymen av digitiserat material.** Materialet från övriga DST-bibliotek uppskattas utgöra ca 60 TB.

Digitaliserat svenskt tryck finns även vid andra organisationer än de som ingår i DST. Dessa organisationer har inte gått igenom i denna undersökning.

Uppdrag och metod

Uppdraget för denna undersökning var **att ge en bild av de digitala samlingarna vid DST-biblioteken, samt att ge underlag för KB:s planering** av det fortsatta arbetet inom DST-satsningen. I uppdraget ingick också att göra en analys av innehållet i **Libris deldatabas DST.**

Uppdraget har genomförts genom en insamling av metadata från **plattformarna Libris, Alvin och SwePub.** Libris har utgjort huvudkällan för arbetet. Inom ramen för uppdraget har också samtal genomförts med medarbetare eller chefer vid DST-biblioteken. Analysen av samlingarna och deldatabasen är mycket översiktlig.

Avgränsningar

Definition av svenskt tryck

En första frågeställning för undersökningen var att definiera vad som egentligen utgör Det svenska trycket. **Avsiktsförklaringen från 2020 sätter som mål att digitalisera och tillgängliggöra hela den svenska tryckproduktionen från 1400-talet och framåt.** Detta innefattar böcker, tidskrifter, rapporter, läromedel m.m. – ”det som traditionellt katalogiseras inom biblioteken” – medan vardagstryck visserligen ingår i den svenska tryckproduktionen men inte kommer att prioriteras.

Denna utredning utgår pragmatiskt, med ett tillägg till formuleringen i avsiktsförklaringen, från ”det som traditionellt katalogiseras *som svenskt* och *som tryckt* inom biblioteken.” Definitionen gör att materialkategorier som ligger på gränsen mellan vad som traditionellt betraktas som tryckt material och inte, t.ex. kartblad och noter, ibland kommer att ingå och ibland inte. Vad som ingår beror på bedömningar som gjorts från fall till fall vid biblioteken.

Att skilja ut vad som är ”svenskt” är heller inte självklart. **En rimlig definition är att materialet är utgivet av ett förlag baserat i Sverige. Denna definition är inte självklart att applicera på t.ex. äldre tryck.** Ej heller kan en avgränsning göras till vad som är utgivet på

svenska. Framför allt äldre tryck kan vara skrivet på latin, tyska eller franska och ändå tillhöra det svenska trycket.

Katalogisering av det svenska trycket

En överblick av det digitaliserade svenska trycket är beroende av hur detta tryck katalogiserats. Där finns ingen gemensam praxis vid de deltagande biblioteken – särskilt har någon sådan gemensam praxis inte funnits historiskt. Ibland har det digitaliserade materialet katalogiserats som en särskilt katalogpost, ibland har en länk till ett digitalt objekt tillförts den katalogpost som beskriver den fysiska förlagan. I plattformen Alvin, som används av universitetsbiblioteken i Göteborg, Lund och Uppsala, är praxis att katalogen beskriver det fysiska objektet, och att adderandet av en digital version helt enkelt talar om att det existerar en digitisering av just detta fysiska exemplar.¹

Länkar till digitalt material har historiskt lagts på existerande katalogposter för äldre litteratur, vilket gör att det inte finns någon självklar matchning mellan t.ex. den organisation, eller den sigel, som upprättat en katalogpost i Libris och den organisation som skapat ett digitalt objekt.

I ett antal fall har material märkts upp som manuskript eller handskrift, trots att det vid en granskning är tydligt att det rör sig om ett tryckt objekt. Sådana misstag uppstår vid manuell katalogisering. Inget av detta gör den bedömning som gjorts i denna utredning meningslös, men man bör vara medveten om att bedömningen inte är eller kan vara exakt.

Digitisering åt andra än det egna biblioteket

Flera av biblioteken inom DST-samarbetet har digitiserat material åt andra organisationer. Exempel på detta är SUB:s digitisering av Filosofisk tidskrift, som gjorts på uppdrag av tidskriften, och digitiseringen av Taxeringskalendern åt ett forskningsprojekt inom nationalekonomi. Detta material ägs och tillgängliggörs inte av SUB. Sådant material har jag därför valt att inte ta hänsyn till i denna undersökning.

Digitalisering vid andra bibliotek än DST-biblioteken

Fler bibliotek än de sex som ingår i DST-samarbetet har arbetat med olika former av digitalisering. Denna undersökning har inriktat sig på de sex bibliotek som ingår i DST-samarbetet: KB och Universitetsbiblioteken i Göteborg, Lund, Stockholm, Umeå och Uppsala.

DST-biblioteken

Biblioteken inom DST lagrar och publicerar idag material på flera olika plattformar. Anledningarna till detta är framför allt historiska. Biblioteken har byggt upp sina digitala samlingar utifrån lokala förhållanden, i något fall har en särskild lösning behövt införas för att hantera en avtalslicens som bara gäller inom Sverige. Infrastrukturen har som man brukar

¹ Alvin formathandbok s. 4, <https://www.alvin-portal.org/info/digitalAssets/Alvin-formathandbok.pdf>

säga ”växt fram organiskt” utifrån de system som funnits tillgängliga, eller som de deltagande biblioteken av olika anledningar redan använt. När ett nytt system införts är det inte självklart att digitaliserat material flyttats tid från en tidigare lösning. Detta medför att det inte är helt enkelt att skaffa en överblick på nationell nivå. Libris spelar en nyckelroll i att knyta ihop dessa olika system, och större delen av det digitaliserade svenska trycket går att identifiera genom Libris. Libris ger dock inte någon tydlig bild av mängden digitaliserat material – stora mängder av det digitaliserade materialet utgörs av tidningar eller tidskrifter vilka beskrivs med en övergripande katalogpost i Libris men utgör ett betydligt mer omfattande material om man till exempel tar sig vidare till KB:s tjänst Svenska dagstidningar.

Göteborgs universitetsbibliotek

Göteborgs universitetsbibliotek driver tillsammans med Lunds universitetsbibliotek och Uppsala universitetsbibliotek plattformen Alvin. Plattformen finansieras av ett konsortium av 19 organisationer, där GUB, LUB och UUB utgör styrgrupp. Vid sidan av det material som finns i Alvin har GUB publicerat digitaliserat material på den egna publiceringsplattformen Gupea.² I Gupea finns 20228 digitaliserade objekt, allt från kyrkoböcker via bibliografier, läroplaner och verk av August Strindberg till ett 80-tal tidskrifter. 6336 av dessa objekt finns katalogiserade i Libris. Åtminstone majoriteten av materialet i Alvin är katalogiserat i Libris och ingår i deldatabasen DST.

Lunds universitetsbibliotek

Lunds universitetsbibliotek publicerar liksom GUB digitaliserat material i plattformen Alvin. Det material som finns i Alvin återfinns liksom vad gäller Göteborgs UB i Libris.³ 21 poster finns för svenskspråkigt material utan att vara uppmärkta för deldatabasen DST.

Tingstensamlingen, som innehåller digitaliserade versioner av Herbert Tingstens samlade verk, återfinns dock inte i Alvin eller Libris.⁴

Stockholms universitetsbibliotek

Stockholms universitetsbibliotek har framför allt digitaliserat doktorsavhandlingar. Avhandlingar från 1906 till 2003 är klara (med undantag av två avhandlingar som där SUB saknar fysiska förlagor). Doktorsavhandlingarna katalogiseras i Diva. På grund av att den avtalslicens som krävs för publicering inte möjliggör tillgängliggörande utanför Sveriges gränser publiceras avhandlingarna på en särskild server hos SUB (pubs.sub.su.se). Det gör att de inte är helt lätta att söka ut, sökningen funkar inte i Diva. Det går dock att söka fram dem i SwePub, och en sådan sökning ger 4916 träffar. Till detta kompletterar SUB muntligt att ca 200 avhandlingar blivit ”äkta” fulltexter i Diva, d.v.s. publicerade direkt i Diva och därför

² <https://gupea.ub.gu.se/handle/2077/30489>

³ (db:dst OR (db:digi AND lang:swe)) ocode:lub AND "urn:nbn:se:alvin"

⁴ <https://www.ub.lu.se/hitta/digitala-samlingar>

inte möjliga att skilja från digitalt födda avhandlingar på annat sätt än att manuellt granska filerna.

Allt annat material som SUB digitaliserat katalogiseras i Libris och publiceras på en egen server (images.sub.su.se), 194 träffar varav 171 ingår i del databasen DST och ytterligare 9 titlar som inte ingår i det svenska trycket ingår i del databasen DIGI. 14 poster är inte uppmärskta som digitaliserat material i Libris, men fortfarande möjliga att söka fram som fritt tillgängligt digitalt material. Dessa 14 poster är samtliga svenskt tryck, två av dem tidskrifter.

Umeå universitetsbibliotek

Umeå universitetsbibliotek har enligt egen utsago allt sitt digitaliserade tryckta material katalogiserat i Libris. Materialet är idag publicerat i systemet Nainuwa, levererat av det Österrikiska företaget Treventus. Doktorsavhandlingar från ett äldre katalogiseringsprojekt är publicerade i Diva. Umeå betraktar Nainuwa som ett visningssystem, och där finns bara övergripande metadata. Mer data kring de digitaliserade objekten finns i Libris. Eftersom digitaliserade objekt märkts inom olika projekt märkts upp olika i Libris, är det inte helt enkelt att söka fram dem – alla ingår inte i del databasen DST eller har bibliografikoden DIGI.

En sökning i Libris på 'digitalisering@UmU OR ("urn:nbn:se:umu:rara" AND (db:digi OR db:dst)) OR ("urn:nbn:se:umu:diva" AND (db:digi OR db:dst))' ger 3594 träffar vilket förefaller inkludera samtliga digitaliserade tryckta publikationer från Umeå UB.

Uppsala universitetsbibliotek

Uppsala universitetsbibliotek publicerar liksom GUB och LUB material i plattformen Alvin. Sökgränssnittet i Alvin gör det svårt att skilja på digitaliserat tryck och på handskrifter i bokformat. Metadata har därför hämtats från Alvin via protokollet OAI-PMH. Utifrån detta material bedömer jag att det material som är intressant för den här undersökningen även återfinns i Libris. Det är däremot inte alltid uppmärskt med DST eller DIGI. Ofta uppmärskt med bibliografikoden SPDT – ”Specialsamlingar (digitaliserat) tryck, Uppsala universitetsbibliotek, Enheten för äldre tryck och samlingar”. Vissa skrifter (t.ex. Noveller och skizzer av Tor Hedberg, utgiven 1889) är inte uppmärskta på något annat sätt än att de innehåller en urn:nbn-länk till Alvin.

Likaså finns ett stort antal katalogposter uppmärskta som tillhörande del databasen DST med länkar som leder till Diva, men utan att något bestånd ligger på UUB. Dessa poster förefaller ha levererats som del av e-plikten. Större delen av det digitaliserade tryck som identifierats från Uppsala, 14176 poster varav 12198 på latin, ingår i denna mängd. Någon motsvarande mängd utan bestånd i Libris förefaller inte finnas för något av de andra DST-biblioteken.

Kungliga biblioteket

KB är det bibliotek som digitaliserat mest svenskt tryck av biblioteken inom DST.⁵ Vid sidan av digitaliseringar som registrerats i Libris med bestånd på KB, finns ett stort antal svenska dagstidningar som digitaliserats av KB men som inte registrerats med en synlig beståndspost.⁶ Dagstidningar ingår i tjänsten Svenska tidningar (<https://tidningar.kb.se>). I nedanstående sammanställning har dagstidningarna lagts till bedömningen av KB:s digitaliserade material.

Sammanställning – bedömning av nuläget

Nedanstående tabell bygger på data från Libris, kompletterat med data för SUB:s digitaliserade doktorsavhandlingar, som inte återfinns i Libris. Att göra en undersökning baserat på Libris är inte enkelt, eftersom beståndsuppgifter kan ligga på flera bibliotek och det är relativt vanligt att flera av DST-biblioteken har beståndsuppgifter på samma post – sannolikt för att få med dessa poster i exporten till sina lokala bibliotekskataloger.

Data från KB, SUB och UmUB går att separera på egna uppmärkningar och sökkoder, liksom det material som GUB publicerat utanför Alvin. Digitaliserat material publicerat i Alvin är dock inte helt enkelt att separera på de olika upphovsorganisationerna genom en sökning i Libris. Jag har försökt göra detta i nedanstående tabell, men det är möjligt att siffrorna för dessa tre bibliotek inte är helt rättvisande. Det är också möjligt att material digitaliserat av någon annan medlem av Alvinkonsortiet kommit med i inventeringen. Likaså har sannolikt ett mindre antal poster från t.ex Litteraturbanken tagit sig in i sammanställningen, då det rör svenskt material och något av DST-biblioteken lagt bestånd på posten.⁷

Längst ned i tabellen finns en rad med enbart det som finns i Libris tillhörande deldatabasen DST.

	Poster i Libris		På svenska, ej db:DST		1901-1999		2000-	Böcker	Tidskrifter	Övrigt
GUB	8713	8036	675	3224	5167	322	8274	140	299	
LUB	1100	1079	21	1072	35	3	1100	0	0	
SUB	185	171	14	156	4064	878	5276	11	0	
UmUB	3594	2971	273	1249	1946	399	3458	19	223	
UUB	16733	16592	182	15990	1126	230	2038	15	504	
KB	35860	29790	6033	16136	19422	302	33945	1915	0	
<i>Summa</i>	<i>66185</i>	<i>58639</i>	<i>7198</i>	<i>37827</i>	<i>31760</i>	<i>2134</i>	<i>54091</i>	<i>2100</i>	<i>1026</i>	
DST	72941	72941	0	41221	28296	3424	60175	2285	15051	
<i>Diskrepans</i>	<i>6756</i>	<i>14302</i>	<i>-7198</i>	<i>3394</i>	<i>-3464</i>	<i>1290</i>	<i>6084</i>	<i>185</i>	<i>14025</i>	

⁵ Undersökningen räknar inte digitalisering av handskrifter eller annat som tydligt faller utanför ramen av svenskt tryck.

⁶ Sökning i Libris webbsök för poster med bestånd hos KB: 'ocode:kb (db:dst OR (db:digi AND lang:swe)) AND "urn:nbn:se:kb" NOT (mat:bild OR mat:karta OR mat:handskrift OR mat:affisch OR mat:artikel OR mat:noter OR mat:annat)', sökning på dagstidningar som ingår i det svenska trycket: 'genre:dagstidningar db:dst'

⁷ Det är inte nödvändigtvis fel att få med andra delar av det svenska trycket än det som digitaliserats vid DST-biblioteken. Denna undersökning utgår dock från en avgränsning till DST-biblioteken.

Utifrån denna tabell kan vi göra en högst rudimentär uppskattning av den mängd digitala filer som finns vid DST-biblioteken, vid sidan av KB: **Vi utgår från att en bok är i genomsnitt 200 sidor, och att en scannad sida i genomsnitt är på 10 MB. Varje bok utgör då 2 GB.** Den samlade digitiseringen av böcker vid de fem DST-biblioteken utanför KB utgör då sammanlagt 40 TB data. Till detta kan läggas 185 tidskrifter. KB:s digitiserade dagstidningar utgör idag uppskattningsvis ca 400 TB. en mängd titlar som digitiserats vid de övriga DST-biblioteken (framför allt GUB) är ca 10% av KB:s. Då det rör sig om tidskrifter snarare än dagstidningar är däremot den faktiska mängden filer med största sannolikhet lägre. Om vi utgår från att mängden digitaliserade tidskrifter utgör 20 TB, blir den totala mängden 60 TB data. KB:s lagringskostnader är idag ca 70 kr per TB och månad, vilket innebär att kostnaden för att lagra det data som idag finns vid de andra DST-biblioteken i ett år skulle bli ca 50 000 kr. Notera att varken lagringskostnad eller datamängd inte är statiska siffror. Lagringskostnaden kan komma att ändras beroende på kostnadsökningar från leverantörer av licens-, service- och supportkostnader.

Analys

Samlingarna vid DST-biblioteken och Libris deldatabas DST

Att få en fullständig korrekt sammanställning av det digitaliserade materialet vid DST-biblioteken låter sig sannolikt inte göras utan att faktiskt samla in materialet på ett och samma ställe. **Dagens distribuerade modell är inte tillräckligt standardiserad för att möjliggöra en fullständig överblick. Ovanstående tabell ger sannolikt en bild som stämmer relativt väl överens med verkligheten.**

Noteras bör att det finns material vid de bibliotek som ingår i DST-samarbetet som inte är katalogiserat som del av deldatabasen DST, inte är katalogiserat med bestånd på det bibliotek som äger materialet, eller över huvud taget är katalogiserat i Libris.

Ca 7200 poster för svenskspråkigt digitaliserat material ingår inte i DST. 14302 poster ingår i DST men förefaller inte beskriva material som hör till något av de bibliotek som ingår i DST. En stor del av dessa poster beskriver bilder (7851 st), eller artiklar som levererats som del av e-plikten från något av de bibliotek som använder Diva. Huruvida digitaliserade bilder ingår i det svenska trycket kan diskuteras (jag har utgått från den definition som återfinns i avsiktsförklaringen, där jag bedömer att bilder utesluts). Att e-pliktslevererade artiklar ingår i svenskt tryck stämmer, men eftersom de förefaller vara digitalt födda snarare än digitaliserade från fysiska förlagor bör de inte ingå i deldatabasen.

Övriga frågeställningar

Vid sidan av det material som digitaliserats av de sex bibliotek som skrivit under avsiktsförklaringen om DST finns digitaliserat svenskt tryck i olika form och kvalitet från bl.a. **Litteraturbanken, från Projekt Runeberg, från Stiftsbiblioteken** m.fl. Huruvida detta material ska betraktas som en del av det digitaliserade svenska trycket förefaller för mig vara en fråga om syftet med DST. Digitalisering sker med olika standard utifrån olika syften.

En annan fråga är hur viktigt det är att få exakta siffror på det digitaliserade svenska trycket. Hur viktigt detta är, och hur materialet ska användas, bör vara avgörande för det fortsatta arbetet. Att hitta en exakt definition av vad som utgör svenskt tryck och vad som inte tillhör denna kategori är sannolikt omöjligt.

Slutsats

KB står för den större mängden av den digitalisering som hittills genomförts av svenskt tryck. En stor del av den digitalisering KB gjort utgörs dessutom av dagstidningar, som volymmässigt utgör klart större mängder digitala filer. **Utmaningen i att samla material centralt hos KB bör därför inte ligga i kostnader för lagring, utan i den komplexitet som finns i att samla material som över tid skapats och beskrivits utifrån skilda rutiner vid olika institutioner.**

Om så önskas kan en mer djupgående beskrivning av innehållet göras. En mer djupgående beskrivning bör dock utgå från tydligt formulerade frågeställningar.

Kungliga biblioteket

Besöksadress: Humlegården
Box 5039, 102 41 Stockholm
Telefon: 010-709 30 00
E-post: info@kb.se
www.kb.se

