

2001-09-14

01-393-2001

Datainspektionen
Box 8114
104 20 Stockholm

Ang. Kungl. bibliotekets arbete med att arkivera det elektroniska kulturarvet
(Dnr 1404-2001)

Kungl. biblioteket (KB) har den viktiga uppgiften att för framtiden spara den information som på olika sätt uttrycks av vår egen samtid. Den lag som KB:s arbete vilar på, är en av de äldsta i världen och härstammar från 1661. Från denna tidpunkt finns allt som publicerats i landet av tankar, idéer och känslor sparat i KB:s arkiv. Få samhällsforskare och humanister har genom århundradena kunnat undvara KB:s nästan oändliga resurser. KB har varit och är en hörnsten i samhällsbygget. Årligen inkommer ca 900 hyllmeter svenskt tryck. Det mesta handlar om olika typer av tillfallestryck. Därutöver erhålles tidningar, böcker och tidskrifter.

Nu händer det emellertid någonting radikalt med tillfällighetstrycket. Myndigheter, företag, föreningar, tidningar och intressentgrupper m.m. väljer i allt högre grad att enbart tillhandahålla sin information via Internet. Materialet kommer på så vis inte att bevaras för framtiden utan hamnar i evig glömska. Detta är synnerligen allvarligt. KB:s funktion urholkas och uppdraget från regering och riksdag och ytterst svenska folket kan inte fullgöras.

Regeringen har också i prop. 2000/2001:3 bl.a. uttalat att det är viktigt att bevara och tillhandahålla yttringar av svenskt liv, svenskt samhälle och svensk kultur, däribland information som görs tillgänglig on-line, samt andra elektroniska dokument. Det slås fast att vårt kulturarv måste säkras, utan kvalitetskriterier och i rimlig omfattning.

För att hantera denna svåra situation initierade KB projektet Kulturarw³, där KB två gånger om året samlar in svenskt material från WWW med avancerad robotteknik och via ett automatiskt program. Urvalet sker på basis av domännamn. Om det bedöms att domänen är att betrakta som svensk utifrån i förväg fastställda kriterier så hämtas informationen oavsett form, innehåll etc. FTP-arkiv med programvara ingår icke. Databasen innehåller drygt 65 miljoner sparade webbsidor. Eftersom en sida vanligtvis är uppbyggd av flera

objekt, är antalet insamlade filer ungefär dubbelt så stort, dvs. ungefär 130 miljoner. Dessa dokument består av olika text- bild- och ljudformat. Totalt har KB hittills funnit 439 olika namn på filtyper bland materialet. I praktiken är en del av dessa dokument av samma typ.

Vid ett närmare studium av materialet vintern 2001 omfattande 30,8 miljoner filer framgick att de mest förekommande formaten är html-filer (ca 16,2 miljoner sidor), gif-filer (ca. 6,2 miljoner bilder) jpeg-filer (6,2 miljoner bilder) och enkel text (ca 800 000 sidor). Dessa fyra vanligaste format utgjorde 95% av allt som var insamlat. Resterande 5% var således fördelade på 435 olika format. Delades formaten sedan upp i tre huvudtyper, bestod ca 17 miljoner sidor av text, ca 12,5 miljoner av bild och ca 1,4 miljoner av övrigt. I den senare gruppen ingick ca. 320 000 sidor PDF-filer, 200 000 audio-filer ca 20 000 videofiler och samtliga övriga filer utgjorde totalt ca 10 000 sidor.

Vid insamlandet av material sker en ofrånkomlig behandling av personuppgifter. Själva behandlingen är en förutsättning men inte något KB eftersträvat. Kulturarw³ är, som framgått, synnerligen betydelsefullt för forskning i vidaste bemärkelse. Det får därför anses ligga i det allmännas intresse att KB samlar in svenskt material från webben. Behandlingen sker enligt 10 § d) personuppgiftslagen, PuL. De grundläggande krav som ställs i PuL 9 § får anses uppfyllda. KB samlar in personuppgifter endast för ovan uttryckligt angivna och berättigade ändamål. De insamlade personuppgifterna behandlas inte för något ändamål som är oförenligt med det för vilket uppgifterna samlades in. Personuppgifterna får inte bevaras under en längre tid än vad som är nödvändigt med hänsyn till ändamålen med behandlingen. Personuppgifterna är en integrerad del av det insamlade materialet som KB har för avsikt att spara för all framtid. KB anser att det för såväl historiska som vetenskapliga ändamål är möjligt i enlighet med 9 § 2 st.

Vidare anser KB att de eventuella känsliga personuppgifter som förekommer i det insamlade materialet kan behandlas i enlighet med PuL 19 §. Kulturarw³ är ett forskningsprojekt och projektet får anses så viktigt att samtycke av enskilda personer inte bör vara nödvändigt. KB:s bedömning är att samhällsintresset av Kulturarw³ klart överväger den risk för otillbörligt intrång i enskildas personliga integritet som behandlingen kan innebära.

KB har utsett personuppgiftsombud och gjort anmälan om det till Datainspektionen. KB följer i övrigt lagens krav vad gäller att lämna besked om personuppgifter i KB:s andra register till den registrerade, om han eller hon ansöker om det. Vad gäller den insamlade svenska webben anser sig KB inte ha skyldighet att lämna sådana uppgifter. Enligt KB:s mening ger artikel 13.2. EG-direktivet stöd för denna tolkning. De allra flesta uppgifterna förekommer i löpande text. KB har gjort den bedömningen att det i praktiken är omöjligt att lämna den registrerade information om behandlingen av uppgifterna. Några åtgärder som rör de registrerade, som samlats in via den svenska webben, vidtas överhuvudtaget inte.

KB har inte mottagit någon begäran om rättelse av personuppgift i enlighet med PuL 28 §. Det torde heller inte vara möjligt att genomföra korrigeringar av personuppgifter eftersom syftet med KB:s insamlande är att materialet skall bevaras i sitt ursprungliga skick.

Tomas Lidman