

MAGISTERUPPSATS I BIBLIOTEKS- OCH INFORMATIONSVETENSKAP
VID INSTITUTIONEN BIBLIOTEKS- OCH INFORMATIONSVETENSKAP/BIBLIOTEKSHÖGSKOLAN
2009:11
ISSN 1654-0247

Massdigitalisering och kvalitativ digitalisering

En jämförelse av digitaliseringen
på nationalbiblioteken i Norge och Sverige

CHRISTOFFER NILSSON



HÖGSKOLAN I BORÅS
VETENSKAP FÖR PROFESSION

© **Författaren**

Mångfaldigande och spridande av innehållet i denna uppsats
– helt eller delvis – är förbjudet utan medgivande.

Svensk titel:	Massdigitalisering och kvalitativ digitalisering: En jämförelse av digitaliseringen på nationalbiblioteken i Norge och Sverige
Engelsk titel:	Mass digitization and qualitative digitization: a comparative study of digitization by national libraries in Norway and Sweden.
Författare:	Christoffer Nilsson
Kollegium:	2
Färdigställt:	2009
Handledare:	Jan Buse
Abstract:	<p>The purpose of this thesis is to compare mass digitization and qualitative digitization, to see how the digitization process and the digitalized material differ. The study emanates from the mass digitization as it is performed at the National Library in Norway, and the qualitative digitization at the Royal Library in Sweden. The method used is document studies combined with mail interviews. The focus is upon the practical operation, to which counts: “purpose & selection”, “preparation, image capturing and processing”, “metadata & text encoding”, “quality assessment”, “storage of digital originals and copies” and “display”.</p> <p>The result showed that the mass digitization automatize the process as far as possible and uses uniform methods of whole collections. They have minimal preparation, processing and quality controls, if it doesn't make the digitization quicker. During qualitative digitization the most parts are done manually, where every moment and object is prepared, processed and controlled considering what is best for the original. The result of the product of digitization – the digital documents – showed that qualitative digitization prefers aspects like caution and security, as they among other things uses the well established file format TIFF, and uses sufficiently high quality in methods to preserve all information from analogue documents, and store both processed and unprocessed versions of the digitization. Mass digitization is more insecure as they use the non-established file format JPEG2000, and the digital master is processed, compressed (lossless) and cut.</p>
Nyckelord:	massdigitalisering, kvalitativ digitalisering, digitalisering, Kungl. biblioteket, Nasjonalbiblioteket, bevarande, tillgängliggörande, digitalisering av kulturarvet

1. INLEDNING	5
1.1 ÄMNESVAL	6
1.2 PROBLEMBESKRIVNING	6
1.3 AVGRÄNSNING.....	7
1.4 SYFTE	7
1.5 FRÅGESTÄLLNINGAR	7
1.6 LITTERATURSÖKNING OCH KÄLLKRITIK.....	8
1.7 METOD	8
1.8 DISPOSITION	11
2. LITTERATUR OCH TIDIGARE FORSKNING	11
3. TEORETISK RAM FÖR DIGITALISERINGEN	13
3.1 TEXT OCH BILD	13
3.2 ANALOGT OCH DIGITALT	13
3.3 OM DIGITALISERING OCH DIGITALISERINGSPROCESSEN	14
3.3.1 Argument för att digitalisera.....	14
3.3.2 Argument mot att digitalisera	15
3.4 MASSDIGITALISERING OCH KVALITATIV DIGITALISERING.....	16
4. TEKNISK RAM FÖR DIGITALISERINGSPROCESSEN	17
4.1 SYFTE OCH URVAL	17
4.2 FÖREBEREDELSE, BILDFÅNGST OCH BEARBETNING.....	18
4.2.1 Bildens kvalité.....	18
4.2.2 Filformat	21
4.2.3 Bildfångstutrustning.....	22
4.2.4 Omvandling till digital text.....	22
4.3 METADATA OCH UPPMÄRKNING	24
4.4 KVALITETSBEDÖMNING.....	26
4.5 LAGRING AV DIGITALA ORIGINAL OCH KOPIOR	28
4.6 FRAMVISNING	30
4.6.1 Bildhanteringssystem	30
4.6.2 Hur och vad användaren får ta del av i framvisningen.....	31
4.7 SAMMANFATTNING AV DIGITALISERINGSPROCESSEN	33
5. RESULTAT	35
5.1 NASJONALBIBLIOTEKET I NORGE.....	35
5.1.1 Syfte och urval.....	36
5.1.2 Förberedelse, bildfångst och bearbetning.....	37
5.1.3 Metadata och uppmärkning	39
5.1.4 Kvalitetsbedömning.....	39
5.1.5 Lagring av digitala original och kopior.....	40
5.1.6 Framvisning	41
5.2 KUNGL. BIBLIOTEKET I SVERIGE	42
5.2.1 Syfte och urval.....	43
5.2.2 Förberedelse, bildfångst och bearbetning.....	44
5.2.3 Metadata och uppmärkning	46
5.2.4 Kvalitetsbedömning.....	47
5.2.5 Lagring av digitala original och kopior.....	48
5.2.6 Framvisning	49
6. ANALYS	51
6.1 BESVARANDE AV FRÅGESTÄLLNING 1	51
6.2 BESVARANDE AV FRÅGESTÄLLNING 2	55
7. DISKUSSION OCH SLUTSATSER	57
8. SAMMANFATTNING	62

KÄLLFÖRTECKNING	64
PUBLICERADE KÄLLOR	64
OPUBLICERADE KÄLLOR.....	69
<i>E-postfrågor</i>	69
BILAGA 1	70
BILAGA 2	71

1. Inledning

I nationalbibliotekens ägo finns mängder av dokument, i form av texter, bilder, affischer, m.m., som är av stort kulturhistoriskt och samhällsligt värde. Tyvärr är merparten av de mest intressanta samlingarna vare sig kända eller tillgängliga för den stora allmänheten. Tack vare digitaliseringen under 1900-talet och internet på 1990-talet har det blivit lättare för allmänheten att ta del av dokumenten. Under de senaste åren har projekt och arbeten vuxit fram runt om i världen och på biblioteken. Att digitalisera sitt material – enstaka eller större samlingar – ses som en praktisk lösning för att dokument ska nå en vidare krets av intresserade samtidigt som dokumenten bevaras som digitalt lagrad kopia och att originalet slipper påfrestningen som hanteringen innebär. Även om digitalisering som arbetsmoment har existerat i flera år innefattar den fortfarande frågor att beakta. Processen att omvandla ett analogt objekt till en digital version lagrad på bibliotekets datorer är lång och består av en rad problem att utvärdera innan digitaliseringsverksamheten startar. Den innefattar syftet med digitaliseringen till urval och hur det analoga verket görs digitalt, vilken metadata som ska medfölja, hur det lagras och sedan görs nåbara för biblioteksanvändarna. Med ”digitalisering” avses hela angivna processen, och inte bara bildfångsten.

Det finns flera uppfattningar om digitalisering och dess förhållande mellan det analoga och digitala. Två sådana förhållningssätt är massdigitalisering (eller kvantitativ digitalisering) och kvalitativ digitalisering (eller kritisk digitalisering). Massdigitalisering innebär att en stor mängd dokument digitaliseras och att institutionen då kan bli tvungen att göra avsteg på kriterier som kvalitet och kontroll, för att hinna digitalisera inom en tid som kan anses rimlig. Kvalitativ digitalisering innebär att det digitala dokumentet är i mycket hög kvalitet och har bearbetats för att passa många användargrupper, men att färre dokument hinner bli digitaliserade, medan resterande analogt material fortsätter åldras. Eftersom digitalisering består av många moment innebär det att ett digitaliseringsarbete kan använda flera olika tekniker, och att man därmed inte alltid kan sägas använda endast mass- eller kvalitativ digitaliseringsperspektiv. En och samma institution kan även praktisera flera digitaliseringsförhållanden beroende av bl.a. syfte och fysiskt tillstånd.

Hur man väljer att förhålla sig mellan det analoga och digitala objektet är av stor vikt och inverkar på hela arbetet, även den digitala kopian av originalet. Digitalisering och dess teknik är i ständig förändring, allt eftersom nya och bättre möjligheter kommer, snabbare bildfångst, lagringsdiskar som klarar långtidsbevaring, format med bättre komprimering, är bara några exempel.

I föreliggande uppsats jämförs Nasjonalbiblioteket (NB) i Norge och Kungl. biblioteket (KB) i Sverige. Båda biblioteken har aktiva digitaliseringsprocesser och jag har valt att utforska delar av deras hela digitaliseringsverksamhet. Det som jag undersöker hos NB är deras mål att digitalisera hela sin samling, där en stor del sker i form av massdigitalisering. Hos KB inriktar jag mig på deras digitalisering av enskilda dokument, när digitaliseringen utförs som kvalitativ digitalisering. Jag undersöker alltså inte institutionernas fullständiga digitaliseringsverksamhet, utan låter delar av deras arbeten fungera som modeller vid jämförelsen av massdigitalisering och kvalitativ digitalisering. NB utför även arbeten som kan ses som mer kvalitativ digitalisering med syfte att skapa högkvalitativa bilder, och på KB har de undersökt möjligheten till arbete som påminner om massdigitalisering.

1.1 Ämnesval

Digitalisering är idag vanligt på många nationalbibliotek och dess betydelse växer allt eftersom världen digitaliseras. Digitalisering som område och ämne utvecklas även den, där tekniker för bildfångst förbättras och lagringsmöjligheterna ökar i omfång och hastighet. Samtidigt vållar detta problem med hur biblioteken ska digitalisera sitt material till format som ständigt befinner sig i rörelse. Den förbättrade tekniken ger lösningar samtidigt som den skapar nya problem när samhället går mot det digitala och många användare förväntar sig digital åtkomst.

Genom att se på de digitaliseringstekniker som idag används på två nationalbibliotek är det därför möjligt att se hur väl dessa står sig sinsemellan i dagens digitala samhälle och vilka fördelar respektive nackdelar dess digitalisering innebär i både processen och det digitala materialet. Genom att se på kvalitativ- respektive massdigitalisering så ger det inblick i framtidsutsikterna för de båda förhållningssätten och hur de kan tillämpas, och vilka eventuella fördelar och nackdelar det finns av momenten i respektive digitaliseringsarbete.

Biblioteks- och informationsvetenskap är ett tvärvetenskapligt forskningsområde och på samma sätt sträcker sig ämnet digitalisering över många vetenskaper. Digitaliseringens betydelse för biblioteken och viss mån biblioteks framtid gör det betydelsefullt att vara medveten om vad de digitala teknikerna innebär och hur bibliotek kan dra nytta av det. Hur bibliotek med digitalisering bäst kan anpassa materialet för en vidare allmän användarskara men fortfarande kunna tillgodo se behovet hos forskare och specialintresserade.

1.2 Problembeskrivning

De senaste åren har det startats många digitaliseringsprojekt. En del berör endast digitalisering av ett fåtal verk, men där den digitala reproduktionen är trogen originalet så långt det är möjligt med hög upplösning och korrekt färgrepresentation i den digitala bilden och med digitala texter som överensstämmer med den analoga versionen. Andra projekt har målet att digitalisera hela, stora, dokumentsamlingar. Generellt kan det förstnämnda betecknas kvalitativ digitalisering och den senare som massdigitalisering. Termerna och fenomenet ska dock inte ses som motpoler utan som två förhållningssätt till att lösa en digitaliseringsuppgift.

Massdigitalisering anses vara en digitaliseringsmetod där arbetsflödet är effektivt, med högt tempo, men att det samtidigt sätts frågetecken för kvalitén på de digitala bilderna. Reflektionerna förs speciellt med tanke på om digitaliseringen utförs med syftet långtidsbevaring och ersättning av ett analogt/original dokument. Den kvalitativa digitaliseringsmetoden görs med huvudsyfte att det digitala dokumentet ska kunna ersätta en analog version och att kvalitén därför måste vara mycket hög. Detta är särskilt kännbart tidsmässigt vid större dokumentsamlingar, eftersom varje dokument tar lång tid att digitalisera och bearbeta.

Kvalitativ- och massdigitalisering skiljer sig därmed åt om vad som värdesätts. Ska digitaliseringen gå snabbt så sänks kravet på högsta möjliga kvalitét, och vill man ha högsta kvalitét så utökas arbetstiden. De flesta digitaliseringsprojekt utförs med en begränsad

budget och måste därför balansera de ekonomiska medlen med hur digitalisering kan utföras av en samling. I ett projekt kan valet därför stå mellan att använda olika tekniker och moment, där enskilda utföranden inte nödvändigtvis fullständigt följer principen om exempelvis massdigitalisering, men att hela digitaliseringsarbetet trots det kan anses som massdigitalisering.

Förhållningssättens problem är att dagens tekniska utrusning inte kan presterar tillräckligt. Det kan vara ett övergående problem som löser sig inom de närmsta åren, men faktum kvarstår, många digitaliseringsprojekt utförs och har utförts med de givna premisserna. Från att bibliotek tidigare använt mer kvalitativa metoder så har de på senare år börjat snegla på massdigitaliseringsarbeten som Google Book Search (Svärd 2006, s.55f.) Vad digitaliseringen innebär för de digitala dokumenten är fortfarande osäkert, men klart är att Google har fått kritik där kvalitén ifrågasatts, men också fått ros för deras vilja att tillgängliggöra litteratur och texter som endast finns på ett fåtal bibliotek.

1.3 Avgränsning

Uppsatsen fokuserar främst på digitalisering av bild, men också i viss mån digitalisering av text och textkodning. Dokumenttyper som berörs är företrädesvis böcker men också fotografier/bilder och i delvis tidningar tas upp.

I uppsatsen kommer jag inte att ta upp ”född digitalt”-material, alltså dokument som från början är digitalt skapat. Dessutom utelämnas digitalisering som rör ljud, som musik, radioprogram m.m., och allt som har med video att göra. Digitalisering av video och ljud är förvisso ett väldigt nytt forskningsområde och därmed intressant, men hålls utanför omfånget för denna uppsats.

1.4 Syfte

Syftet med föreliggande uppsats är att undersöka två olika förhållningssätt vid digitalisering av kulturarvet såsom det framträder vid Nasjonalbiblioteket (NB) i Norge och Kungl. biblioteket (KB) i Sverige. Jag undersöker dessa nationalbibliotek i avseende på bitar av deras fullständiga digitaliseringsarbete, nämligen: NB:s massdigitaliseringsarbete och KB:s kvalitativa digitaliseringsarbete. Undersökningen fokuserar på hur digitaliseringsprocessen och arbetsflödet skiljer och hur den digitala texten och digitala bilden då särskiljer sig ifrån varandra i avseende på de två metoderna.

1.5 Frågeställningar

- Hur skiljer digitaliseringsprocessen sig åt när den sker med mass- respektive kvalitativ digitalisering?
- Hur skiljer de digitala dokumenten tekniskt och bildmässigt sig åt när digitaliseringsprocessen sker med mass- respektive kvalitativ digitalisering?

I termen ”digitaliseringsprocessen” inkluderar jag: syfte, urval, förberedelse, bildfångsten, bearbetning, metadata, uppmärkning, kvalitetsbedömning, lagring och framvisning.

1.6 Litteratursökning och källkritik

Litteratursökning för uppsatsen gjordes inledningsvis i högskolans egna samlingar. Jag läste andra magisteruppsatser som behandlade digitalisering och hämtade inspiration från litteraturen de använt. Jag har dock försökt stå fri i hur deras uppsatser är strukturerade och vad de i texten tagit upp. Utifrån böcker om digitalisering och deras referenser har jag utökat min litteratursökning. I databaser som Library and information science (LISA) och ACM Digital library utfördes sökningar, men i uppsatsen är de få referenser som härrör från dessa. Diskussionslistan BIBLIST har varit till hjälp, där texter och hänvisningar om digitalisering frekvent postas. Jag har parallellt med magisteruppsatsen läst en 15hsp- kurs vid högskolan - ”Digitalisering av kulturarvet”. Mycket av innehållet i den kursen återfinns även i min uppsats, främst litteratur men också uppsatsen inriktning mot massdigitalisering och kvalitativ digitalisering har inspirerats av kursen. Jag har också fått tips på relevanta texter av lärare på högskolan.

Både NB och KB har egna manualer och rapporter med digitaliseringsrekommendationer för bibliotek i respektive land. Då dessa manualer är en del av mitt studieområde så har jag undvikit att använda dem i avsnittet om digitaliseringsprocessen. De behandlas istället i resultatavsnittet, tillsammans med artiklar från respektive lands bibliotekstidskrifter.

1.7 Metod

I detta avsnitt beskrivs metoden som används i uppsatsen, hur jag valt och vilket urvalet är till det empiriska materialet och hur jag förhåller mig till det. Informanterna jag använder i undersökningen tas upp och hur frågorna till dem formuleras och använts. Vidare nämner jag hur den komparativa analysen ska utföras.

Uppsatsen är en komparativ studie av massdigitalisering och kvalitativ digitalisering utifrån två nationalbibliotek och deras befintliga digitaliseringsarbete. I uppsatsen används främst dokumentstudier av både officiella och interna dokument rörande digitalisering. När dokumenten inte ger tillräckligt med information används e-postintervjuer, vilka även används för att verifiera överensstämmelsen av dokumentens innehåll med faktiska förfaranden. Institutionernas egna webbplatser beskrivs, och på vilket sätt digitaliserat material visas och vilken åtkomst användarna har.

De två nationalbibliotek som valts är Kungl. biblioteket i Sverige och Nasjonalbiblioteket i Norge. Valet föll på dessa bibliotek av flera grunder. Att som i Nasjonalbiblioteket ha målsättningen att digitalisera samtliga verk är väldigt ambitiöst, och är enligt deras utsago det första europeiska nationalbiblioteket med dessa planer (Skarstein 2006). Det gör det intressant att se vilket tillvägagångssätt de har. Att låta Kungl. biblioteket stå för den kvalitativa digitaliseringen är resultatet av att det är i Sverige jag som uppsatsförfattare bor och känner till biblioteksförhållandena, men även för det faktiska digitaliseringsarbetet som bedrivs på nationalbiblioteket. Jag kan även nämna att jag som ettårig utbytesstudent i Norge också kom nära deras biblioteksverksamhet. Det bör förtydligas

att båda institutionerna använder sig av flera digitaliseringsförfaranden. Det är inte så att NB:s digitalisering endast är massdigitalisering, och att digitaliseringen hos KB endast är kvalitativ digitalisering, men jag har valt dessa avsnitt av deras verksamhet.

Kungl. biblioteket har digitaliserat i många år och med olika förfaranden. Ett av deras digitaliseringsarbeten måste därför väljas ut och fungera som exempel och informationsunderlag i undersökningen. Den digitalisering jag valt att använda är den som gjordes av *Codex Gigas*, även kallad *Djävulsbibeln*, ett av KB:s senast digitaliserade verk. Valet bygger delvis på ett förslag jag fick av KB, då det för dem var en speciell digitalisering, som krävde särskilda lösningar.

Codex Gigas betyder ”den väldiga boken”. Verkets storlek är 92x50,2 cm och väger ca.75 kg. Den består av 312 pergamentblad, alltså 624 sidor. Boken är ifrån 1200-talet och är skriven helt på latin. Den består av bl.a. texter från gamla och nya testamentet. Verket togs av svenskarna som krigsbyte i Prag under trettioriga kriget. Sedan 1649 har det tillhört KB:s samlingar och är troligen ett av Kungl. bibliotekets mest välbekanta verk (Codex Gigas; Official website of the Czech Republic 2007).

Då uppsatsen ämnar ta upp skillnader och likheter av digitaliseringen så är det viktigt att institutionerna får beröra samma områden och moment inom digitalisering. Det finns en klar risk i att dra felaktiga slutsatser utifrån att dokumenten och intervjuerna behandlar momenten olika mycket och att institutionernas arbete egentligen är mer djupgående än vad dokumenten ger sken av. Då digitaliseringsprocessen redan från början är väldigt omfattande så måste beskrivningar göras av vad jag kan ta upp. Detta ska dock inte behöva ske på bekostnad av att skillnader och likheter försummas.

Litteraturen som har används i det deskriptiva avsnittet om digitaliseringsprocessen är inte fullständig på så sätt att alla åsikter och vinklingar dryftats. Mycket av materialet rör utförandet, där många av författarna instämmer med varandra. De mer tekniska aspekterna skiftar p.g.a. snabbare utveckling, och sådana rekommendationer skiljer också i litteraturen.

Från början var tanken att jag ensidigt skulle använda textdokument om institutionernas arbeten och förhålla mig till dess text och innehåll som sanning. Förhållningssättet orsakade dock snabbt problem i och med att texterna inte alltid är entydiga med hur arbetet utförs och att det skiljer mellan rapporter och utgivningsår. Detta har troligtvis helt naturliga orsaker p.g.a. att digitalisering och momenten utvecklas och skiftar i snabb takt. Något som stöds av bibliotekens egna dokumenttexter där de anger att arbetet ständigt förändras. Min utgångspunkt till dokumenten har därför förändrats till att jag kontrollerar texternas innehåll och frågar institutionerna om dess aktualitet, om vad som stämmer överens med digitalisering som i dagsläget utförs. Problem som kan förekomma vid dokumentstudier är att texterna kan skönmåla det egna arbetet och undviker att berätta förekomna problem. Hur sanningsstatusen är på dokumenten som används i uppsatsen kan därför vara svårt att uppskatta eftersom informanterna kan undvika att ta upp problem. Det kan även föreligga en skillnad i vad de säger att de gör/bör göra och hur de faktiskt går tillväga, och att dokumenten och informanternas svar därmed inte helt stämmer överens med verkligheten. Det kan göra resultatet skevt, med påföljd att det i analysen och diskussionen dras ofullständiga och/eller felaktiga slutsatser.

Hur representativt resultatet i undersökningen är för övriga digitaliseringsarbeten som utförs i världen, oavsett massdigitalisering eller kvalitativ digitalisering, kan diskuteras. Ett digitaliseringsarbete består av många val av bl.a. metoder och utföranden. Ett arbete som säger sig vara det ena eller det andra behöver inte stämma överens med resultat och påföljande analys av just KB eller NB:s, där man kan ha tagit olika beslut och ha olika kriterier. Däremot kan uppsatsen ge en viss indikation om de olika digitaliseringsförhållandena och dess inverkan och användningsområden.

Jag har haft en kontaktperson som utsågs av de själva på vardera nationalbibliotek. De har besvarat mina frågor och i KB:s fall bistått med interna dokument och andra textkällor. Hos KB blev jag även hänvisad till en person med mina frågor om projektet *Codex Gigas*.

Informant A: är anställd hos NB och har stort inflyttande över digitaliseringsarbetet.
Informant B: är anställd hos KB och har arbetat länge med digitalisering och lagring.
Informant C: är anställd hos KB och bistod med svar om arbetet med *Codex Gigas*.

Uppsatsens karaktär gör det dock troligt att de rådfrågat fler personer inom digitaliseringen. Om så inte har skett kan svaren som framkommer på frågorna vara något vinklade p.g.a. att de inte på detaljnivå är insatta i alla delar av digitaliseringen. Då jag i studien redogör för faktiska digitaliseringsutföranden ser jag inga andra risker i vem som ligger bakom svaren. Om mer personliga uttryck och kommentarer angående digitaliseringen hade legat i fokus skulle däremot det vara möjligt. Olika enheter för digitaliseringen kan ha skilda åsikter om hur det utförs, men är alltså inget som tas i beaktande i uppsatsen.

Frågorna har fokuserat på områden där textdokument saknat information eller varit otydliga om förfarandet. Frågorna har skickats och blivit besvarade med e-post från respektive kontaktperson. Då förekomsten av information skiljer har biblioteken inte ställts inför samma frågor. Målet har dock varit att varje avsnitt ska beröras av biblioteken. Typen av frågor varierar, från öppna till mer slutna. Min förhoppning var att institutionerna skulle beskriva sin verksamhet och val mer utförligt än vad som ibland blev fallet, där exempelvis ett svar kunde bestå av ”ja” när jag egentligen ville veta mer om deras kvalitetskontroller. Jag måste därför förtydliga och dela upp frågan i mindre bitar. Det var främst svaren från NB som var kortfattade och att de inte alltid var villiga att gå in på detaljer. Många av frågorna bygger på mina tidigare ställda frågor och svar. Jag bifogar dem därför inte i uppsatsen då jag bedömer att de kan vara svåra att tyda utan de tidigare svaren. En del frågor liknande mer påstående av typen: ”Stämmer det att ni går tillväga ..?”. Till KB ställde jag få frågor men angav istället ämnesområden där de skulle delge mer information, ex. om vilken metadata som används, hur deras lagringssystem är uppbyggt, speciellt med åtanke långtidsbevaring. Frågor som exempelvis ställdes till NB var:

”Är det ett politiskt önskemål från regeringen att NB skulle digitalisera dokument rörande nordområdet?”.

”Skannas alla böcker i 400 dpi oavsett dokumentstorlek, detaljrikedomen i text eller illustrationer?”.

”Vilka delar av samlingen kan bli aktuell för avancerad strukturanalys, som det nämns om på sida 8?”.

”Vad för sorts kvalitetskontroll sker av den digitala bilden? Hur ofta sker kontrollen och vad är det som undersöks?”.

”Har accessbilderna för nätvisning förändrats för att höja/förenkla läsbarheten?”.

Biblioteken har även fått möjlighet att läsa igenom resultatavsnittet om deras arbeten, främst för att se att det stämmer med deras förfaranden. Ingen av dem gjorde några tillägg.

Uppsatsens analys sker i ljuset av texterna från dels kapitel tre, om bl.a. digitalisering och analoga och digitala skillnader, och dels från kapitel fyra med det deskriptiva avsnittet om digitaliseringsprocessen. Svaren på fråga ett presenteras för vart och ett moment. De tydligaste särdragen mellan förhållandena grupperas därefter i kategorier. Eftersom uppsatsen fokuserar på jämförelsen av digitaliseringsmetoder så ser jag detta sätt som mest lämpligt. Den andra av uppsatsens frågeställningar rör konsekvensen av digitaliseringen och resultatet av den digitala filen. Jag jämför då de digitala dokumenten, med tanke på framtiden och dess användning.

1.8 Disposition

Det första kapitlet inleder uppsatsen och beskriver uppsatsen med problembeskrivning, syfte och uppsatsens frågeställningar. Vidare formuleras hur uppsatsen genomförs med litteratursökning och ett källkritiskt förhållande. Inom samma kapitel formuleras även uppsatsmetoden, dels dokumentstudien men också e-postkontakter. Nästa korta kapitel tar upp de viktigaste texterna jag använt i uppsatsen. I nästkommande kapitel beskrivs digitaliseringen i sin helhet och vad som ses som nackdelar respektive fördelar. I kapitlet finns också avsnittet som beskriver skillnaden mellan begreppen massdigitalisering och kvalitativ digitalisering. Därefter beskrivs hela digitaliseringsprocessen och alla moment den innehåller. Denna indelning används även i påföljande resultatavsnitt. Resultaten för undersökningen presenteras uppdelat på de två institutionerna. I kapitel 6 analyseras resultaten och bearbetar vad det innebär utifrån frågeställningarna, och i näst sista kapitlet diskuteras analysen och slutsatser dras av studien. Uppsatsen avslutas med en sammanfattning av hela undersökningen och uppsatsen.

2. Litteratur och tidigare forskning

Nasjonalbiblioteket publicerade i september 2007 det 14-sidiga dokumentet *Digitalisering av böcker i NB: Metodikk og erfaringer* som har legat till grund för mycket av mitt material om NB:s digitalisering. Den tar upp hur digitaliseringen av böcker utförs och beskriver sambandet mellan förväntningar och erfarenheter av arbetet, hur upplägget för digitalisering ändrats under tiden verksamheten varit igång. Den nämner även motgångar och problemen som de haft.

Material som finns angående Kungl. biblioteket och deras verksamhet är av mer blandad karaktär, men mer omfattande. KB har publicerat ett par större rapporter, med un-

dersökandekarakteristik om vilka förfaranden som är aktuella att använda och inte främst hur de faktiskt går eller har gått tillväga. Informationen om hur digitaliseringen har genomförts har till stor del utgått ifrån en intern manual kallad *Digitaliseringsmanual*, v2. Det är ett 114-sidigt dokument (inkl. bilagor) som tar upp tillvägagångssättet vid digitalisering som används vid projektet *Öppna samlingar*. Dokumentet innehåller referenser från tidigare projekt och hur digitaliseringsarbeten vid KB generellt ska utföras, vilka krav som fastställts, m.m. Det finns inget entydigt datum för manualen, men dateringen av sammanfattningen är ifrån 2007-04-10 och är författad av Viktoria Enmark.

Catrin Persson och Annevie Tångemar skrev 2006 magisteruppsatsen *Varför digitalisera?: En studie av tillkomsten av Kungl. Bibliotekets digitaliserade samlingar*. De går igenom KB:s digitaliseringsprojekt och granskar syftet för var och ett av dem, tillsammans med intervjuer av medverkande. De diskuterar även urvalet i större omfattning. De kom fram till att KB har digitaliserat i övervägande bevarandesyfte men även tillgängliggörande, speciellt av sköra dokument och eftertraktade studieobjekt.

Fler dokument som används om KB:s digitalisering tas upp i samband med resultatet.

Moving theory into practice: Digital imaging for libraries and archives (2000), redigerad av Anne R. Kenney och Oya Y. Rieger är en bok med bidrag från ett flertal författare. Den tar upp digitalisering och hur teorin kan bli praktiserad. De skriver om teman t.ex. urval, bevarande och åtkomst. De vill med boken utveckla ett kritiskt tänkande i en teknisk värld hos bl.a. bibliotekarier, arkivarier, teknologer och andra som arbetar med kulturell lagring. Detta för att de ska kunna utvärdera accepterade teoretiska konstruktioner och implementera strategier som reflekterar institutionens uppdrag och kapacitet (2000b, s.2).

I *Digitizing collections: strategic issues for the information manager* (2004) tar Lorna M. Hughes upp digitaliseringsprocessen och hur den strategiskt ska organiseras och utföras. Boken behandlar hur digitaliseringsprojekt och policys ska utformas och hur kostnaderna ska hanteras. En stor del ägnas också åt copyright och rättigheter i samband med digitalisering.

Digital futures: strategies for the information age (2002) av Marilyn Deegan och Simon Tanner är en introduktion och översikt till digitaliseringen och digitala bibliotek. De går igenom digitala samlingar och hur de byggs upp, ekonomiska frågor, metadata, standarder, digitaliseringsprogram, digitalt bevarande och hur biblioteksrollen förändras i och med den digitala utvecklingen. Boken redogör även för olika biblioteksprojekt och hur de arbetat med de digitala medierna och digitalisering.

Oya Y. Rieger har i *Preservation in the age of large-scale digitization: A white paper* (2008) utvärderat storskaliga digitaliseringsarbeten och hur de inverkar på tillgänglighet och användbarhet över tid och de digitala böcker som projekten skapar. Rieger tar upp projekten Google Book Search, Microsoft Live Search, Open Content Alliance och Million Book Project. Hon beskriver deras digitaliseringsstrategier och behandlar dem utifrån bl.a. kvalitet, bildfångst, förpliktelser och genomföranden av arkivinstitutioner och hur villiga de är att samarbeta. Hennes paper försöker förutse den påverkan som storskalig digitalisering utgör på boksamlingar och konkluderar med att kulturella institutioner måste samarbeta eftersom inget enskilt bibliotek har råd att utföra ett arbete i liknande skala som, ex. Google Book Search.

I artikeln *Mass digitization of books* från 2006 tar Karen Coyle upp massdigitalisering, dess bakgrund, syfte och utförande, och redogör även för ett par digitaliseringsprojekt, speciellt Google Book Search. Hon ser möjligheten till att digitalisera i stor skala p.g.a. av förbättringarna av bildfångsten och konverteringen till digital text (Optical Character Recognition). Bildfångst av böcker kan enligt Coyle ske utan att boken tas isär och utan skaderisk eller andra påfrestningar. Med en digitalkamera är det möjligt att ta bilder ovanifrån och den kan därefter automatiskt korrigera böjningar och vridningar av boksidan, så att den blir rak. Bearbetningar av exempelvis upplösning kan även utföras av mjukvaruprogram. P.g.a. minskade mänskliga moment kan en timmes bildfångst konvertera mellan 1200-3000 sidor. Massdigitaliseringsarbeten och dess teknologi och managementfrågor måste enligt Coyle studeras mer. Hon lyfter speciellt fram: "workflow", "output" och bokstruktur, användargränssnitt, standarder, bevarande, samt att biblioteken måste visa större hänsyn till objekts egenhet vid digitalisering. Artikeln konkluderar med att trots många massdigitaliseringsprojekt existerar så finns det få idéer om användandet av digitaliserade böcker, vilka som tjänar på digitaliserade bibliotek, hur tjänar användarna på det och hur reagerar systemet när sökningar sker i en samling på 10 miljoner böcker? Coyle tycker sig också se tendenser att biblioteksledare lockas av massdigitalisering men är oförmögna att ge åtkomst till böckernas innehåll. Om massdigitalisering är bästa sättet att tillgängliggöra så anser Coyle att förlikningar måste ske mellan ekonomin av massdigitalisering, och tillfredsställa biblioteksanvändares behov.

3. Teoretisk ram för digitaliseringen

I kapitlet redogörs för dokumentmaterial som en bibliotekssamling kan innehålla och hur analoga dokument skiljer sig ifrån digitala. Varför institutioner väljer att digitalisera eller att låta bli. Massdigitalisering och kvalitativ digitalisering tas sedan upp, både som begrepp och vad de innefattar.

3.1 Text och bild

I ett nationalbiblioteks samling förekommer material av skiftande omfång och form. Textdokument har under åren skrivits antingen för hand eller med hjälp av tryck. Använda material har exempelvis varit: papper, papyrus, trä, sten. Formaterna varierar från tidningar, tidskrifter till brev, musikaler, m.m., där typsnitten varierar. Stillbilder och visuellt material existerar på material som: papper, glas, textilier, sten och kanvas. Dessa innehåller t.ex. målningar, teckningar, manuskript bilder, fotografier, kartor, glasmålning, satellitbilder. Alla olika material har sina egna förutsättningar och krav på bildfångsten (att konvertera till digitalt format) och vilka tekniker som bäst anses återskapa originaldokumentet. Dokument är dessutom ofta hybrider och kan t.ex. innehålla både bild och text, enligt Hughes krävs då flera tekniska lösningar (2004, s.255f.).

3.2 Analogt och digitalt

Digitalisering är att konvertera analogt material till digitalt, även kallad digitaliseringsprocessen. I begreppet innefattar jag följande moment: syfte, urval, förberedelse, bild-

fångst, bearbetning, metadata, uppmärkning, kvalitetsbedömning, lagring och framvisning.

Det analoga objektet benämner jag i uppsatsen som *originalet*. Produkten som blir av bildfångsten betecknas som *digitala originalet*, men kan även kallas *digital master*, *masterfil*, *masterbild* eller *digitala kopia* (av analoga originalet).

Det analoga materialet, eller mediet som Hughes kallar det anser hon kännetecknas av tre punkter:

- Bundna till fysiskt medium, vilket innebär ett linjärt förlopp, en start och slutpunkt.
- Bunden till en representation som bestäms av upphovsmannen.
- Materialet försämras vid kopiering.

De digitala mediernas kännetecken anser Hughes är:

- Går att länka till andra medier och på så vis skapa multimedier.
- Inte bundna av rumsliga eller temporära barriärer och kan därför lagras och förflyttas på många sätt.
- Materialet kan kopieras oändligt många gånger utan att försämras.
- Det är möjligt att komprimera.
- Lätt att browsa och söka i (2004, s.4).

3.3 Om digitalisering och digitaliseringsprocessen

Begreppen digitalisering och digitaliseringsprocessen (även kallad digitaliseringsprojekts livscykel) innefattar moment som gör ett analogt objekt till digitalt format. Det innebär enligt Lee urval av material, materialförberedelse, bildfångst, efterbearbeta, framvisa, ge support och underhåll av lagring av filer och system (2001, s.8). Hur troget originalet representeras av den digitala kopian är helt beroende av originalets storlek, form och tillstånd (Hughes 2004, s.255).

En viktig fråga om digitalisering är varför man ska göra det, vilka fördelar ger det? I detta avsnitt redogör jag för de vanligast förekommande åsikterna och även varför eller när man inte ska digitalisera.

3.3.1 Argument för att digitalisera

Det finns i huvudsak ett par återkommande svar om varför man bör digitalisera. Där tillgängliggörandet till en vidare krets människor är bland det första som nämns. Tanken är också att det digitala materialet ska locka intresse för biblioteket och resterande samlingar. Objektet som digitaliseras är många gånger unikt och existerar endast i ett fåtal exemplar. Om det görs digitalt (och publiceras på internet) ökar möjligheten till studier av verket, trots att de inte håller objektet fysiskt i handen. Demokratiaspekten är uppenbar, alla kan nu ta del av ett värdefullt dokument (Hughes 2004, s.9; Lee 2001, s.5). Det andra skälet för digitalisering är bevarandenaspekten. Om inte dokumentet tas tillvara så

vittrar det sönder och blir till slut obrukbart. Genom att digitalisera verket så bevaras det, om än i digital variant. Det har däremot konstaterats att om ett dokument görs digitalt för internetvisning, så ökar också kunskapen om dess existens vilket drar större intresse till originalet. Originalet kan då bli mer använt än vad det blev innan digitaliseringen (Lee 2001, s.6). För att motverka överanvändningen av den analoga versionen bör den digitala kopian vara i tillräckligt hög kvalitet för att fylla de flesta intressenters behov. Det råder dock en viss skepsis till att se digitalisering som bevarande då tekniken ständigt utvecklas och där datalagringen av digitalt material inte håller tillräckligt länge. Hughes uppskattade exempelvis 2004 att efter ca. 3 år måste digitalt materialet flyttas till nytt lagringsmedium (s.7, 11). Vid digitalisering kan samlingar som fysiskt är ifrån varandra, men ursprungligen hör ihop, återförenas. När material blir digitalt innebär det möjligheter att länka och bygga struktur till övriga digitala källor och på så vis göra nya sammankopplingar (Hughes 2004, s.4, 11f.). Deegan och Tanner nämner ytterligare fördelar som, att det lättare går att söka i texterna och kanske mer kontroversiellt, att redigera och förbättra (eller förändra) en bild, t.ex. att ta bort smuts, färgstick. Något som troligtvis inte fanns med i den ursprungliga bilden (2002, s.32f.)

En institution kan välja att digitalisera p.g.a. strategiska och institutionella fördelar. Om digitaliseringen sker utifrån ett tema eller speciell profil så kan institutionen stärkas och bl.a. ge tillström av finanser (Hughes 2004, s.13f.). När bibliotekets material blir digitaliserat kan det innebära minskad påfrestningen för biblioteket och övrig utrustning. Biblioteken behöver t.ex. inte hantera böckerna och kopiering av dem. Personalen utvecklar också nya färdigheter när de handskas med uppgifter som tidigare inte ingått i deras verksamhet (Lee 2001, s.26ff.). Hughes tar upp hur digitalisering kan användas för forskning och utbildning, ifall man presenterar det med riktat upplärings syfte med t.ex. särskilda teman. I användandet av digitala kopior inom forskning ses den största fördelen och möjligheten ligga inom den dynamiska virtuella miljön. En miljö där olika typer av medium samverkar och leder till nya slutsatser. Dessutom nämns exempel på hur digitaliserat material med avancerade fototekniker fått fram text från dokument vars innehåll varit osynligt för det mänskliga ögat (Hughes 2004, s.16).

Magisteruppsatser skrivna vid Högskolan i Borås som studerat enskilda bibliotek och museer och vilka argument biblioteken angett till att digitalisera visade att vanligaste svaret var bevarande och tillgängliggörande (Emanuelsson 2006, s.42; Andersson & Nilsson 2006, s.52f.).

3.3.2 Argument mot att digitalisera

Viktigaste orsaken till att inte digitalisera är den ekonomiska belastningen, då digitalisering är mycket kostsamt (Besser 2003, s.31; Lee 2001, s.6f.). Utrustningen som används är dyr och digitalisering är ett område som fort förändras, vilket gör att utrustningen snabbt blir föråldrad. En stor del av kostnaden ligger på bevarande av de digitala dokumenten, för vilka det krävs mer än för analogt material. Då man inte vet vad som ligger i framtiden för digitalisering så kan det medföra att digitalisering som görs är med fel metoder och senare måste göras om, vilket leder till stora kostnader (Deegan & Tanner 2002, s.36, 57). De Stefano menar att digitaliseringen inte ska ske utifrån syftet att bevara dokument, då digitala filer fortfarande har en kort livslängd i jämförelse med mikrofilm och papper (2000, s.22). I de fall där tekniken inte är fullt utvecklad kan det vara bättre att avstå ifrån digitalisering. Upphovsrätt och copyright aktualiseras vid digitali-

sering då dessa frågor inte ses på samma sätt som bibliotekets fysiska utlån av dokument. Rättigheter kan behöva klargöras vid konverteringar till varje digitalt lagringsmedium (Deegan & Tanner 2002, s.205). Deegan och Tanner påpekar dock i sin bok att någon typ av kopia av originalet är bättre än ingen alls, i fall originalet förstörs (2002, s.187).

3.4 Massdigitalisering och kvalitativ digitalisering

Digitalisering har under en längre tid existerat och utförts med varierande krav och föreställningar om vad dess syfte är. Digitalisering som metod kan därför inte utföras eller betraktas som enhetlig massa. De senaste åren har det dock börjat utkristalliseras digitaliseringsmetoder och vad metoden innefattar och står för. Två metoder är ”massdigitalisering” och ”kvalitativ digitalisering”, vilka även kan kallas ”kvantitativ digitalisering” och ”kritisk digitalisering” (Dahlström & Hansson 2008, s.112f.). Jag ska i detta avsnitt beskriva vad metoderna vanligtvis associeras och karaktäriseras av.

Efter att Google 2004 lanserade projektet Google Book Search har också termen ”massdigitalisering” varit dess följeslagare. Hur frekvent förekommande termen varit innan Googles projekt vet jag inte, men förefaller utifrån Coyles artikel blivit ett begrepp och en digitaliseringsmetod (2006, s.641). Syftet med Googles digitaliseringsarbete är att tillgängliggöra litteratur, att det ska vara möjligt att ta del av det mesta digitalt (Google). Böckerna som Google skannar tillhör ett antal partnerbibliotek, mestadels amerikanska. För att det skulle bli tekniskt och ekonomiskt möjligt att inom rimlig tidsram utvecklades digitaliseringsprocessen till massdigitalisering (Coyle 2006, s.641).

Massdigitalisering syftar till att konverteringen från analogt till digitalt sker i industriell skala med enhetlig metod och med stora kvantiteter material. För att arbetet ska anses ekonomiskt lönsamt är arbetsprocessen hög, vilken innebär att mänskliga (manuella) procedurer undviks och utförs istället automatiskt och maskinellt (Coyle 2006, s.641f.). Av praktisk hänsyn undviks tolkande moment och processen görs två-dimensionellt linjär (Dahlström & Hansson 2008, s.112). Den digitala filen i massdigitaliseringsprojekt blir sällan kvalitetskontrollerad. Det enda som möjligtvis görs är enstaka stickprov. Den digitala text (OCR-text) som skapas av den analoga texten är sällan tillrättalagd och bearbetad, och om det körs någon kontroll så är det på maskinell väg. När materialet blivit digitalt är det ovanligt att det struktureras i samlingar (Coyle 2006, s.642; Rieger 2008, s.22f.).

”Kvalitativ digitalisering” eller ”kritisk digitalisering” är ett nytt begrepp och som bland annat förespråkas av Dahlström och Hansson, som använder båda termerna. Kvalitativ digitalisering menar de involverar djup textkodning, kritisk bild- och textredigering, och rik informationstilldelning vid urvalskriterier och en strategi tillämpad för enskilda dokument (2008, s.112). Karen Coyle väljer i sin artikel om massdigitalisering att endast kalla det ”non-mass digitization”. Med det avser hon digitalisering med ett noga övervägt urval med syfte att skapa representativa kopior utav originalet. Vidare beskriver hon hur denna typ av digitalisering kan producera en OCR-text rik på uppmärkning, vilket innebär att texten och delar av den kan användas i flera sammanhang (2006, s.642). Coyle skiljer på massdigitalisering och storskaliga digitaliseringsarbeten (large-scale digitization) där hon menar att de sistnämnda utförs med ett visst omdöme om

skapande av samlingar, hela samlingar. Dessa arbeten utförs ofta i ett lägre tempo, men att det fortfarande rör sig om stora samlingar (Coyle 2006, s.642).

Av texten ovan framkommer att skiljelinjen mellan massdigitalisering och storskalig digitalisering är otydlig. Rieger använder i sin artikel termerna ”mass digitization” och ”large-scale digitization” ekvivalent, jag väljer samma linje men kallar digitaliseringen endast massdigitalisering (2008, s.4). Att som Coyle använda termen ”non-mass digitization” uppfattar jag som otympligt och följer istället Dahlström och Hansson linje och använder ”kvalitativ digitalisering” (2008, s.112f.).

4. Teknisk ram för digitaliseringsprocessen

I detta kapitel redogörs digitaliseringsprocessen, då processens logiska struktur följs, tillika ordningen ett digitaliseringsarbete vanligtvis utförs. Varje moment i processen bygger vidare på den föregående och inverkar på nästkommande beslut. Digitalisering är komplex och består av mycket samarbete med folk ifrån flera sfärer, bl.a. konservering, digitaliserare och organisatörer (Kenney & Rieger 2000b, s.6; Hughes 2004, s.36). Att arbetet är komplext innebär att vissa avsnitt i kapitlet går i varandra och att upprepningar därmed förekommer, t.ex. så görs kvalitetsbedömningar ofta genomgående under digitaliseringen, men jag väljer att beskriva det i ett eget kapitel.

4.1 Syfte och urval

I inledningsskeendet måste varje digitaliseringsprojekt bestämma vad dess syfte är. Beslutet att digitalisera kan enligt Lee komma från utomstående, externa krav, ex. politiker, vilket leder till *reactive digitization*. Då sker digitaliseringen utifrån behov hos utomstående aktörer, och ämnar främst göra dem nöjda. Digitaliseringsbesluten ligger då mer eller mindre utanför institutionen och beställarnas behov och krav styr digitaliseringsarbetet. Den mest extrema formen är - on-demand-digitalisering. Det motsatta är *proactive digitization*, vilket innebär att institutionen har funnit behov för att starta ett digitaliseringsprojekt, och att beslutet om utförande ligger hos institutionen. Vanligast är ett kombinerat synsätt, t.ex. att ekonomisk hjälp ges men att urvalet och tillvägagångssättet bestäms av institutionen (Lee 2001, s.11).

Det uppgavs tidigare i uppsatsen att digitalisering kan innebära att biblioteket når fler användargrupper. Att digitalisera verk som inte uppfyller detta mål kan anses som dålig investering av tid och pengar. För att veta vilka dokument användare vill se digitalt är det viktigt att veta dess värde. Välanvända analoga dokument kan man även anta blir det i digital form. Om den digitala bilden görs i så hög kvalitet att originaldokumentet inte behöver användas så bevaras den ifrån onödigt slitage, även om kopian drar till sig större intresse än tidigare (Hughes 2004, s.42). I fall den digitala kopian är av otillräcklig kvalitet så leder det till både ökat intresse av verket och originalet, och därmed ökad påfrestning. Det är också populärt att fokusera på verks innehållsliga, historiska och intellektuella värde. Dessa verk är nödvändigtvis inte välanvända och kända men digitaliseringen kan öka intresset för dem. Dessutom finns verk vars fysiska värde eller skador gör att allmänhet inte får ta del av dem. Om dessa verk digitaliseras så kan intresserade ta del av objektet och innehållet, utan att originalverket riskerar skador (de Stefano 2000, s.15; Lee 2001, s.21f.)

Om digitalisering sker utifrån syftet att bevara sköra verk så måste man beakta att vissa inte klarar alla typer av bildfångstmetoder och skadas om de utsätts för viss sorts ljus (Hughes 2004, s.39). Bevarandet av dokumenten blir efter digitaliseringen en dubbelbörda då både den analoga och digitala versionen måste upprätthållas (de Stefano 2000, s.22).

Institutioner kan strategiskt använda digitalisering för att lyfta fram sin position inom sina specialområden. När verk av särskilt studieintresse digitaliseras dras blickarna till resten av arbetet. Utomstående aktörer kan därmed bli mer villiga att skjuta till pengar till projektet (Hughes 2004, s.47). Det finns också möjlighet att tjäna pengar genom att sälja kopior av den digitala filen, då de är enkla att distribuera (Lee 2001, s.27).

Vid urvalet är uppgiften att bestämma vad som digitaliseras och i vilken ordning det sker. Urvalet är med andra ord nära sammanbundet med syftet, där ett syfte som: ”tillgängliggöring av poesi från 1700-talet” medför att den givna samlingen av poesi ska digitaliseras i en viss ordning. Ordningen kan t.ex. vara utifrån populära författare, verk, kvalitet på de analoga dokumenten, praktfulla verk m.m. Vad som anses relevant kan m.a.o. bero på många aspekter (de Stefano 2000, s.11). En metod som går under benämningen ”cherry-picking” innebär att man digitaliserar de verk eller bitar av verk som är av störst intresse - guldkornen. Det kan t.ex. vara en populär dikt ifrån en bok eller artikel från en tidskrift. Man vet då att dokumenten blir använda och man får valuta för pengarna som lagts ner i projektet. Lee ifrågasätter däremot en sådan digitaliseringsmetod då han anser att hela verk bör digitaliseras samtidigt, men inser att det på grund av resursbrist ibland inte är möjligt (2001, s.14f.).

En viktig aspekt att utreda i början av ett digitaliseringsarbete är om materialet är upphovsrättskyddat, då kan tillstånd från utgivare vara nödvändigt, speciellt vid webbpubliserad. Tillstånd kan också krävas i framtiden vid ex. konvertering till andra medier (Deegan & Tanner 2002, s.238).

4.2 Föreberedelse, bildfångst och bearbetning

Förra sektionen handlade om digitaliseringssyften och urval. Där beskrevs hur ett angivet syfte påverkar hur resten av digitaliseringsarbetet fortskrider. Det är därför oundvikligt att skriva om bildfångst utan att de tidigare besluten kommer till ytan. Bildfångst är kärnan i digitaliseringsprocessen, i det momentet omvandlas (mer korrekt, kopieras) analogt till digitalt.

4.2.1 Bildens kvalitet

Det finns olika sätt att digitalt återge bilder, vanligast är s.k. rasterbilder (även kallad: bitmap) men det finns också vektorgrafik. I rasterbilder är bilden uppbyggd av en matris med enskilda pixlar, vilket innebär att gradvisa skillnader i

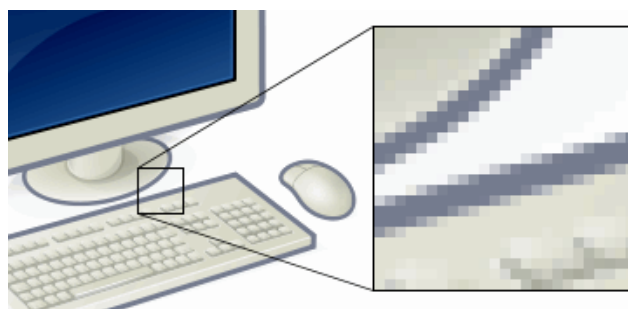


Fig. 1. Exempel på bilduppbyggnad med pixlar (Wikipedia-Pixel)

bildtoner kan visas. Däremot försämras bilden om storleken förändras. I vektorgrafik består bilden av matematiska formler som anger en ytas storlek och färgen den ska fyllas med. Då bilden består av formler innebär det att bildstorleken kan ändras utan försämring. Loger och teckensnitt är exempel vilka är uppbyggda av vektorgrafik. (Price-Wilkin 2000, s.117; Besser 2003, s.68, 82)

Bildkvaliteten anges och styrs av ett par begrepp för bildfångsten. När dokument är digitalt så representeras den av en lång sträng tecken bestående av "1" och "0", där ett tecken kallas för bit. För att kunna bearbeta och läsa strängarna krävs datorer (Deegan & Tanner 2002, s.6; Lee 2001, s.3f.)

Det finns flera upplösningstyper och där sammanhanget avgör vad som menas, skärmupplösning, skrivarupplösning, skannerupplösning m.m. Upplösning är ett relativt mått och inget absolut värde och har endast betydelse i sin egen kontext (Besser 2003, s.14). Bildupplösningen anger hur många pixlar/punkter bilden består av. Om en bild har storleken och upplösningen 400x600 så betyder det att bildens bredd är 400 punkter och längden är 600 punkter och innehåller 240.000 punkter. Vid skanning anges upplösningen vanligtvis i dots per inch (dpi) eller points per inch (ppi), hur många pixlar som går på en inch (1 inch = 1 tum = 25,4 mm). Högre upplösning innebär fler pixlar och vanligtvis bättre kvalitet och fler detaljer (Deegan & Tanner 2002, s.6f.). När en digital bild ses i sin helhet ser de flesta ingen skillnad på upplösningen 600 dpi och 1200 dpi. Däremot om det zoomas i bilden framkommer skillnaden tydligare och de enskilda pixlarna blir synligare i lägre dpi (Kenney 2000, s.29).

Varje punkt innehåller ett färgvärde: svart, vitt, grönt, grått etc. Ju fler färger en punkt kan ha desto större färgdjup har bilden. Färgdjupet beskrivs i bit, förkortning för bitmap. 1-bit betyder att pixeln kan ha två färger (svart & vitt), vilket är vanligt vid OCR-skanning, då man vill få tydliga bokstäver i ett textdokument. När en bild består av 4-bit kan varje pixel representeras av, 2^4 , 16 olika färger. Det kan förklaras med att fyra siffror ger sexton varianter då värdena utgår från binär skala. På samma sätt ger 8-bit, 2^8 , 256 färger och 24-bit, 2^{24} , ca. 16 miljoner kombinationer, och färger. (Lee 2001, s.38f.)

Ljus som delas av en prisma ger uppkomst till flera miljarder färger, människan uppfattar sju till tio miljoner av dem. Detta spektrum av nyanser tillsammans med intensiteten av ljuset ger de enskilda färgerna. Färgen skapas också av ytan den faller på, kontraster

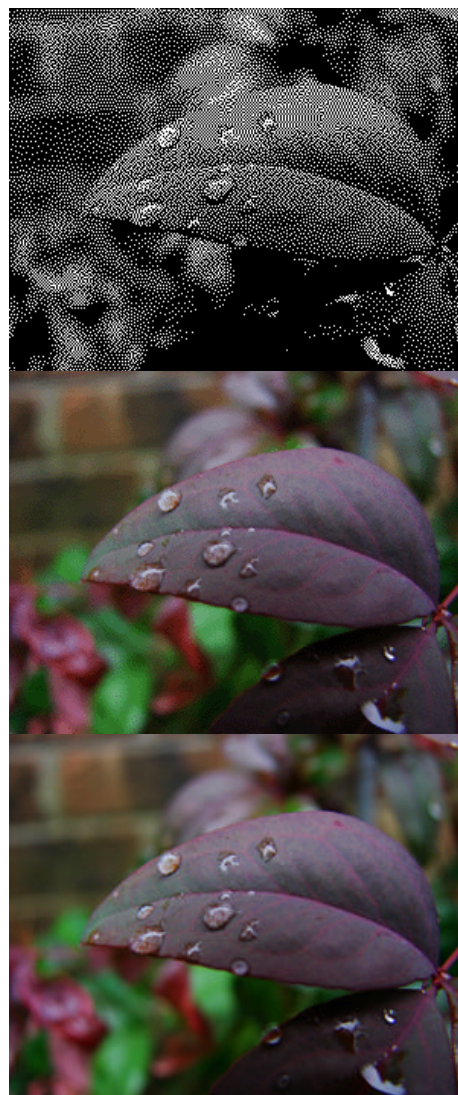


Fig. 2. Färgdjup. 1-bit, 8-bit & 24-bit (Wikipedia-Color-depth)

och färgmättnad. Färger kan därmed blandas på olika sätt men i slutändan resultera i - vad vi uppfattar - samma nyans (Rieger 2000a, s.64, Besser 2003, s.9).

Att representera färger i den digitala världen är svårt och har gett upphov till olika s.k. färgrymder, där färgerna visas och beskrivs på skilda sätt, vanligast är RGB (Red, Green, Blue) och CMYK (Cyan, Magenta, Yellow, Black). Färgrymder har olika användningsområden, RGB är lämpad för digitala medier, CMYK för tryckt material. RGB är uppbyggd av tre färgkanaler: röd, grön och blå. De tre färgerna har sin egen intensitet och blandas ihop till en enhetlig färg för varje pixel. Om en RGB-bild är i 24-bits färgdjup så tillhandahåller varje kanal 8-bit, alltså 256 färger. Standard RGB (sRGB) är den vanliga versionen för färgåtergivning inom RGB-mönstret. Den har dock fått kritik för att inte kunna återge vissa tryckta färger, speciellt inom cyan, blå och grönt. Det kan medföra problem vid utskrifter eller att skanna bild inom de färgtonerna (Rieger 2000a, s.64). Vid tryck är tekniken annorlunda då appliceras trycket på (oftast) vitt papper och att ljusreflektioner då inverkar på färgerna. Färgmodellen som då används är CMYK som använder cyan, magenta, gul och svart. När färgerna trycks läggs de enskilda färgkanalerna i lager över varandra och bildar med små punkter den nya färgen. Papperet absorberar ljusvåglängder och reflekterar resten av färginnehållet, som blir den färg ögat uppfattar (Besser 2003, s.9f.; Rieger 2000a, s.64).

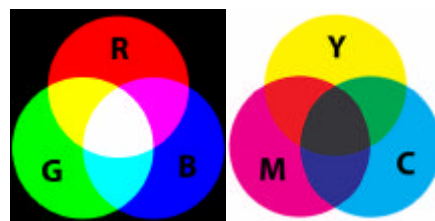


Fig. 3. Färgmodeller, RGB & CMYK (Wikipedia-RBG color model; Fotoguiden)

Filstorleken som fås i och med bildfångsten beror delvis på besluten som tas utifrån kvalitetskriterier för bilden och skanningsinställningarna. För att räkna ut vad filstorleken blir av en bild/dokument när det skannas används denna formel: (Kenney 2000, s.32).

$$\text{filstorlek (byte)} = (\text{höjd} \times 0,39) \times (\text{bredd} \times 0,39) \times \text{färgdjup} \times \text{dpi}^2 / 8$$

Genom att multiplicera med 0,39 omvandlas talet från cm till inch (Manuels Web). För att exemplifiera formeln. Ett fotografi med storleken 14x20cm skannas med färgdjup 24-bit i 600 dpi, ger ca filstorleken 46 Mb. Om samma bild skannas i 36-bit och 1000 dpi blir filstorleken ca 1,5 Gb.

För att minska bildens filstorlek är det möjligt att komprimera. Komprimering har länge inneburit försämrade kvalitet, men idealet är att bilden ska kvarhålla den. Komprimeringsalgoritmerna existerar i två varianter, förstörande (lossy) och ickeförstörande (lossless). De förstörande algoritmerna tar bort bildinformation och medför försämrade kvalitet. Algoritmen tar bort, för mänskliga ögon, svårupptäckta bildelement ur bilden och ersätter det med färgvärde som överensstämmer med pixelns omgivning. Den ickeförstörande komprimeringen bibehåller all bildinformation, och kan minska filstorleken med 40-60%. (Besser 2003, s.19f.; Lee 2001, s.43f.)

Eftersom syftet med digitaliseringen skiftar medför det att bildfångsten görs utifrån olika kriterier och skiftande fokus. Om den utförs för användning i bibliotekskatalogen och visa hur bokomslaget ser ut, räcker antagligen låg bildkvalitet, om det är för fotoanalys, ställs troligen högre kriterier för acceptabel kvalitet. Många handböcker (ex. Besser

2003, s.43) rekommenderar att bildfångsten ska generera en bild av mycket hög kvalitet och att denna bildfil får bli s.k. digital master. Utifrån mastern skapas s.k. accessbilder, bilder som anpassas för specifika användningsområden som: utskrifter, webben, mobiltelefoner m.m. Skapas bildfilen med rikare bildinformation än vad det direkta behovet av projektet kräver så kan framtida ändamål för bilden möjliggöras som från början varit ovissa. Vid skanningen anser Lee att 300 dpi är minimikravet, men att 600 dpi är idealiskt, och att det sker i 24-bit för färgbilder, 8-bit vid gråskalor. Besser rekommenderar att den digitala mastern har 36-bitars färgdjup (Besser 2003, s.44f.; Lee 2001, s.69).

Att i framtiden behöva skanna om material innebär onödiga påfrestningar och kostnader. Det råder viss diskussion om hur vida den digitala mastern över huvud taget ska få förändras. Besser anser att justeringar som medför att den digitala kopian motsvarar den analoga är acceptabelt, som att klippa bort bakgrund, ändra ljusstyrka, kontraster och färger, men påpekar samtidigt att det är bra om den redigerade varianten av mastern sparas tillsammans med ursprungliga mastern. Idealet anses dessutom vara att den digitala mastern inte komprimeras, men att ickeförstörande komprimering kan godtas. Det är inte alltid skannrar klarar av en analogs bild dynamiska omfång, alltså hur väl de mörka och ljusa tonerna representeras i bilden. För att underlätta justeringar av färgerna och det dynamiska omfånget är det vanligt att vid skanning inkludera referensverktyg som visar kända färgpunkter det mörkaste av svart och ljusaste av vitt. Vid bildredigeringen kan bearbetningen utgå från stickans färger för bland annat färgmättnad, då vet man att resten av bildens färger är korrekt representerade (Kenney 2000, s.25, 37f.; Besser 2003, s.44f.; Lee 2001, s.69).

4.2.2 Filformat

Många filformat har utvecklats för att beskriva digitalt innehåll för bilder och texter. I början hade varje utvecklare av mjukvara och hårdvara egna standardformat. En del av dessa har försvunnit och andra ersatts av nya. Vissa format är dock vanligare än andra och rekommenderas i digitaliseringssammanhang (Lee 2001, s.45). I bildfilerna inkluderas automatiskt teknisk metadata om bilden, såsom pixlar och färgdjup (Besser 2003, s.21).

TIFF (Tagged Image File Format) är det vanligaste bildfilformatet för lagring av den digitala mastern. I TIFF lagras bilden med LZW-algoritmen, vilket är en algoritm som kan komprimera i ickeförstörande och förstörande bildformat. Populariteten för TIFF gör att den stöds av de flesta bildprogram, men har också medfört att flera TIFF-varianter utvecklats varav alla inte stöds, eller läses olika av mjukvaruprogrammen (Besser 2003, s.21f.). I TIFF-formatet går det spara metadata, som är av typerna deskriptiv, administrativ och strukturell¹. TIFF stöder dessutom flera färglägen, ex. RGB och CMYK. TIFF klarar av färgpaletter upp till 64-bit, och 8-bit i gråskala (Kenney 2000, s.52).

JPEG2000 är ett relativt nytt filformat och är en vidareutveckling av JPEG. JPEG2000 använder sig av waveletkomprimering och innebär större grad av kontroll över hur komprimeringen utförs. JPEG2000 tillhandahåller två typer av filformat, JP2 och JPX.

¹ Mer information om dessa typer i avsnittet 4.3 *Metadata och uppmärkning*.

Där JPX har bättre stöd för avancerad XML metadata (Besser 2003, s.22). JPEG2000 uppvisar många fördelar framför TIFF och JPEG och tros därför bli en betydelsefull standard. Formatet kan spara både förstörande och ickeförstörande komprimering. Samma filformat kan anpassas efter målgrupp och användningsområde. Vid bearbetningen av en bild kan regioner som är av mindre betydelse komprimeras hårdare än delar av stor betydelse (Acharya 2005, s.139ff.). JPEG2000 klarar av 48-bitars färgdjup. Formatet är inte väletablerat, och förekommer sällan på internet (Rieger 2008, s.20).

JPEG (Joint Photographic Expert Group) är det vanligaste formatet för webben, mycket p.g.a. komprimeringen som gör filen liten men med fortsatt hög bildkvalité. Komprimeringen är förstörande och gör att bildinformationen minskar vid varje sparning. Det gör filen olämplig som digital master (Lee 2001, s.46). JPEG har en färgpalett på 24-bit, och 8-bit för gråskala. JPEG kan med tillägget SPIFF (Still Picture Interchange File Format) lagra metadata (Kenney 2000, s.52; Besser 2003, s.22). Det som refereras som JPEG är egentligen formatet JFIF (JPEG File Interchange Format) då JPEG är en komprimeringsmetod (Besser 2003, s.22).

GIF (Graphics Interchange Format) tillåter endast 8-bits färgdjup, alltså 256 värden, vilket har gjort den välanvänd på webben i samband med illustrationer och gråskalor. Eftersom formatets färgdjup är lågt så ger det mindre filstorlek. Formatets komprimering är med LZW-algoritmen, alltså icke-förstörande (Lee 2001, s.46).

PDF (Portable Document Format) är ett filformat som klarar av både bild och text inom samma dokument. Då text och bild samsas i ett och samma format ger det stora filer. Formatet är populärt och har stort stöd, men för att läsa dokumentet krävs det att användaren har Acrobat Reader eller Adobe Reader installerat (Lee 2001, s.48, 138f.).

4.2.3 Bildfångstutrustning

Vid bildfångsten kopieras det analoga och blir digitalt. Flera varianter existerar för hur det kan utföras. Bland utrustningen finns skillnader, där utrustningen och källdokumentet är i kontakt med varandra (ex. flatbäddskanner) och där dem inte är det (ex. kameror). Dokument som t.ex. är skrynkliga eller att boken inte helt går att vända upp utsätts för påfrestningar vid den typ av skanning (Hughes 2004, s.182f.). Andra typer av skannrar än flatbäddskanner är automatisk dokumentmatrare (Sheet-feeders) som är lämpade för lösa dokument och stora kvantiteter. Den matar och skannar materialet självgående och kräver därmed inte personal som vaktar arbetet (Lee 2001, s.54). Vid användandet av kamera monteras den på en ställning där höjden justeras för att hela dokumentet ska få plats på en bild, för det mesta oberoende av dokumentets storlek. Ljuset måste organiseras för att undvika skuggor på objektet. Att använda kamera är därmed mer krävande än med skanner, men ger också större kontroll över bildfångsten. Digitala kameror kan få problem med analoga material av hög detaljrikedom, ex. kartor (Kenney 2000, s.32; Lee 2001, s.56ff).

4.2.4 Omvandling till digital text

Filformat och skanningstekniker är irrelevant för textdokument om man ändå inte digitalt kan söka och använda texten. Det finns två sätt att omvandla analog text till digital.

Den första är transkribering och innebär att originalkällan skrivs manuellt, tecken för tecken, i ett skrivprogram. Det är tidsödande och betyder nödvändigtvis inte att texten blir korrekt återgivet. Eftersom det är manuellt kan både tryck- och stavfel inträffa (Price-Wilkin 2000, s.110).

Den andra är optical character recognition (OCR), som maskinellt omvandlar ett inskannat textdokument till digital text, vilket möjliggör maskinläsning. För att avläsa texten ifrån bilddokumentet måste kvalitén på bilden vara tillräckligt hög så att minsta tecken tydligt kan avläsas. Typsnitt och språk har stor inverkan på resultatet om OCR-skanningen blir framgångsrik (Lee 2001, s.136f.).

Ett dokument kan innehålla olika layoutmässiga detaljer och textstrukturer som kan vålla problem för OCR-avläsaren, speciellt tidningar anses svåra (Price-Wilkin 2000, s.112). Ett test gjort av Price-Wilkin visade att en upplösning på 600 dpi nästan alltid gav bättre OCR-resultat än vid 300 dpi. Komplex sidlayout och typsnitt innebar sämre resultat, och att suddig och dålig bildfångst gav sämre digital text (2000, s.114).

Skanningen av bilden har länge skett i bitonalskala, för att bokstäverna lättare framträder och håller filstorleken nere. Men allt eftersom datorernas minneskapacitet och OCR-programmen förbättras så blir färgskanning mer attraktivt. Frågetecken gällande OCR är hur tillförlitlig och korrekt den digitala texten blir. Om OCR sägs vara 98 % korrekt så betyder det att två bokstäver av 100 är fel. En typisk OCR-översättning har 98-99,5 % korrekthet. Vid analogt handskriven text överensstämmer däremot den digitala texten endast med 85-95 % (Price-Wilkin 2000, s.111). Enligt Hughes är den önskade korrektheten 99,95% (2004, s.259). OCR-programmen har svårast att hantera icke-alfabetiska bokstäver och kursiva tecken som i arabiskan (Deegan & Tanner 2004, s.495).

Det förekommer flera sätt på vilket en OCR-skannad text kan förbättras och blir mer korrekt. Texten kan automatiskt granskas tillsammans med stavningskontroller, ordlistor och tesaurier, men också förbättras med hjälp av manuella granskningar (Deegan & Tanner 2004, s.495). Programmen kan lära sig att tolka tecken med träning på nya texter och obekanta tecken. Det kräver till en början mänsklig medverkan men är enligt Deegan och Tanner värt mödan. En OCR-skanning anser de ger bäst resultat om originalet är modernt och i gott tillstånd och att bra kvalitetskontroll finns. De påpekar att projekt måste överväga fördelar och nackdelar med mänskligt arbete, då det är dyrt och tidsödande att korrigera en OCR-text, även när den är relativt korrekt (2004, s.495). Willett påpekar att en text medvetet kan innehålla ”fel”, eller att det är skrivet dialektalt och om OCR-behandlaren rättar upp sådant så överensstämmer inte den ursprungliga och digitala texten (2004, s.246).

Om slutanvändaren endast är intresserad av faksimilen så är enstaka fel av smärre betydelse, men kan fortfarande ge problem vid textsökning (Lee 2001, s.137f.; Price-Wilkin 2000, s.110). Ett annat alternativ är att använda sk. fuzzy sökning, sökmotorn använder då lingvistiska regler och kontextuella ordlistor när den söker i inkorrekta texter och kan bl.a. hitta och returnera alternativa stavningar till sökfrågan. Hybridlösningar av både OCR-korrigerings och fuzzy sökning kan även användas (Deegan & Tanner 2004, s.496).

4.3 Metadata och uppmärkning

Metadata har i sin enklaste form beskrivits som ”data om data” eller ”information om information” (Hughes 2004, s.196). Lagoze och Payette anser det vara för stor simplificering, och betonar istället att det är ”strukturerad data om data”. Alltså, att det är strukturerad information som beskriver information. De påtalar att struktur och normer är en viktig del av metadata och att metadata först då har möjlighet att göra påtaglig verkan (2000, s.84). Syftet med metadata anges vara att hitta information och hur informationen kan nås och användas, men också att dokumentera innehåll, kvalitet, egenskaper hos informationen och ”fitness for use” (Deegan & Tanner 2002, s.113).

Metadata's funktion kan likställas med katalogisering, där en katalogpost komprimerat beskriver en bok. Väldefinierad metadata besparar många mödor och ger välformulerad data som kan ligga till grund för långtidsbevarande. Deegan och Tanner skriver att skapa av digitala objekt måste ta hand om metadata lika mycket som själva datan. De påtalar också existensen av flera nivåer av metadata, från att beskriva en hel samling med samma metadata, till en bok, till enskilda sidor (2002, s.114f.). Vilken nivå ett digitaliseringsarbete väljer att lägga sig på styrs av bl.a. lokala regler och användarbehov. Dokumentregistrering utan lämplig metadata blir snabbt ohållbar och försvårar återfinnande och användning. Besser anser metadata av kvalitet är bättre än ett välutvecklat schema som är dålig ifyllt, då det innebär svårigheter av använda (2003, s.6).

Skapandet av metadata lyfter enligt Lee fram frågan: ”hur katalogiserar man digitala bilder/texter?”. Lee ser två lösningar på frågan (som kan kombineras). Den ena lösningen (och den vanligaste) påminner om vanlig standardkatalogisering där viktiga element registreras i enskilda fält som upphovsman, utgivningsår, skanningsinformation, innehåll, m.m. Systemet är strukturerat och mängden information det innehåller varierar utifrån behovet. Det fungerar i både analoga och digitala världar och är välbekant för användare och katalogisatörer. Samtidigt öppnar det för subjektivitet, och katalogiseringsarbetet är tidsödande (2001, s.103f.) För att underlätta arbetet, främst med indexering och sökning, har det forskats på automatisering och hur man maskinellt kan få fram information ur bilder, som färger, former och mönster, för att i slutändan få datorn att ”förstå” vad en bild innehåller. Forskningen benämns content-based image retrieval (CBIR) men är än så länge i sin linda. Systemets fördel ligger i snabbheten och att bilderna alltid avläses och uppfattas likadant, men systemet har svårt att veta vad som egentligen står i texten eller vad bilden innehåller. CBIR används därför i väldigt liten omfattning i digitaliseringsprojekt (Lee 2001, s.104f.).

Informationen i metadata kan vara av vitt skilda slag, men delas vanligtvis in i tre övergripande kategorier: deskriptiv-, strukturell- och administrativ/teknisk metadata (Besser 2003, s.7; Deegan & Tanner 2002, s.116). Det finns fler metadata-kategorier, som utgår ifrån skilda synsätt, exempelvis nämns ”bevarande metadata”, vars innebörd är att information bibehålls i ett långtidsperspektiv. Perspektivet innefattar följande tre kategorierna, men också rättighetsinformation (Rieger 2008, s.20).

Deskriptiv metadata är beskrivande information om objektet, ex. titel, upphovsman, ämne, datum, ämnesord, abstracts m.m. Det är information som beskriver texter, bilder, ljud m.m., attribut som ofta förekommer i ordinär katalogpost (Deegan & Tanner 2002, s.116f.). Med den informationen kan användaren söka efter särskilda dokument och förstå vad det beskriver (Besser 2003, s.7). Ämnesord för bilder ger uppkomst till fler

problem än vid texter då bedömning av bilder anses mer subjektiv, där tolkning av innehållet kan skifta avsevärt (Deegan & Tanner 2002, s.116f.).

Strukturell metadata beskriver strukturen och relationen inom digitala objekt - definierar digitalt hur analoga dokument är uppbyggda. Ett analogt objekt har sin struktur (för det mesta) genom att den hålls samman av en bokrygg och det visuella. I en digital värld existerar dokumenten som fristående objekt. Med strukturell metadata kan dessa objekt bindas ihop. Ett komplicerat dokument som tidningar innehåller flera typer av objekt på en sida, ex. artiklar, bilder och reklam. För att upprätthålla strukturen mellan objekten används strukturell metadata. Det kan också användas för sektions- och kapitelindelningar (Deegan & Tanner 2002, s.117f.). I större digitaliseringsprojekt är det inte alltid möjligt eller kostnadseffektivt att registrera detaljerad strukturell metadata av detta slag (Rieger 2008, s.22). Strukturell metadata kan ses i relation till OCR (vilket beskrivs i avsnitt 4.2.4 *Omvandling till digital text*), då strukturell metadata kan uppfattas vid OCR-skanningar och registrera komplexiteten och känna igen strukturen hos sidan för att där efter återskapa det digitalt. Något som har genomförts i exempelvis projektet The Metadata Engine Project (METAe).

Administrativ/teknisk metadata är information om digitaliseringen och det digitala objektet. Det kan bl.a. beskriva: skanningen, filformat, komprimering, använd utrustning, skanningsdatum. Vars information automatisk kan generas av hård- och mjukvara, dock inte allt, som exempelvis bildproduktion. Metadata innefattar också copyrighträttigheter, betalning, autentisering och varför objektet digitaliserades. All information som möjliggör enklare bevaring av det digitala objektet för framtiden är av intresse och relevans i administrativ metadata (Deegan & Tanner 2002, s.120f.; Besser 2003, s.7; Rieger 2008, s.21). Metadata bör berätta för läsaren när, var och av vem verket skapades. Dokumentationen ska göras så att en utomstående lätt förstår vad och hur digitaliseringen utförts, och enkelt fortsätta arbetet (Anderson 2006, s.76).

Metadata ska idealiskt skapas och uppdateras automatiskt när något sker med filen, att den flyttas eller på annat sätt förändras, men detta ses som mer otänkbart för digitala poster, då det anses svårt och omfattande (Shape 2007).

För att få struktur på metadata används s.k. scheman, vilket också definierar innehållsmängden. Ett metadata-schema måste klargöra hur mycket information som ska registreras, då det kan vara närmast oändligt, speciellt när metadata används av många yrkesgrupper där deras behov och förväntningar skiljer sig åt (Lee 2001, s.107f.). Strukturen av schemat gör det inte bara enklare att registrera uppgifter utan förenklar även för datorerna att läsa materialet. Det vanligaste schemat inom bibliotekskatalogisering är Machine-Readable Cataloging (MARC) med Anglo-American Cataloguing Rules 2 (AACR2) som definierar hur schemat fylls i. Vad som är möjligt att genomföra och beskriva med MARC har inom digitala medier visat sig bristfälligt, då de främst är avsedda för deskriptiv metadata. Lagoze och Payetto ser automatiska rättigheter och kontroll av åtkomsten som de största svårigheterna, men nämner även systemadministration och bevarande. MARC anser de vara oflexibelt, där ett (1) schema anpassas för alla dokumenttyper (2000, s.85).

Metadata kan fördjupas längre än vad det tidigare avsnittet avslöjade. I stället för att nöja sig med att i en digital text markera vad som är kapitel, rubriker, m.m., är det möjligt att gå in på enskilda ord, meningar, uttryck och beskriva vad det är, vad det innehål-

ler, ex. om det är ett citat, ort, versrad, tal, eller vem som sagt vad. Detta kallas för textuppmärkning (eng. text encoding). När meningar och ord i texten märks upp så beskrivs innehållet, men inget av det syns vid ex. webbpublicering. Fördelen med att märka upp ex. citat i en text är att man sedan kan bestämma att alla citat ska vara kursiverade. Textuppmärkning kan med andra ord styra hur innehållet framställs vid presentationen. Uppmärkning kan också användas för att automatisk skapa innehållsförteckning, förenkla för sökmotorer, då de får möjlighet att "förstå" vad texten faktiskt innehåller. Fler fördelar finns att läsa hos Renear (2004, s.222). Att använda detta förhållningssätt till texten kräver intellektuellt tankearbete och funderingar på vad texten egentligen innehåller (Renear 2004, s.224f.).

För att kunna märka upp en text används scheman som är flexibla men enkla att strukturera. Vanligast och närmast att betrakta som standard är eXtensible Markup Language (XML). XML är plattformsoberoende. Då texten definieras av olika uppmärknings möjliggör det att maskiner kan avläsa innehållet och strukturera det (Ray 2003, s.2; Lee 2001, s.139). Vad och hur en text kan märkas upp i XML styrs av dokumenttypsdefinition (DTD). Det finns flera olika DTD, anpassade för uppmärkning av olika texttyper. Text Encoding Initiative (TEI) är en DTD vars inriktning är texter från böcker, poem, drama m.m. och är internationell standard för bibliotek, museum och akademiska institutioner för vetenskapliga texter (Ray 2003, s.288). TEI definierar inte vad som ska uppmärkas i texten utan ger möjlighet/redskap att märka upp texten. Vilka objekt är upp till textkodaren att bedöma (Renear 2004, s.234f.). Encoded Archival Description (EAD) är en DTD vars fokus är arkivmaterial, men är inte speciellt utvecklad med tanke på digitalt material, men anses fungera väl med både analogt och digitalt material (Deegan & Tanner 2002, s.127f.). Ytterligare en DTD är den som utvecklats av Library of Congress och Digital Library Federation, nämligen Metadata Encoding & Transmission Standard (METS). Schemat är skapat för att hantera metadata för bevarande, visning av digitala objekt. Schemat ska klara av avancerad digital information och består av fem komponenter: deskriptiv metadata, administrativ metadata, filgrupper, strukturell karta och beteende metadata (Taylor 2004, s.95f.). DTD kan var på väg att överges till förmån för XML Schema (XSD eller WXS). Det stöds och rekommenderas av World Wide Web Consortium, vilka definierar standarder i internetsammanhang (Wikipedia – *XML Schema (W3C)* [2009-03-05]).

Project Gutenberg är ett ideellt arbete där analoga texter konverteras till digital och görs nåbara som e-text. Grundaren av projektet, Michael Hart, är kritisk till uppmärkning av text p.g.a. att det är tidskrävande och han tror att texten klarar hårdvaru- och mjukvarubyten bättre när den är minimalt uppmärkt. Andra ser möjligheter med uppmärkning då det sker med internationella standarder och underlättar långtidsbevarande (Hart 1992; Willett 2004, s.244f.).

4.4 Kvalitetsbedömning

Uppsatsens tidigare kapitel har tagit upp väsentliga avsnitt för digitaliseringsprocessen. Texterna visar på betydelsen av kvalitet och hur detta uppnås för vissa digitaliseringsmoment. I detta kapitel beskrivs kvalitetsbedömningen noggrannare. Enligt Besser bör varje moment i digitaliseringen genomgå kvalitetsutvärdering, såväl utrustning som digitala filer. En bedömning som bör ske fortlöpande i digitaliseringen (2003, s.52).

Kvalitetskontroll anser Rieger inkluderar proceduren och tekniken som verifierar kvalitet, trovärdighet och konsistens av den digitala produkten: bilden, OCR och övrigt metadata. Bilden ska särskilt kontrolleras utifrån upplösning, färger, toner och sammanfattande bilduppfattning (2008, s.22).

För att undersöka kvalitén hos arbetet fastställs ett program – standard, för vad institutionen uppfattar som tillräcklig kvalitet och hur mycket arbetet får avvika ifrån den normen. Att fastställa programmet är mödosamt och tar mycket tid och resurser i form av specialkompetens. Bedömningar måste även göras av eventuella effekter det innebär på hela arbetet om material skulle behöva skannas om. Rieger konstaterar dock att ett grundläggande och enkelt kvalitetsprogram dämpar effekten av dåligt material (2000a, s.61). Om kvalitetskontroll inte görs kan materialet inte garantera konsistens och integritet av den digitala filen (Besser 2003, s.52).

Vad som kontrolleras skiljer sig utifrån målet och syftet med digitaliseringen. Om arbete görs för att troget återge en analog version så krävs bättre dokumentkvalité än om arbetet främst syftar till att visa utdrag från böcker. Vid bedömning måste det fastslås vad som jämförs med vad, vilket inte är lätt alla gånger. Mikrofilmer har t.ex. redan genomgått en skanning, ska den digitala filen jämföras med mikrofilmen eller hur ett eventuellt original såg ut? Och om filmnegativ skannas, hur vet man att det återges korrekt? Det är frågor som är svåra att besvara. Konverteringen sker därför med vetskapen om begränsningar av användningen, teknologin och kunskapen (Rieger 2000a, s.63).

Kvalitetskontroll kan i större arbeten inte utföras manuellt och arbetet måste delvis genomföras med automatiska verktyg. Men då ett vältränat öga anses bedöma bättre än maskiner förespråkas det att manuella kontroller genomförs frekvent. Hur ofta det kontrolleras preciseras i kontrollprogram. Även om automatiska verktyg används för kvalitetskontroll av ex. namngivning och integritet så har det svårare att upptäcka saknade sidor och bildförvrängning (Besser 2003, s.52; Rieger 2000a, s.69; Rieger 2008, s.22).

Oavsett om kontrollen sker manuellt eller automatiskt så måste utrusningen som används kontrolleras så att digitaliseringen och de digitala dokumenten förblir konsekventa. De största problemen är att system och applikationer läser och översätter den binära koden olika vilket försvårar avgörandet av inställningar. Det ska dock stämma överens bland institutionens egen utrusning så att digitaliseringen på så vis genomförs enhetligt (Rieger 2000a, s.69). Hårdvaran är inte konstant utan påverkas av användningen och hur länge den har varit i bruk. Digitala filer jämförs med analoga versioner och monitoren har då en direkt påverkan på om skanningen kommer att anses lyckad. Skärmbedömningen påverkas av hårdvarukonfigurering, bildvisningsmjukvara, skärminställning, färgkontrolleringsinstrument, färghantering, betraktningförutsättningar och hur människor uppfattar ex. färger (Rieger 2000a, s.66f.).

Skanningen kan ge uppkomst till optiska bildförvrängningar. Brus är ett exempel på det som uppkommer vid ojämn bildfångst och karaktäriseras av kornigt mönster eller spräcklighet och är vanligast i mörka områden. Andra mönster som skanningen kan medföra är *morié* (vågigt mönster), *banding* (varierande ljus- och mörkhet) och *newton rings* (cirkulära mönster) (Rieger 2000a, s.71).

Bildens egenskaper och kvalitet bestäms av upplösningen. Ett textdokument granskas utifrån distinkta kanter, läsbarhet, tonåtergivning, högdager, mörkhet, skarphet, skevhet,

kontrast m.m. Vid granskning av bilder fokuseras även färgdjup och dynamiskt omfång. Jämförelsen mellan den analoga och digitala bilden och de nämnda egenskaperna försvåras av det fysiska materialet vilket inverkar på den mänskliga uppfattningen av dokumentet (Rieger 2000a, s.73).

För att bedöma om en bild återger egenskaperna från analog till digital bild korrekt används flera verktyg. De är aldrig en garanti för att det blir korrekt, men hjälper till vid bedömning för både maskiner och människor. För att jämföra RGB i olika värdenivåer används färg- och gråskalestickor – referensverktyg. Ett histogram används för att utvärdera *signal-to-noise ratio*, tonbeskrivning och spatial upplösning. Vanliga färgkorri-geringsverktyg är: färgkalibrerare, densiometer, colrimeter, spectrophotometer (Rieger 2000a, s.68).

4.5 Lagring av digitala original och kopior

Lagringen av de digitaliserade bilderna ligger i hjärtat av digitalisering och möjliggör dess bevarande. Som tidigare nämnts har digitalt bevarande uppfattats på olika sätt: bevarande ska ske för lång tid framöver, andra ser det som en evig bevaring medan ytterligare en åsikt är att dokumenten ska bevaras så länge de innehar värde (Rieger 2000b, s.135).

Bevarande och lagring syftar till att säkerhetsställa att innehållet skyddas från odokumenterade förändringar och fokuserar därmed på autencitet, integritet och att data inte blir korrupt. Förändringar av data kan enligt Graham uppkomma på tre sätt: oavsiktlig förändring, avsiktlig förändring (med gott syfte) och avsiktlig förändring (med dåligt syfte). Oavsiktliga förändringar kan förorsakas av olyckor vid ex. förflyttning av data eller att uppdateringar missar vissa sektioner av ett dokument. Avsiktliga förändringar (med gott syfte) har två varianter: nya versioner och revideringar har tillkommit eller strukturella uppdateringar som förändrar informationsinnehållet. Avsiktliga förändringar (med dåligt syfte) är ex. att förändra och dölja egna dokumentförändringar eller för att manipulera innehållet (Graham 1998, s.84).

För att upptäcka om data har förändrats kan kontrollsummor (eng. checksum) användas. Avstämning kan göras efter lagringsskifte (beskrivs nedan) för att se om förändringar förekommit. En enkel variant räknar antalet bits i filen före och efter. Avancerade kontrollsummor kan peka ut var förändringen har skett. För att öka nivån av integritet används krypteringsfunktioner (Besser 2003, s.69; Wikipedia - *checksum* [2009-03-05]).

Det finns flera sätt att lagra data på, men lagringsmedierna har olika försättningar betraktande bevarandekrav och fysisk kondition. De varierar också i fråga om lagringskapacitet och åtkomstbehov (Jones & Beagrie 2001, s.128). Ett magnetiskt lagringsmedium är t.ex. magnetiska band och hårddiskar. De använder magnetiska egenskaper i metallmaterial. Det är en allsidig och billig lagringsmetod som har förbättrats de senaste åren. Lagringsutrustningen ska inte utsättas för magnetisk störning, då det leder till dataförlust. Lagringen är kompakt och små störningar kan leda till förlust av data. Jones och Beagrie rekommenderar att lagring på magnetiskt band sker i två versioner, från två olika utvecklare. De menar ytterligare att utrustningen som används för åtkomst till magnetiska banden noga kontrolleras då det är en av de vanligaste orsakerna till att magnetiska band skadas (2001, s.128f.).

Optiska lagringsmedium är t.ex. CD-ROM och DVD-ROM. De använder laserljus för att avläsa datalagret, vilket består av gropar och platåer i metall, som utanpå skyddas av plast. En DVD-skiva kan lagra 4.7-18Gb. Optiska diskar är populära som lagringsmetod. Läsaren och skiva vidrör inte varandra och mekaniska fel hos utrustningen drabbar därför sällan lagringsobjektet (Jones & Beagrie 2001, s.129f.).

Optomagnetiska (eller magnetoptiska) lagringsmedium är ett mellanting av magnetisk och optisk lagring. Precis som med optiska medier skrivs den och läses med laserteknik, men bygger på det magnetiska. Det gör den okänslig för magnetisk störning och har lång livslängd. Dess lagringskapacitet anses mycket bra, men är samtidigt något långsammare än vanliga hårddiskar (Kungl. biblioteket 2008b; Prismas IT-ordbok).

Då digitala lagringsmedium fortfarande anses instabila förekommer att digitala dokument skrivs ut. På så sätt finns de även lagrade i analog form (Jones & Beagrie 2001, s.103). Digitala kopior kan förvaras på flera geografiska platser. Om en förstörs så existerar andra kopior att hämta tillbaka informationen ifrån, och mindre risk för att digitala versioner försvinner och digitaliseringsprocessen måste återupprepas. Ett sådant arbete utförs av LOCKSS (Lots Of Copies Keep Stuff Safe), som samlar in material från 40 bibliotek och 30 publicerare. I Norden är endast Lunds Universitet delaktiga. LOCKSS är dock inget arkiv utan ett redskap för att garantera åtkomst till material, främst tidskrifter (LOCKSS; Deegan & Tanner 2002, s.206).

Det finns flera sätt att hantera lagrad data då det med jämna mellanrum behöver förflyttas till ett nytt medium. Detta för att teknik slits ut och blir omodern. I digitaliseringsarbetet måste det därför accepteras att ingen optimal lagringsteknik finns. Optimal på så sätt att ingen informationsförflyttning är nödvändig. Lagringen ska därför göras så att det enkelt kan överföras till bättre och nyare teknik (Erway 1998, s.165f.). Hughes uppskattade 2004 att materialet måste flytta till nytt lagringsmedium ca. var 3 år (s. 11). För att data även i framtiden ska vara åtkomlig finns det tre vanliga tekniker och synsätt på hur uppgradering till moderna lagringenheter kan utföras (Jones & Beagrie 2001, s.130.).

Refreshing: Med refreshing menas att innehållet kopieras från ett lagringsmedium till ett nyare. Varje bit av filen är därmed identisk med varandra. Rieger betonar att det inte ska ses som säker långtidsbevaring utan att det är ett arbete som utförs i förebyggande syfte då medier föråldras. De ser refreshing som en integrerad del av ett fortlöpande upprätthållande av digitala samlingar (2000b, s.146). Refreshing möjliggör bättre tillvaratagande av filerna då det lagras på utrymmen som är stabilare och mer effektiva. Tekniken är rak med låg risk för förstörda filer eller andra skador (Deegan & Tanner 2002, s.196).

Migration: Migration är den populäraste och mest använda tekniken. Datainformation förflyttas från en hård- och mjukvara till en annan, eller från en datorgeneration till en senare. Med denna metod lagras den digitala information och dess intellektuella värde på ett sätt och i ett medium som följer teknikutvecklingen (Jones & Beagrie 2001, s.103f; Rieger 2000b, s.146). Detta möjliggör att data som är svår att läsa kan räddas. Migration kräver noga övervägande, planering och analys p.g.a. komplicerade förflyttningar och val av mjukvara. Informationsförlust kan bli oundviklig då gamla och nya lagringar inte alltid är kompatibla. Det kan också leda till korrupt data (Jones & Beagrie

2001, s.103; de Stefano 2000, s.312). Vid migrering ser de Stefano med viss oro på hur få utredningar som finns om hur det kan genomföras, vad de kostar, och att det inte finns någon kvalitetskontroll för metoden (2000, s.321).

Emulering: Emulering är en metod där det eftersträvas att likna och i viss mån återskapa den ursprungliga tekniska miljön och utseende för den digitala samlingen, både gällande hårdvara och mjukvara. Originaldata lämnas orörd så långt det är möjligt. För att det ska bli möjligt bör samma teknikutrustning användas som datan ursprungligen utvecklades för. Där det inte är genomförbart kan information och kunskap om miljön hjälpa till att återuppbygga miljön (Rieger 2000b, s.149). Metoden har mötts med viss skepsis men kan vara bra i ett riktigt långt bevarandeperspektiv (Deegan & Tanner 2002, s.200). Kostnaderna och att det krävs mycket kunskap sätter dock ett frågetecken för hur praktisk genomförbar den är (Jones & Beagrie 2001, s.105).

Emulering snuddar vid fenomenet teknisk bevarande där fokus ligger på tekniken och där data ska köras ifrån den ursprungliga utrustningen och inte bara försöka återskapas som vid emulering. Tekniskt bevarande ses dock inte som ett realistisk långtidsbevarande alternativ. Att spara och vårda all skapad teknisk utrustning för att kunna läsa originalfiler skulle kräva omfattande kunskaper och fysisk plats (Jones & Beagrie 2001, s.106).

Dessa tre är de vanligast förekommande och en blandning av dem är möjlig och trolig inom ett digitaliseringsarbete. Det har dessutom tidigare nämnts i uppsatsen att tekniken än så länge inte har möjlighet att erbjuda den stabilitet som krävs för att lagra under lång tid på samma lagringsenhet. Hybridlösningar förekommer med att lagra digitala dokument på mer traditionsbundna och säkrare sätt, ex. mikrofilm eller som utskrifter (de Stefano 2000, s.312f.).

4.6 Framvisning

Framvisning handlar om hur de digitala bilderna visas för användaren, vilka valmöjligheter användaren har och hur mycket information den får ta del av. Framvisning är därmed ett viktigt medel för hur institutioner väljer att tillgängliggöra sitt digitala material. Ett starkt band finns också mellan bevarande och tillgängliggörande, då man bevarar för att det i framtiden ska fortsätta att vara nåbart (Jones & Beagrie 2001, s.122; Besser 2001, s.54).

För att få framvisningen framgångsrik måste den enligt Besser baseras på: användarkrav, gränssnittdesign, konsistens och kvalitet av metadata, val av bildhantering eller presentationsmjukvara och teknisk infrastruktur (2003, s.54).

4.6.1 Bildhanteringssystem

Ett bildhanteringssystem kan bland annat användas för att registrera lagringsposition för masterbilder och accessbilder, söka och återvinna bilder, ge en kontextuell miljö för den digitala bilden, spåra och kontrollera källmaterial och kontrollera användning. Hur en institution väljer att lösa framvisningen bestäms av användningsmönster. Vid upprättande av bildhanteringssystem styrs valen av samlingen som ska digitaliseras och

dess syfte, storlek, komplexitet och ombytlighet, funktioner som autentisering och granskningskrav, förväntad prestanda, tillgänglig teknisk infrastruktur och kostnader. Hitle påpekar också att ett bildhanteringssystem endast blir så bra som metadatan är (2000, s.119).

Strukturen som används inverkar på hur användaren når fram till dokumenten. Witten och Bainbridge beskriver två strukturvarianter: hierarkiskt strukturerade dokument och plattstrukturerade. Den strukturerade varianten tar nytta av den redan inbyggda strukturen som finns i böcker: kapitel, underrubriker, m.m. Användaren presenteras för bokens delar och kan med exempelvis en länk komma direkt till kapitel fyra. Om denna struktur inte fanns tvingas användaren titta i den digitala bokens innehållsförteckning och därefter förflytta sig till kapitel fyra. För att gå vidare till kapitel sju måste användaren återvända till innehållet, notera sidnummer och sedan klicka vidare. Den strukturen kallas plattstrukturerade dokument (2003, s.81ff.).

4.6.2 Hur och vad användaren får ta del av i framvisningen

I ett digitalt bibliotek kan användaren ta del av dokument på två sätt. Antingen genom att använda sökmotorer eller att browsa. Sökmotorer används när användaren vet vad den vill åt, browsa när sökandet sker mer planlöst och det är oklart vad man är ute efter (Witten & Bainbridge 2003, s.112). Vid användning av sökmotorer sker letandet i hela samlingen, både metadata och OCR-texten. Det är en stor fördel jämfört med ordinära bibliotekssystem som vanligen använder endast metadata. Sökning görs med söktermer och returnerar förekomsten av dessa termer i dokumenten, och var de påträffats. Vilka sökmöjligheter som ges i sökmotorn kan variera från mycket enkla där inga specifikationer kan göras till mer avancerade. Hur sökfrågan formuleras särskiljs mellan två typer, boolesk och rankad. I boolesk sökning kombineras termerna med hjälp av AND, OR och NOT. Resultatet av sökningen är de dokumentet som uppfyller kriteriet i sökfrågan. Rankad sökning behandlar söktermerna som ett eget litet dokument och söker efter texter som är lika, rankade gradvis. Operationen kan även implementeras oavsett sökfrågetyp. Sökformuleringar kan konstrueras så att man t.ex. söker med hela fraser - frassökning - eller att trunkeringstecken finns så att den söker på resterande stavningar, s.k. stemming (Witten & Bainbridge 2003, s.99ff.).

Browsa är att söka/leta efter något på ett ostrukturerat sätt, att låta blicken snabbt glida över böckerna tills den finner något av intresse. För att browsing ska vara möjligt krävs metadata, ty utifrån dess innehåll kan struktur för planlöst sökande skapas. Enklaste exemplet är en ordnad alfabetisk lista utifrån titel eller författare. En annan möjlighet är att utgå ifrån klassifikationskoden och få titlarna ämnesfördelade. Witten och Bainbridge nämner också så kallad ”frabrowsing” vars innebörd är att ämneskategorisering görs automatiskt av termer som förekommer i fulltextdokumenten och sedan struktureras för användaren (2003, s.112ff.).

Det förekommer flera sätt där användaren får ta del av digitaliserat material och ingen lösning kan betraktas som mer korrekt än andra. Istället förekommer det bara hypoteser för lämplig framvisning. Lösningar bör dock vara flexibla utifrån användaren och kunna erbjuda bilder i flera format, standarder och kvalitéer. Med flexibel menas också att det ska vara enkelt att inkorporera framtida behov i den befintliga framvisningen, då tekniska kunskaper och krav hos användaren tros öka (Price-Wilkin 2000, s.101f.). Om fram-

visningen inte är tillräckligt bra eller otymplig så finns risk att användaren struntar i tjänsten och använder andras tjänster (Deegan & Tanner 2002, s.165).

Det finns flera varianter av framvisning och hur digitalt material delges och anpassas efter behov hos institutionen och användare. Enklast är att publicera den digitala bilden, utan att ta hänsyn till om den innehåller text eller inte, men då blir inte texten sökbar. En annan variant är att leverera hela texten, fullt sökbar med kontrollerad transkribering. Formatet kan då användas för t.ex. textanalyser eller kopiering. En annan variant visar sidbilden av texten, med en fulltextversion dold i bakgrunden. Texten används då enbart för sökning. När en användare söker förs den med hjälp av en länk till sidbilden av den aktuella texten. Vidare finns varianten där både sidbilder och fulltext visas. En sista variant är när bild och fulltext visas och att texten dessutom är uppmärkt, med ex. TEI. Användaren har då en text rik på valmöjligheter (Hughes 2004, s.259ff.). Oavsett vilken variant som väljs så är enkelhet och enkel navigering viktiga beståndsdelar (Willett 2004, s.251f.).

Oftast får användaren inte ta del av all dokumentinformation genom framvisning. Delvis för att det inte tros intressera, men också för att institutionen inte vill visa all information, om t.ex. skanningsförfaranden eller bilden med bäst kvalitet (digitala mastern). Av rättighetskäl förekommer också begränsningar i vad som är tillåtet att visa över internet eller att bara vissa användare delges förtroende till materialet (Lee 2001, s.130f.).

Eftersom man inte helt fritt vill delge sin digitala master används istället accessbilder på olika nivåer för att tillfredställa användaren. Om masterbilden skapades i bevaringssyfte kan den vara mycket stor och i hög kvalitet och institutionen vill då inte att den otillåtet ska kopieras. De vill i stället skydda den och se till att användaren kan vara säker på att det är en ursprunglig digital kopia av originalet, att den inte har manipulerats. Det är viktigt att användaren upplever informationen trovärdig, då mycket av dess användning och förtroende bygger på detta. Accessbilderna är dessutom behändigare för användare då det inte kräver lika mycket av deras utrusning. I visningsläge av bilderna kan t.ex. zoomning användas för att kunna studera detaljer i bilden, eller ett par bildvarianter med skild detaljrikedom (Price-Wilkin 2000, s.103f.; Deegan & Tanner 2002, s.166; Lee 2001, s.132).

För att institutionen ska veta hur och av vem materialet används kan exempelvis registrering och lösenord förekomma. Åtkomsten till bättre accessbilder kan då lättare ordnas så att det t.ex. bara ges till utvalda (Hirtle 2000, s.121). Anderson uppfattar det däremot etiskt problematiskt, då alla inte fritt kan ta del av informationen (2006, s.95f.).

Digitala vattenstämplar används för att i viss mån skydda digitala bilder, även om de kan kopieras så kan också bildens ursprung spåras. Om den digitala stämpeln är synlig i accessbilden tros den dock skyddas något bättre ifrån kopiering (Lee 2001, s.143ff.).

4.7 Sammanfattning av digitaliseringsprocessen

Ett digitaliseringsarbete startar med att utröna varför det görs, vilket syftet är, varför den digitala reproduktionen görs. Vanligt förekommande orsaker är att bevara material eller att kunna göra det lättillgängligt för flera. Det kan också ha mer strategiska syften som att belysa olika dokument som är i bibliotekets ägo eller att profilera sig med digitaliserade kopior av speciella dokument. Syftet överskuggar därefter alla delar inom digitaliseringsprocessen. Vid urvalet blir det dock extra tydligt vilket syftet var. Ett urval kan exempelvis fokusera på dokumentets innehållsliga, historiska, ekonomiska eller intellektuella värde.

Bild- och textfångsten är moment där analogt material kopieras och överförs till digitalt format. Hur kvalitén för den digitala kopian blir avgörs av bildfångstutrustningens inställningar. Bilden kan aldrig göras bättre än ursprungsmaterialet. Beroende på originaldokumentets material och dess fysiska standard väljs vilken bildfångstmetod som ska användas. Flatbäddskannrar och digitala kameror är vanligast. Inställningar som kan göras av utrustningen är till exempel: bildupplösning, färgdjup och dynamiskt omfång. Det påverkar i sin tur filens kommande storlek och hur mycket plats den då tar vid lagring och distribuering. Vad man väljer beror på kommande användning. Bilder i bevarandekvalité är t.ex. mycket detaljerade. Den skannade bilden kallas därefter för digital master och förblir vanligtvis oförändrad. Utifrån masten skapas accessbilder, som har specifikt användningsområde, ex. webben och utskrifter. Dessa kan bearbetas och anpassas för att bättre passa in i användningen. Det finns flera filformat men TIFF är absolut vanligast för masterbilder. Filformatet för accessbilderna varierar beroende på hur det framvisas. När textdokument skannas måste ytterligare moment göras – optical character recognition (OCR), en metod för att överföra texten i bilden till digital text. Texten kan därmed användas för bland annat webbsök.

Metadata beskrivs som strukturerade data om data. Metadata är tänkt att underlätta informationshantering då den kortfattat beskriver ett dokumentets innehåll. Den hjälper till med att hitta och är därför mycket betydelsefull i ett längre bevarandeperspektiv. Hur mycket och vilken information som registreras varierar, alltifrån enskild sida till hel bok eller samling. Metadata delas vanligtvis in i tre kategorier, deskriptiv – som beskriver objektet, strukturell – som sätter in objektet i sitt sammanhang, ex. omkringliggande sidor, administrativ/teknisk – information som berättar hur digitaliseringen utförts och hanterats. När enskilda texter beskrivs extra noga benämns det textuppmärkning. Markeringar kan i en text utgöras av exempelvis vad som är ett platsnamn eller text som i originalet står i kursiv stil. Texten blir innehållsrik och mer anpassningsbar till skilda aktiviteter.

I ett digitaliseringsarbete bör kvalitetskontroll av teknisk- och mjukvaruutrustning utföras. Ett dagligt användande sliter på utrustningen och fel uppkommer. Det är också viktigt att skannade dokument granskas så att de uppfyller kraven som ställs. Vad och hur den digitala kopian ska jämföras med är inte alltid lätt av att avgöra. Vanligen sker det mot det skannade verket. Fysikaliska egenheter hos dokumentmaterialet inverkar på betraktarens öga och kan i jämförandet med en digital representation medföra svårigheter. Till hjälp använts ett flertal verktyg, men som många gånger överträffas av ett mänskligt vältränat öga.

Lagring handlar om att bevara de digitala dokumenten, bildfiler och metadata, för att det även i framtiden ska vara åtkomligt. I lagringen innefattas även att materialet ska skyddas från förändringar, speciellt oavsiktliga. När förändringar görs bland filerna bör det dokumenteras så att en historik av dessa skapas. Att lagra digitalt innehåll kräver ett övervägande av vilka fysiska lagringseenheter som ska användas. Främst finns det två varianter: magnetisk och optisk lagring. Dessutom förekommer en kombination av dessa - optomagnetisk lagring. Ingen teknisk infrastruktur är än så länge evigt bestående utan måste förnyas med jämna mellanrum. Refreshing, migration och emulering är tre metoder som försöker överbrygga detta.

Framvisning fokuserar på hur digitalt material tillgängliggörs för användaren. Hur en institution väljer att gå tillväga baseras på många aspekter som utgår ifrån användaren och materialet. Det som håller samman visningen är metadata, vars utformning således påverkar visningen. Strukturen på dokumenten kan organiseras hierarkiskt eller platt, där man i det första tillvaratar ett dokumentets struktur i form av kapitel m.m. och underlättar för användaren att hitta. Användaren har vanligen bara två möjligheter att nå digitalt material, via sökmotor eller att browsa. Den senare när sökaren inte vet vad den vill åt. Anpassning av de digitala filerna görs för att förenkla, som att exempelvis ha flera bildnivåer eller möjlighet att zooma i materialet. Systemstrukturen bör också vara flexibel för nya implementeringar. Vid visning av materialet kan man dölja OCR-texten och endast använda den som sökkälla och låta all läsning sker utav bilden, eller också kan både OCR-text och bild visas.

5. Resultat

Följande kapitel redovisar resultaten som framkommit i undersökningar av dokument och e-postkontakt rörande nationalbibliotekens digitaliseringsverksamhet. Styckeindelningen följer den som användes i kapitel 4, fördelat på varje institution.

5.1 Nasjonalbiblioteket i Norge

Den 29 mars 2006 markerade den norske kulturministern Trond Giske starten för Nasjonalbibliotekets (NB) digitaliseringssatsning – att digitalisera allt material, och därefter tillgängliggöra det på webbplatsen NBdigital. NB blev därmed första nationalbibliotek i Europa att ha som mål att digitalisera alla dokument – oavsett medietyp - i bibliotekets samlingar. Projektet innefattar konvertering från analogt material men även att samla in ”född digitalt” material som ligger på domänen .no (Skarstein 2006). NB har även slutit avtal med bokförlag om att bevara elektroniska förlagan inför boktryck, vilket innebär att de böckerna inte behöver genomgå en digitaliseringsprocess (Grave 2008).

NB:s vision är att vara ett multimedialt kunskapscenter för levande minnen, ett center som fokuserar både på bevarande och förmedling. Det digitala nationalbiblioteket ses i den visionen som en viktig komponent och att biblioteket ska bestå av mycket digitalt material, både historiskt och modernt kulturarv, för att på så vis vara tillgängligt för många användare när och där informationen behövs (Nasjonalbiblioteket 2007, s.2). Vid invigningen av massdigitaliseringsarbetet angav generalsekreteraren i Norsk Faglitterær forfatter- og oversetterforening Trond Andreassen att digitaliseringen har betydelse för norskt språk och nationell identitet i en värld som är globaliserad och mycket information endast är tillgänglig på engelska (Myrvang 2006). Nationalbibliotekarien Vigdis Moe Skarstein är av åsikten att Norge som ung nation och med nationella samlingar i överkomlig storlek har bra möjligheter att digitalisera allt under rimlig tid (2006).

Nasjonalbiblioteket har innan nuvarande massdigitaliseringsprojekt digitaliserat i ca. 10 år. Objekt som främst digitaliserades var fotografier, ljud och tidningar (från mikrofilm). Det digitala materialet uppgår till ca. 150 000 timmar radiosändningar, över 300 000 fotografier och mer än 1 000 000 tidningssidor. Dessutom har över 25 000 böcker digitaliserats. NB digitaliserar också enskilda verk med syfte att åstadkomma en bild av mycket hög kvalitet. Digitaliseringssatsningen som NB nu arbetar med omfattar ca.:

- 450 000 böcker
- 2 000 000 tidskrifter
- 4 700 000 tidningar (60 000 000 sidor)
- 1 300 000 bilder (foto och postkort)
- 60 000 plakat
- 200 000 kartor
- 4 000 000 handskrifter
- 200 000 noter
- 1 900 000 småtryck
- 80 000 musik

250 000 timmar film och teve (Nasjonalbiblioteket 2007, s.3; Informant A 2008a)

För att kunna digitalisera alla dokument har NB utvecklat en digitaliseringsprocess som är effektiv och snabb. De har så långt det varit möjligt försökt automatisera processen gällande dataflyttning och behandling av den digitala boken, men vill också vara flexibel så att nya delprocesser enkelt kan inkluderas. Till processen räknar de följande moment: urval och beställning, hämtning av material från magasin, transport till digitalisering, hämta metadata från katalog, digitalisering (bildfångst), OCR-behandling och strukturanalys, formatkonvertering, generera bevaringsobjekt, lägga in i digitalsäkerhetsmagasin, meddela katalogen om det digitala objektet och indexera OCR-texten och metadata i sökmotorn (Nasjonalbiblioteket 2007, s.6). Att produktionslinjen skulle automatisera i så hög grad som möjligt visade sig vara svårare än NB från början hade väntat sig och att digitaliseringen var mer omfattande och mer komplex än de först antagit. Utvecklingen av digitaliseringsprocessen och NBdigital befinner sig i ständig förändring, allt eftersom arbetet fortskrider, ett tillvägagångssätt NB använt från första början vid upprättandet av det digitala biblioteket (Nasjonalbiblioteket 2007, s.11).

Nasjonalbiblioteket uppskattar att hela digitaliseringsarbetet ska ta mellan 15-20 år. I dagsläget digitaliseras 2000-3000 böcker/månaden och hela arbetet beräknas kosta ca. en miljard NOK (~ 1,2 miljard SEK). Av totalkostnaden beräknas 60 % gå till digital lagring, köp av utrustning och programvara, utveckling och integrering av system i digitalisering och efterbearbetning, samt löner till utförandet av själva digitaliseringen och en del "oppdragsmidler". Resterande 40 % är för registrering av material, etablering av nödvändig metadata och återvinning, samt hämtning av material från samlingar, konservering och återlämning av material (Nasjonalbiblioteket 2007, s.4, 8). Nasjonalbiblioteket har lagt ut en del av verksamheten till en firma i Tyskland som digitaliserar och OCR:ar NB:s mikrofilmade tidningar (Nasjonalbiblioteket 2007, s.5).

Vid digitaliseringsprojekt som ger fri tillgång till materialet är det viktigt att rättigheter-na till dokumenten klargörs. Av 450 000 titlar i NB:s samling är 5000 helt fria att publicera (Nasjonalbiblioteket 2007, s.7). NB:s mål att även digitalisera och tillhandahålla nyare dokument har inneburit att kontrakt upprättats med rättighetsinnehavare med tillåtelse om tillgängliggörande via internet. Ett sådant avtal har slutits i samband med digitalisering av verk som berör nordområdet. Avtalet med över 6000 upphovsrättsinnehavare innebar att ca. 1400 verk blev fria att studera i sin helhet. Dessa verk ingår i ett pilotprojekt för att bedöma användarfrekvens och användningsområde av NBdigital. Undersökningen, som avslutades i oktober 2008, ska därefter ligga till grund för framtida permanenta avtal med rättighetsinnehavare (Nasjonalbiblioteket 2007, s.3f.; Grave 2007, s.2; Informant A 2008a).

5.1.1 Syfte och urval

Motivet för NB:s digitalisering anges i tre punkter:

- att nå ut till så många som möjligt
- säkra bevarandet av innehållet även om originaldokumenten förstörs
- att för eftervärlden säkra material som är "född digitalt" (Skarstien 2006).

Nasjonalbiblioteket har till en början fokuserat på digitalisering av böcker². Urvalet kan NB låta ske utifrån interna eller externa förfrågningar. NB har ett systematiskt urval, där äldst material tas först. Rättigheterna har då oftast upphört och är fria att publicera. NB prioriterar även material bundet till kultur- och samhällspersoner som kan vara aktuella vid ex. jubileum. NB digitaliserar även utifrån efterfrågan av material, dessa förfrågningar får högre prioritet än den systematiska digitaliseringen. Ett sådant avtal har slutits mellan NB och en rad organisationer om att digitalisera böcker och tidsskriftsartiklar som berör nordområdet, alltså Nordnorge och Arktis. De senaste åren har nordområdet uppmärksamats för ett flertal intressen, bl.a. politisk, strategiskt, klimatmässiga och kommersiella (Nasjonalbiblioteket 2007, s.4, 9; Bakken 2006; Støre 2007, s.83).

Norges utrikesminister 2007, Jonas Gahr Støre, rosar Nasjonalbibliotekets val att digitalisera dokument utifrån fokus nordområdet, och beskriver det som en viktig kunskapsutveckling och säger därefter:

Et nasjonalbibliotek forvalter en svært viktig ressurs og er en informasjonsbase for hele nasjonen. Jeg er derfor glad for at Nasjonalbiblioteket nå satses på nordområdene, og det på en måte som gjør informasjonen tilgjengelig også for de yngre generasjoner, som i første rekke henter kunnskap elektronisk på nettet. Kunnskap er navet og Norge skal være ledende på kunnskap om nordområdene. (Støre 2007, s.84)

Att NB digitaliserar och tillgängliggör dokument om nordområdet ses av Støre som en satsning på norskt kunskapsfrämjande. NB betonar dock att valet att digitalisera objekten inte är en beställning från regeringen utan en egen värdering hos NB om vad som är relevant (Informant A 2008a).

Nasjonalbiblioteket har utvecklat ett system för urval av material som ska digitaliseras. Systemet väljer automatiskt objekt utifrån om NB har tillräckligt med exemplar och startar då med det äldsta. Genom hela digitaliseringsprocessen kan speciellt prioriterat material ges särskild företräde. Systemet vet dock inget om dokumentets fysiska hälsa. Verk kan därför väljas ut för digitalisering trots att bättre exemplar finns i samlingen (Nasjonalbiblioteket 2007, s.9, 12).

Att allt ska digitaliseras innebär konkret att allt som finns i Nasjonalbibliotekets samlingar ska digitaliseras. Det innebär även ett verks olika utgåvor, men om andra redan digitaliserat verket så görs ingen ytterligare digitalisering av NB (Informant A 2008a; Informant A 2008b).

5.1.2 Förberedelse, bildfångst och bearbetning

Digitaliseringen vid Nasjonalbiblioteket utförs för att göra bevarandet av samlingen mer effektiv och mindre utsatt för fysisk nedbrytning. Vidare skriver de att: ”digitaliseringen må gjøres i en kvalitet som er høy nok til at man gjennom digital bevaring på et senere tidspunkt kan gjenskape egenskapene til originalen på en tilfredsstillende måte.”, alltså att NB utifrån den digitala varianten kan tillfredställande återskapa egenskaperna hos originalet. Verken ska reproduceras i en kvalité som motsvaras av det skannade originalet. Materialet ska av användaren upplevas som en korrekt representation av verket, baserat på värderingar av kvalité som: färg, upplösning, funktionalitet: som att sidorna är i

² Notera att resultatredovisningen av Nasjonalbibliotekets digitalisering främst handlar om digitalisering av böcker.

rätt ordning och att man kan bläddra i boken. För objekt som existerar i många kopior, (vilket är vanligt för nya verk) läggs det i kvalitén betoning på läsbarhet. För mer unika verk viktläggs lojalitet mot originalet vid skanningstidpunkten (Nasjonalbiblioteket 2007, s.6; Informant A 2008c).

Nasjonalbibliotekets förberedelse är att böckerna demonteras i fall NB har minst tre exemplar i sitt egna bibliotek. Demonterade exemplar blir efter digitaliseringen kasserade. Är exemplaren färre skannas de manuellt med bläddring, vilket sker två sidor i taget. De allra sårbaraste dokumenten skannas under övervakning av konservator. I nuläget uppskattar NB att ca. en fjärdedel av alla titlar i boksamling kan demonteras före skanning. För att öka andelen finns det planer på att erbjuda landets bibliotek att skicka exemplar av böcker NB har få av, och att dessa demonteras för att på så vis höja farten på hela digitaliseringsarbetet (Nasjonalbiblioteket 2007, s.7f.).

Vid demontering av böcker använder NB två specialsaxar. Vid bildfångsten används tre pärmskannrar (i2s Copibook) och två skannrar med automatisk bladmatning (Agfa S 655). För bläddringsskanning används typen i2s Suprascan där fem stycken är A2-skannrar för normalt material och en A0-skanner för speciellt material (Nasjonalbiblioteket 2007, s.10). Skannrarna som digitaliserar demonterade böcker skannar bägge sidor på varje ark i en operation. Det betyder att två olika digitaliseringsenheter används på de två sidorna. NB har haft väldigt svårt att kalibrera dessa helt lika, vilket resulterat i färgskillnader på sidorna. Ett problem de försökt lösa, men som fortfarande kvarstår (Nasjonalbiblioteket 2007, s.12).

Bildfångsten av böcker sker med upplösning 400 dpi och ett färgdjup på 24 bit, och som bevarandeformat används JPEG2000, med ickeförstörande komprimering. Detta sker oberoende av dokumentstorlek och detaljrikedom i text och illustrationer. Den digitala mastern är då i filstorlek ca. 20 Mb. NB har gjort bedömningen att de genom att använda JPEG2000 som filformat och inte TIFF reducerar lagringsbehovet med 50 %, vilket innebär en besparing på 70 miljoner NOK. Efter egna praktiska försök har NB kunnat visa att det är möjligt att konvertera tillbaka till ickekomprimerat TIFF utan att förlora någon information. De är medvetna om att ett bitfel kan förstöra hela bilden i JPEG2000, där samma fel i TIFF endast ödelägger den enskilda pixeln, men bedömer risken minimal (Nasjonalbiblioteket 2007, s.6f.; Informant A 2008a).

Vid bildfångsten är det för fotografier specialanpassade lokaler med bl.a. ljus och monitorer. Lokalerna för bokdigitalisering är fysiskt tillrättalagt med exempelvis bord och luftrengöring av damm (Informant A 2008b).

Masterbilden som bevaras är bearbetad. Materialet som genomgår automatiskskanning färgkorrigeras för att förenkla framvisningen. Manuellt skannat material korrigeras emot originalet. Böcker som skannas blir digitalt beskurna och boken döljs därmed, vilket även gäller för masterbilden. Fotografier genomgår färgkorrigerings och dammbortplockning, vilket sker utifrån subjektiva kvalitetsvärderingar. I tillägg används specialiserade programvaror (Informant A 2008a; Informant A 2008b; Informant A 2008c).

5.1.3 Metadata och uppmärkning

Metadatatyperna som används av NB är deskriptiv, teknisk, bevarande, strukturell samt OCR-texten. På varje bok som är registrerad i bibliotekskatalogen finns en unik streckkod som sammanbinder boken med katalogsystemet Bibsys och den enskilda katalogposten. Från den posten hämtas bokens bibliografiska metadata och kopplas ihop med ett digitalt id och allt läggs därefter i en XML-fil. NB registrerar teknisk/administrativ metadata om bl.a. produktionssätt, datum, process, men inte om vilka personer som är involverade. All metadata organiseras enligt METS-schemat, där metadata för varje boksida registreras (Informant A 2008a).

Efter bildfångsten läggs den digitala boken och tillhörande metadata i en temporär lagerplats. Böckerna måste därefter manuellt importeras in i programmet docWorks, som används för strukturanalyser och OCR, en process som görs helt automatiskt. Texten indexerats tillsammans med metadata i sökmotorn. Det sker en automatisk strukturanalys där eventuell innehållsförteckning uppmärksammas och där sidnummer i boken ordnas så den korrelerar med visningsgränssnittets paginering, så att de kan återfinna kapitel, avsnitt och sidnummer. Programmet som används, docWords, meddelar om processen inte utförts korrekt enligt förbestämda gränskriterier och feltoleranser (Nasjonalbiblioteket 2007, s.8, 10; Informant A 2008b).

I docWorks är det möjligt att göra avancerade strukturanalyser, men i nuläget bedömer NB det vara för komplicerat att använda då det skulle krävas omfattande manuell efter- och kvalitetskontroll. Enligt NB skulle efterkontroll för varje sida ta 15 sekunder, resurser de inte har möjlighet till. Texturor och illustrationer med tillhörande text får heller ingen speciell behandling. Utvalda delar av samlingen kan däremot bli aktuell för mer avancerad strukturanalys, exempelvis värdefullt material, material av stor aktualitet, eller speciellt intressanta tidningar. Tillsammans med METS-filen används ALTO (Analyzed Layout and Text Object) som gör det möjligt att beskriva positionen i bilden för alla ord (Nasjonalbiblioteket 2007, s.8, 13; Informant A 2008a; Informant A 2008b; Informant A 2008c).

NB hade större förhoppningar på OCR-texten men har fått sänka kraven till ett absolut minimum för att kontrollen inte ska ta för mycket tid, då avancerad strukturanalys kräver manuell arbetsinsats. Vid OCR-bedömning har latinska bokstäverna haft störst fokus och precisionen på dem anser NB vara acceptabel. För frakturskrift är det sämre, NB anser dock att det går att använda för fritextsökning med någorlunda tillfredställande resultat. Hela OCR-systemet tros kunna förbättras allt eftersom programmet lär känna bokstäverna (Nasjonalbiblioteket 2007, s.13).

5.1.4 Kvalitetsbedömning

I testfasen av projektet gjordes kvalitetssäkring för alla digitaliserade sidor. Därefter har NB utvecklat kvalitetsbedömningar utifrån material och vilket värde och ålder de har och det finns kvalitetsrutiner för varje dokumenttyp och moment. Det görs även osystematiska kontroller av materialet som är tillgängligt över nätet i NBdigital (Nasjonalbiblioteket 2007, s.12; Informant 2008a).

För böcker finns en viss subjektiv kvalitetskontroll av struktur, sidnumrering och bildkvalité. Vid automatisk skanning används ColorFactory för färgkorrigering, där flera korrigeringsvarianter satts upp, beroende på den automatiska värderingen av varje enskild sida. Det innebär att sidorna blir normaliserade till vad NB beskriver som ”god visningskvalité”, och inte direkt en återgivning av originalet vid skanningstidpunkten. De böcker som skannas automatiskt genomgår en operatörskontroll under skanningsprocessen som värderar utförandet. Av färgkorrigerade sidor görs stickprovskontroller. För manuell skanner och för skannrar med automatisk ”turn-page” sker justeringar av skannern för att bilden ska vara lojal mot originalet vid tidpunkten för bildfångsten. Processen övervakas av operatörer som även utför osystematiska stickprov av sidorna. Äldre böcker kontrolleras mer noga än nya. OCR-texten omfattas dock inte av kvalitetskontrollen, utan där saknas den helt. Det finns heller inte några planer på att granska OCR-texten, den godtas som den är, så länge som texten primärt används för sökningar. En viss kontroll finns i och med upptäckta fel vid användning av NBdigital. Fotografier genomgår en noggrannare kontroll än böcker, och så gott som alla fotografier granskas. Vid digitalisering av tidningar från mikrofilm görs en automatisk kontroll för att bl.a. se om det saknas sidor, korrekt namngivning, om all relevant information finns med. Det görs också en subjektiv kontroll av bildkvalitén och filstorleken (Informant A 2008a; Informant A 2008b; Informant A 2008c).

Nasjonalbibliotekets rutin är att kalibrera utrustningen, och det finns dagliga rutiner som inkluderar rengöring av dem (Informant A 2008a).

5.1.5 Lagring av digitala original och kopior

Nasjonalbiblioteket använder sig av digitala säkerhetsmagasin vars infrastruktur ska garantera att innehållet bevaras över lång tid. Där ska allt förvaras som digitaliseras av NB. Det digitala säkerhetsmagasinet frikopplar det digitala innehållet från den använda teknologin för lagring. Detta för att förenkla vid migrering till nya generationer av lagringstekniker utan att systemet för hämtning av digital data berörs (Nasjonalbiblioteket 2007, s.8). NB räknar med att de blir tvungna att migrera minst vartannat eller vart tredje år, något som ska ske automatiskt. För att bibehålla autenticiteten och att materialet inte ska förändras oavsiktligt används kontrollsummor av typen MD5 (Informant A 2008a; Nasjonalbiblioteket 2008a).

Det digitala innehållet lagras i tre kopior på två olika lagringsmedier i det digitala säkerhetsmagasinet. En av kopiorna lagras på disk medan de två andra lagras på band (Nasjonalbiblioteket 2007, s.8). NB har idag tre maskinhallar ämnade för digital lagring på två skilda platser. Lagringen och kontrollen av digitaliserat material ingår i NB:s vanliga verksamhet för att skydda och hantera data (Informant A 2008a; Informant A 2008b).

När böcker hamnar i det digitala säkerhetsmagasinet betyder det att dokumentets digitala id läggs i bibliotekskatalogen. Därefter indexerar materialet automatiskt och läggs i det digitala biblioteket. En del av NB:s material innefattas inte av rättigheten att publiceras på internet, men när det placeras i det digitala säkerhetsmagasinet sker det automatiskt. För att undvika att material blir fritt tillgängligt för allmänt bruk läggs dessa digitala objekt på sk. mellanlagring. Materialet är dock lagrat med samma policy som för materialet i säkerhetsmagasinet (Nasjonalbiblioteket 2007, s.13f.).

5.1.6 Framvisning

NB använder NBdigital för framvisning, vilken bygger på sökmotorer istället för traditionell databaslösning. I NBdigital är det möjligt att söka i både metadata och fulltext, vilket sker tvärs över materialtyperna. I metadatan går det att avgränsa sökkriterier och i realtid bygga upp alternativa vägar till materialet (Nasjonalbiblioteket 2007, s.8f.).

Startsidan för NB: s digitala arkiv - <http://www.nb.no/sok/search.jsf> - ger åtkomst till materialet med en sökmotor som är utvecklad av företaget Fast. Den är baserad på, vad Lervik och Brygfeldt kallar, den tredje generationens sökteknologi, en teknik som ska ge möjlighet att söka och hantera NB:s stora samlingar med fler än 30 databaser, detta oavsett materialtyp. Sökresultatet ska bygga på en kombination av webbaserade sökmotorers enkelhet men med nya förbättrade relevansmodeller, byggd på kontextuell relevans (2006, s.14f.). Kontextuell sökning baseras på 1: intelligent text och informationsåtervinning som utgår ifrån koncept modeller användandet av mönster och relationer mellan entiteter. 2: Flexibel ämnesstrukturering med XML-schema och förbindelser av innehåll med strukturerade och ostrukturerade källor. 3: Att snabbt samla, processa och återvinna informationen av stora kvantiteter, utan att användaren upplever fördröjning (Olstad & Seres 2005, s.10).

Vid söktillfället finns ingen möjlighet till avancerad sökning. Det val användaren kan göra är ” ” för frassökningar och * för trunkering. Det förefaller inte finnas möjligheter till boolska operatorer, undantag för AND/OCH som automatiskt används mellan termerna (Nasjonalbiblioteket 2008b). När man söker skrivs termerna i en sökruta och resultatet visas därefter i en lista som innehåller alla dokumenttyper. Till vänster om dessa är det möjligt att specificera sökkriterierna på exempelvis materialtyp, tema, upphovsman, ämne, år, platser, bibliografier och klassifikation enligt Dewey. Vill man ha material av digitalt innehåll måste detta också preciseras. När man klickar på någon av dessa kategorier visas materialet enligt det smalare kriteriet. Se bilaga 1 för exempel på sökningen ”noreg”, nynorsk stavning av ”Norge”³. I databasen förekommer texter på bokmål, nynorska och nordsamiska, men sökmotorn har inget tillrättalagt för att hantera dessa på något speciellt sätt. En sökning på ”Norge” skulle inte hitta texter där Norge genomgående benämns ”Noreg” eller ”Norga”(nordsamiska för Norge) (Informant A 2008b).

I resultatlistan för digitalt material visas kortfattat författare och titel, och ett kort utdrag från OCR-texten från vars avsnitt söktermen upptäcktes. Genom att klicka på ”Les mer” förflyttas man till den aktuella delen av boken. Här visas endast den digitaliserade boksidan. Var på sidan termen står preciseras inte.

I visningsläge för boksidan finns till vänster möjlighet att ta fram bibliografisk information om boken med länken ”Vis opplysninger om boka”. Under den kan man skriva in sidnummer och direkt förflyttas dit. Vidare så finns direktlänkar till eventuell innehållsförteckning. För en del fotografier ges en kortfattad beskrivning, eller berättar var fotot är taget, liksom objektets material och storlek.

³ Jag sökte med gemener. Korrekt stavat skulle ha varit med stor första bokstav – Noreg. Sökmotorn gör dock ingen skillnad på stora och små bokstäver.

När användaren väljer att se en skannad bild så hämtas bilden ifrån det digitala säkerhetsarkivet och med komprimeringsalgoritmer skapas i realtid en enklare visningskopia av det digitala originalet. Men denna metod kan NB enkelt byta ut algoritmen om en ny och bättre algoritm utvecklas. NB har tre eller två kvalitetsnivåer för användaren att välja på. Dessa accessbilder är i övrigt inte förändrade jämfört med masterbilden, men det finns idéer om ändring så att läsbarheten i dessa kan förenklas och höjas (Nasjonalbiblioteket 2007, s.7, 14; Informant A 2008a).

Det finns planer på att anpassa visningsgränssnittet för olika användare, med fler funktioner, att använda så kallad rollbaserad tillgång till det digitala materialet. Tillgång till icke fritt material kan ges till universitet och högskolor i Norge. Det är också institutionerna som styr autenticiteten av användarna. NB behöver därför inte känna varje person som använder deras material. Detta är än så länge inte sjösatt (Informant A 2008a; Nasjonalbiblioteket 2007, s.9).

5.2 Kungl. Biblioteket i Sverige

Kungl. biblioteket (KB) började under perioden 1998-1999 digitalisera i mindre projekt. De första dokumenten var affischer, okatalogiserade tryck, projektet ”svenskt tryck före 1700” och tidningar. Verksamheten avtog därefter, främst p.g.a. bristande ekonomiska medel. Projekten ledde till insikten att KB behövde mer samordning emellan enheterna i digitaliseringen. KB har därför satsat på projekt som syftar till att organisera digitaliseringsverksamheten. Sverige och KB har digitaliserat förhållandevis lite material, enligt KB är orsaken bristande resurser och bristande samordning (SOU 2003:129 2004, s.182f.). För att få arbetsro och stabilitet i digitaliseringen anser KB att långsiktig finansiering från uppdragsgivare eller andra finansiärer, såsom årligen återkommande statliga medel, är det enda som kan underlätta byggandet av en stabil och komplex verksamhet (Scherman 2005, s.11).

Finansieringen till många av KB:s långsiktiga digitaliseringar har varit från överskottsmedel som biblioteket fick disponera under 1998 och 1999. Resurser har därefter varit svårt att få tag på. Under 2002 och 2003 i samband med digitaliseringen av verket *Suecia antiqua* användes medel inom ramen av KB:s verksamhet. Digitalisering av tidningar har delvis finansierats av olika organisationer, men KB stod fortfarande för de största kostnaderna. KB anser att statsbidrag bör tilldelas för att de ska få möjlighet att digitalisera viktiga delar av samlingar och hålla de tillgängliga för studier och forskning (SOU 2003:129, s.189). I slutet av 2008 kunde KB dock rapportera att de fått 8.3 miljoner extra som ska läggas på digitalisering i forsningsändamål med bevarande av material som riskerar att förstöras (Kungl. biblioteket 2008c-12-16). Annars ges idag inga pengar öronmärkta för digitalisering (Informant B 2008d).

1999 gjordes ett utredande projekt som sammanställdes i rapporten *Plattform för bild-databaser*. Där undersöktes förutsättningar för en gemensam infrastruktur, standarder för digitalisering och långtidslagring av databaser. Tanken var även att oavsett om det analoga verket är textburet eller bildburet så ska de digitala bildfilerna samlas i en gemensam bild-databas och sökas från en gemensam plattform. På sikt skulle det kunna integreras i katalogen Libris system (Gram & Kjellman 2000, s.5, 6).

DIGSAM (Digitalisering och dess samordning inom KB) är namnet på ett KB-projekt som pågick mellan 2003-2005. Projektets övergripande målsättning var att utreda hur ett större digitaliseringsprojekt påverkar KB:s organisation, samt vilka verktyg som behövs för att etablera omfattande digitalisering som en del av KB:s verksamhet. Deluppgifter inom projektet var att utreda och utarbeta rutiner för digitaliseringsverksamheten inom urval, kvalitetsnivåer, flödesscheman samt teknisk och administrativa metadata. Det innefattades även undersökningar av administrativa system, digitala utställningsverktyg och hur inomnationella samarbeten kan göras. DIGSAM-rapporten innehåller även checklistor för varje delområde som ska användas på objekten för att bl.a. bedöma aktualitet för digitalisering och om KB har vad som krävs för att genomföra det (Scherman 2005, s.15, 28ff.).

KB har gjort flera små digitaliseringsprojekt varav endast en del är nåbara över internet. De som nämns på KB:s webbsida för digitala samlingar är följande: ”Resor genom tiderna”, ”Linnés nätverk”, ”Codex Gigas”, ”Affischer”, ”Suecia antiqua” och ”Kartor”.⁴ Varje digitalisering som gjorts har varit unik på så sätt att inget objekt/samling varit lika och att variationer krävts, bl.a. i hur samarbetet avdelningarna emellan försiggått (Persson & Tångemar 2006, s 26, 70). Den digitalisering som bl.a. gjordes av affischer i mitten och slutet av 1990-talet med dåvarande teknik och kompetens skulle idag inte anses var av tillräcklig kvalitet (SOU 2003:129 2004, s.187).

KB har konstaterat att deras nuvarande högkvalitativa utrustning är otillräckligt för mer omfattande digitaliseringsarbeten utöver kundbeställningar, att digitalisera allt skulle ta mycket lång tid. KB har därför undersökt vilka möjligheter det finns i att upprätta produktionsenheter på andra platser och organisationer och även möjligheter till massdigitalisering, men mycket av materialet som i dagsläget är högt prioriterat är så ömtåligt, unikt och värdefullt att KB anser att de själva bör göra digitaliseringen p.g.a. de höga krav som de sätter (Scherman 2005, s.39; SOU 2003:129 2004, s.189).

5.2.1 Syfte och urval

Kungl. biblioteket anser att medborgarna har rätt att få ta del av nationalbibliotekets samlingar och att deras övergripande mål med digitaliseringen är att öppna det för omvärlden (nationellt & internationellt). Digitaliseringen ses även som ett skydd för att kommande generationer ska kunna ta del av samlingarna. För att det ska bli möjligt drivs verksamheten med ett långtidsperspektiv. Tillgängligheten anser man bäst uppfylls med internet som distributionskanal. Främsta målgruppen med digitaliseringen är utbildningssektorn i dess vidaste mening, vilket innebär lättåtkomliga samlingar för elever, studenter, lärare, forskare och den intresserade allmänheten. KB ser sitt uppdrag även i bevarandaspekter och att material förhoppningsvis blir mindre utsatt för slitage i och med digitaliseringen (Scherman 2005, s.21f.). I Persson och Tångemars studie framkom att argument för att digitalisera inom KB:s projekt hade en lutning åt bevarande framför tillgängliggörande, 12 projekt gjordes med bevarande som främsta argument och 7 för tillgängliggörande (2006, s.67f.). Många av KB:s digitaliseringar är testprojekt av olika metoder och tekniker som ska ligga till grund för senare, större projekt. Digita-

⁴ Persson och Tångemar räknar i sin uppsats upp 15 verk/samlingar som då var nåbara över internet (2006, s.25).

liseringen har avsikten att den digitala bilden ska vara i så hög kvalitet att det kan ersätta tillträde till originalet (Persson & Tångemar 2006, s.57, 60, 65)

KB saknar än så länge riktlinjer för vilket material som ska väljas ut för digitalisering och därefter presenteras på webbplatsen. Något som skapar problem vid fortsatt digitalisering, riktlinjer sägs dock vara på gång (Digitaliseringsmanual, s.14; Informant 2008f). Även om KB saknar tydliga riktlinjer för urval så existerar kriterier för materialet som kan bli aktuellt för digitalisering, såsom hela samlingar/block av litteratur, bevarande av sköra och unika samlingar/böcker, tillgänglighet från både nationellt och internationellt, information till forskningsvärlden – att visa väg till nytt material (Digitaliseringsmanual, s.15). Vid KB:s bevarandeenhet finns fyra kriterier som ett objekt ska definieras utifrån: informationsvärde (objekts innehållsliga), originalvärde (objekts värde som original som administrativt, historiskt, juridiskt eller p.g.a. upphovsman), inneboende värde (objekts symboliska eller sentimentala) och ekonomiskt värde (Digitaliseringsmanual, s.16f.). Majoriteten av urvalet som hittills digitaliserats är bildmaterial. Detta är medvetet och är för att bilddigitalisering har av KB uppfattats vara enklare och att bilder har starkare koppling till dataskärmen (Persson & Tångemar 2006, s.31f., 59).

Persson och Tångemars intervjustudie visade också att material har digitaliserats då de varit i dåligt skick och inte skulle tåla en framtagning/utlåning. Objekt som varit efterfrågade av användare och dokument som anses spännande har också digitaliserats (2006, s.57). Speciella händelser kan aktualisera digitalisering av berörande dokument, ex. 300-årsjubilet av Carl von Linné år 2007 (Digitaliseringsmanual, s.16).

Vid digitaliseringen av *Suecia Antiqua et Hodierna* valde KB ut vilket exemplar de skulle använda, då de har 10 olika versioner. Verket som digitaliserades var det i bäst skick, där hela verket digitaliserades, och inte bästa delarna från varje exemplar (Persson & Tångemar 2006, s.38).

Valet att digitalisera *Codex Gigas* togs i samband med att verket lånades ut till Tjeckien på politiskt initiativ, de ville ge Tjeckien en gåva för att åtgärda att Sverige tog verket som krigsbyte. Den digitala versionen fick även bli en form av säkerhetskopior (Persson & Tångemar 2006, s.31; Informant C 2008a).

5.2.2 Förberedelse, bildfångst och bearbetning

Målet med digitaliseringen är att den digitala kopian ska kunna ersätta användningen av originalet, fungera som substitut för originalet (Persson & Tångemar 2006, s.57; Scherman 2005, s.41). DIGSAM-rapporten beskriver KB:s definition som att masterfilen ska vara det optimala som utrustningen kan skapa. En utrustning som ska följa utvecklingen. Kvalitén hos filen ska kunna tillgodose samtliga behov som kan komma efterfrågas, exempelvis enkla arbetskopior, webbvisning, tryckkvalité och faksimiler (Scherman 2005, s.41).

Dokumentet som väljs ut för digitalisering hamnar hos bevarandeenheten som beslutar hur bildfångsten ska förberedas. Konservatorer ska både innan och efter bildfångsten se över materialet. Förberedelser kan vara att rengöra material, åtgärda skador, eller andra skyddsmanövrar för att dokumentet inte ska ta skada. Vid projektet *Öppna samlingar* gjordes rengöring såsom, avlägsning av smuts, torrengöring, fuktbehandling för utslät-

ning, revor och lösa sidor säkrades och tejp plockades bort. Bevarandeenheten tillsammans med fotoavdelningen utformar hur lösningen för produktionen ska vara och hur en bildfil av hög kvalité kan göras utifrån objekten och deras förutsättningar. Återkommande frågor mellan bevarande och foto anges vara bokstöd (hur objektet ska placeras), bakgrundsmaterial (bakgrundsfärger och dess storlek), färgreferenser (svart/vit-punkt, linjal och färgreferenser) och hantering (bläddring och flexibiliteten hos objektet) (Digitaliseringsmanual, s.24f.; Scherman 2005, s.22).

KB gör därefter beräkningar för hur många fotografiska exponeringar som behövs av hela objektet, där frampärmen är nr. 1 och insidan som nr. 2, o.s.v. Om objektet finns i dubletter så ska endast ett av dem väljas (Digitaliseringsmanual, s.56).

KB har till sitt förfogade två kameramodeller (Sinar M & Nikon D200) och två skannrar (Microtek ScanMaker 100XL & Kobolt Bron Lumax SB 24). Skanningen/fotograferingen ska ske rakt ovanifrån, och ljuset ska falla jämt över hela objektet. Den använda skalan är 1:1 och upplösningen 400 ppi för objekt större än A5, men med maximal storlek ca. 50x70cm. Om objektet är mindre än A5 ska bildfångsten ske i 800 ppi. Färgdjupet för KB:s digitalisering är 16-bit/kanal, totalt 48-bit. I samband med bildfångsten får ingen skärpning förekomma (Digitaliseringsmanual, s.57f.).

Inkluderat vid bildfångsten är referenshjälpmedel för att underlätta efterbearbetningen. Det som infogas är linjal, färgpaletten Gretag Macbeth Mini ColorChecker Chart, vit- och svartpunktsredskap för kalibrering av tonomfånget. Vid skanningen ska objektet placeras rakt och med referenshjälpmedel på avståndet 20-40mm från objektet (Digitaliseringsmanual, s.57).

Efterhanteringen av analog objekt som digitaliserats är placering i skyddande förvaring med etikett om att objektet har digitaliserats och att användning hänvisas till den digitala versionen (Digitaliseringsmanual, s.25).

Masterfilen som genereras i samband med bildfångsten ska vara korrekt exponerad och snygg och alla toner ska vara representerade, så att histogrammet är sammanhängande och utan hål. Filen ska vara fri från damm. Filen konverteras till arbetsfärgrymden ProPhoto RGB, och inga korrigeringar får förekomma. Filen sparas därefter i okomprimerat TIFF. Utifrån masterfilen skapas en faksimilfil/arbetsfil, som korrigeras för att överensstämma med originalet. Med referenshjälpmedlen anpassas bildens färger och ljus, och sparas därefter i TIFF som egen fil (Digitaliseringsmanual, s.58f.).

Vid projektet *Codex Gigas* var förberedelsen att bland annat bygga en speciell bokvaga, så att kameran kunde monteras 255cm ovanför objektet. Bildfångsten som sådan gjordes i 300 ppi, och inte som vanligt 400 ppi då kamerans sensor inte tillät det, och storleken på den digitala bilden blev då upp emot 650 Mb. Tillsammans med de 1256 bilderna blir det en stor samling och kräver stort utrymme. Varje exponering varade 1,5 min. *Codex Gigas* anpassades till framvisning så att färgerna är korrigerade och filstorleken ca. 50Mb och sparat i filformatet pyramidtiff och slutligen förbättrades skärpan (Flemming 2007, s.30f.; Informant C 2008a). Pyramidtiff är ett format som sparar bilden i flera kvalitetsnivåer. Om fotot betraktas på första nivån så är kvalitén sämre än om man zoomar in, då bilden visas i detaljerad nivå. Formatet kan därmed sägas bestå av flera versioner av en och samma bild.

5.2.3 Metadata och uppmärkning

Bibliografiska metadata hämtas primärt från katalogposten i KB:s bibliotekssystem. De betonar om vikten av en referens tillbaka till den bibliografiska posten i systemet, för om dessa ändras så kan posterna i magasinet och arkivet enkelt uppdateras (Informant 2008b).

Hur mycket teknisk metadata som lagras skiljer mellan vilka hårdvaror och mjukvaror som används. Kamerorna som används vid bevarandedigitalisering tar upp väldigt lite teknisk metadata, medan den skanner som är inköpt för digitalisering av dagstidningar fångar upp desto mer. Övrig teknisk metadata läggs till manuellt (Informant B 2008b; DIGSAM 2005, s.42). I ett av KB:s arbetsdokument om metadata, som främst baseras på tidningar men delvis kan användas för övrig bevarandedigitalisering förekommer en mängd metadatauppgifter som ska kan registreras. Där nämns vid sidan av basala bibliografiska uppgifter som namn och nummer bl.a. var den fysiska tidningen kommer ifrån, fel eller skador, originalets typsnitt, accessrättigheter, precisering av tidningens innehåll (som kan göras ner till enskild sida). Teknisk metadata som anges är t.ex. beskrivning av programvara med bl.a. version, hårdvarubeskrivning, komprimering, checksumma, färgprofil, referenser. OCR-texten ska kunna innehålla, antal säkra och osäkra tecken för hela numret och per sida, ordlista som använts (Informant B 2008b; Kungl. biblioteket 2008c, s.1ff.).

Vid OCR-tolkningen använder KB programmet Abby FineReader OCR XIX, men modul för frakturskrift. Som förberedelse inför texttolkningen studeras ett par sidor av dokumentet för att avgöra användarmönster, vilka bokstäver som riskerar att förväxlas med varandra. Där risker för fel finns görs särskilda regler för dessa, s.k. bokstavsbilder (Digitaliseringsmanual s.74). Textuppmärkningen görs endast för enstaka utvalda objekt, och är inget som är generellt för allt material (Informant B 2008e).

OCR-skanningen görs i 400 ppi i 8 bitars gråskala. Alla sidor tas med även de som saknar text. Med hjälp av en ordlista utförs därefter tolkning. Om felprocenten blir för hög eller har återkommande fel så görs förändringar i förberedelsen och valda bokstavsbilder. OCR-läsaren lägger automatiskt in element i texten som indikerar sidbrytning. Den anger också hur många osäkra tecken det finns, en variation som hos KB är 2-15%. Vid redigeringen av text är principen att det ska motsvara originalet och att tryckfel därmed kvarstår, men kan i samband med textkodningen markeras med speciella koder (Digitaliseringsmanual s.75ff.).

När KB märker upp texter är det enligt standarden som är utarbetad av TEI. Uppmärkningen görs enligt en viss grammatik - DTD. KB använder egenutvecklade DTD:er som mall för uppmärkningen (Digitaliseringsmanual s.83f.). KB har definierat och använder två nivåer vid uppmärkning av böcker. I den första nivån eftersträvas att formatering och layout ska efterlikna originalet, och följande märks upp: strukturen (dvs. delar, kapitel, avsnitt, och deras rubriker), sidbrytningar, styckeindelning, radbrytningar, titelsida, paginering, sidhuvud, fotnoter, tabeller, register, kustoder, arksignaturer, avstavningar, formateringar som kursiv text, hängande indrag, KB:s exlibris, och slutligen anteckningar av bibliotekarie. En stor del av den uppmärkning görs automatiskt. I uppmärkningen enligt nivå två märks sådant upp som är av relevans för sammanhanget. Där exempelvis orsnamn i projektet *Tema resor* märks upp och blir sökbara via databaser. Det kräver med andra ord mer tolkande processer, om vad som är relevant i texten. Vid

avancerade och komplicerade texter anser KB att expertis bör rådfrågas i fråga om transkriberingen (Digitaliseringsmanual s.72, 78f.).

Texterna ska kunna presenteras på två sätt, på webben och i PDF-format; för nedladdning. Det föreligger en skillnad dessa emellan. Webbversionen behåller originalets sid-, stycke och radbrytningar. Detta för att enkelt kunna orientera sig mellan de digitala bilderna och e-texten. Typografins enhet imiteras inom vissa ramar. I PDF-versionen bevaras inte originalets form då det leder till dålig typografi. KB tror inte att kravet är möjligt och inte heller önskvärt. Markeringar som visar sidbrytningar m.m. kan möjligen göras i texten (Digitaliseringsmanual s.86f.).

Vid digitaliseringen av *Codex Gigas* genererade inte mycket information ifrån kameran. Informant C minns bara att det var upplösning och filformat, men är lite osäker på om det helt stämmer. För *Codex Gigas* har det inte gjorts någon OCR. Fotografen Pelle B Adolphson förde under arbetet en loggbok med noteringar om mätinställningar och intressanta iakttagelser av verket (Flemming 2007, s.31.; Informant C 2008a).

5.2.4 Kvalitetsbedömning

På grund av enkelheten i att kopiera och förändra digitalt material så anser KB det vara av största betydelse att garantera materialets autenticitet, där kvalitetskontroll i alla led är en viktig del för att säkerhetsställa äktheten och tillförlitligheten, så att slutresultatet kan inneha kvalitetsförsäkringen. För att uppfylla reliabiliteten för arbetet utförs tester av kvalitetskontrollerna och en uttömmande dokumentation av använda metoder. Hur tillförlitliga metoderna är kan de sedan i efterhand utvärdera mot resultatet. Kvalitén ska vara på sådan nivå att originalet kan spärras för framtagning (Digitaliseringsmanual s.59).

Processer och delprodukter som underkastas kvalitetskontroll är masterfiler, övriga bildfiler/accessbilder, teknisk metadata från bildfångsten, bibliografisk metadata och katalogposten, uppmärkt text och presentationen (Digitaliseringsmanual s.60f.).

KB har vissa grundkrav och frågor gällande kvalitén på de digitala objekten. Kravet är att hela objektet är fullständigt återgivet så att inget har förlorats längs vägen, att alla bilder finns med, och då endast en gång och att inget hoppats över. Att hela objektet finns på bildrutan och att inget hamnat utanför. I textdokumenten ska alla bokstäver vara läsbara, även handskrivna text. För bilder betyder läsbarheten att kvalitén ska vara satt på rimlig nivå med detaljskärpa och korrekt färgåtergivning. I tidigare avsnitt beskrevs att efterbearbetningar gjordes och att dessa utfördes för att möta trogenhet hos originalet. Det ska också återge papperets textur, graderingar av bläckfärg osv. OCR-texten ska rättas vid förekomsten av fel vid inläsningen och textuppmärkningen ska stämma överens med originalobjektet i både textstruktur och innehåll (Digitaliseringsmanual s.60).

Kvalitetskontrollen ska göras jämförande med den fysiska förlagan tillhands. KB använder kontrollsteg, vilka är anpassade för varje materialslag. Kontrollstegen innefattar att försäkra om att rätt objekt digitaliserats, kontrollerar att antalet exponeringar stämmer, att höger och vänstersidor följs åt, att bilder inte avviker i exponering i ljushet eller mörkhet, att textraden nederst på sidan är rak, kontrollera att digital sidpaginering

stämmer överens med metadata och att alla texter är läsbara (Digitaliseringsmanual s.62ff.).

Att använda sig av e-texter, OCR:ade texter, är förhållandevis nytt på KB och befinner sig i ett utvecklingsstadium. Kvalitetsförsäkringar, kontrollkriterier och nivåer av texter är därför inte alltid helt fastställda. KB ser det som en möjlighet att tidigt i arbetet med texter kunna fokusera på frågorna om kvalitet av standarder. Det innebär också att arbetet ständigt revideras (Digitaliseringsmanual s.68). Fel som KB tar upp som kan inträffa i samband med OCR-läsningen är kända, exempelvis att programmet feltolkar tecknen, att manuell transkribering medför missar. Möjliga felkällor är också att korrekturläsningen och korrigeringar utförs på fel sätt. Är det äldre texter och stavningar så kan de innebära svårigheter att förstå texten och att fel således görs. Om fler personer deltar i arbetet kan korrigeringar utföras olika. I textuppmärkningen finns det risker att det blir felaktigt eller ofullständigt, m.m. (Digitaliseringsmanual s.69f.).

I samband med projektet *Öppna samlingar* korrekturlästes all publicerad text. Det ses som en integrerad manuell kontroll av textens trogenhet till originalet och förekomsten av felinställningar i OCR-programmet. KB ser det som önskvärt att texten korrekturläses av flera personer så att varje textavsnitt går igenom av minst två personer. Korrektoren får dessutom agera förebyggande för hela texten, inklusive textuppmärkning. Kontrollen studerar därmed använda databaser, bibliografisk metadata, e-texten specifika metadata samt e-texternas ID. Av presentationerna kontrolleras html-versionerna och utskriften, att de utförs korrekt. På så vis kontrolleras även stilmallarna och konverteringsverktygen från XML (Digitaliseringsmanual s.71f.).

När texter och bilder väl ligger tillgängliga på nätet ska kontroller göras av materialet och att sidorna fungerar som det är tänkt. Bland annat att gränssnittet passar avvikande material, att zoomen når alla detaljer. Försäkra om fungerande kopplingar mellan Libris bibliotekskatalog och verk, mellan faksimil och text, mellan faksimil och bläddring av text. I början görs kontroll av allt, men att det i ett senare skede ska räcka med stickprov (Digitaliseringsmanual s.73).

Kvalitetskontroll av fysisk utrustning görs huvudsakligen i samband med upphandlingen. Då kontrolleras att utrustningen uppfyller krav och givna specifikationer. När utrustningen är i drift sker löpande kontroller av bl.a. belysning och färgåtergivning. Kameror kontrolleras på morgonen när de aktiveras. Flatbäddsskanner kontrolleras ungefär två gånger per år, såvida akuta problem inte upptäcks. Dataskärmarna har en automatisk påminnelse att genomföra en kontroll var 4:e vecka (Informant B 2008a; Informant B 2008c).

5.2.5 Lagring av digitala original och kopior

KB håller för tillfället på att bygga upp en ny struktur för att hantera inkommande digitalt material, vilket inkluderar det egna digitaliserade materialet men även sådant som levereras av andra. Lagringssystemet ska kunna hantera allt digitalt material oavsett format⁵. Produktionen med tillhörande system kommer när det är klart vara skild från

⁵ Resultatavsnittet om lagring tar upp det som gäller digitaliserat material, dock stämmer det till viss del även för digitalt material generellt (inkluderat född-digitalt material).

systemen som lagrar och presenterar materialet. Processen blir följande: producent -> depositsystem -> magasin (repository) och arkiv. Producenten skapar ett paket (digitala materialet) enligt överenskomna kriterier som sedan överlämnas till depositsystemet. Där genomgår paketet automatiska tester för att se om det uppfyller det överenskomna kontraktet, t.ex. att all metadata ingår, rätt filformat. I denna del av processen kan objektet tillföras mer metadata, ex. egen identifikation. Om paketet klarar testet skickas det vidare, om inte - återvänder det till producenten (Informant B 2008a; Informant B 2008c).

Paket lämnas efter depositsystemet till två olika tjänster, arkiv och magasin. I arkivet hamnar det som ska långtidslagras och samtidigt omformas det till ett arkivpaket. Arkivpaketet består av metadata och dokumentet som ska arkiveras. Exempelvis kan det vara en ZIP-fil innehållande metadata i XML samt dokument i form av PDF-fil. Till magasinet (repository) lämnas paketet för att tillgängliggöras, från vilket bl.a. visnings-tjänster byggs upp (Informant B 2008a).

KB:s nuvarande lagring är att masterfilen lagras i bandarkivet, där det lagras på två band, den ena av dessa tas ur och lagras på annan geografisk plats. Accessfilen lagras endast på disk, men det finns en säkerhetskopia av den på samma plats. KB:s lagringsmålsättning är dock att få upp ett enhetligt lagringssystem för allt material med plats för en kopia på disk, en kopia på band på samma geografiska plats och tredje på band på annan plats (Informant B 2008a; Informant B 2008c; Informant B 2008g).

KB:s bedömning är att migrering mellan olika typer av fysiska medier är nödvändigt med 7-8 års mellanrum. Hittills har arkivet varit igång i 8 år utan att någon förflyttning gjorts. De har inte gjort någon migrering mellan filformat då ett formats livslängd är svårt att bedöma. Men de utesluter inte att emulering så fall kan vara aktuellt (Informant B 2008c).

När materialet läggs i arkivet så skapas kontrollsummor av innehållet, som gör att förändringar kan upptäckas. KB har också användarhantering av rättigheter för tillgång till olika typer material (Informant B 2008c).

5.2.6 Framvisning

På adressen <http://www.kb.se/samlingarna/digitala/> visas KB:s digitala samlingar. Det är sex övergripande verk och samlingar som presenteras. Genom att klicka på dem går man djupare ner i verkets hierarki eller kategorisering. Tillsammans med verket finns en kort text som beskriver dess innehåll och placerar det i ett historiskt sammanhang, vilket även kan gälla enskilda bilder. Det finns också möjlighet att söka i bibliotekskatalogen efter digitaliserade verk. Där återfinns KB:s alla projekt, exempelvis från *Öppna samlingar*, men även projekt de inte deltagit i (Kungl. biblioteket 2008b).

Codex Gigas har en egen sida, <http://www.kb.se/codex-gigas/Svensk-Codex-Gigas/> där verket i digital form presenteras. Objektets historia beskrivs och även hur verket tros ha tillverkats. Anekdoter och legender om *Codex Gigas* berättas, bl.a. om dess många namn. Då verket tros ha tillkommit i det medeltida Böhmen innehåller webbplatsen även redogörelsen för dess samlingar och om klostrens funktion och verksamhet. Innehållet i *Codex Gigas* presenteras och bibelböckerna beskrivs, bl.a. olika bibelöversätt-

ningar. Övriga texter beskrivs och från vilken historisk kontext de härstammar, och för texten "Besvärjelser" finns tillhörande transkribering och översättning.

Webbsidan redogör även för verkets fysiska form, med bl.a. omfång, material, typografi, bindning, m.m. En ordlista förklarar termer och begrepp, och Ortsnamn olika stavningar, och placerar dem på en karta. KB hänvisar slutligen till bibliografier och texter de själva använt för att beskriva verket.

Överst på webbplatsen ligger en sökruta som söker i de skrivna kommentarerna. Resultatet presenteras i en söklista med en länk och ett kort textutdrag. Hela verket kan betraktas och bläddras i. Boken delas upp i kategorier, med t.ex. bibelböckerna. Kända motiv som djävulen har sin egen kategori. När bilderna visas är hierarkin synlig överst i bilden. KB använder visningsprogrammet FSI Viewer som möjliggör zoomning av detaljer. Längst ner finns en länk för nedladdning av bilden, om än i lägre bildupplösning. Se bilaga 2 för exempel på webbvisningen.

Hela *Codex Gigas*-webbplatsen är på svenska men finns också översatt till engelska och tjeckiska.

6. Analys

Digitalisering är ett omfattande ämne och i en uppsats av denna övergripande karaktär finns ingen plats för detaljerad behandling av varje moment. Analysen fokuserar på uppsatsens två huvudfrågeställningar, och övriga diskussioner som hade varit värt att uppmärksamma nämns endast kortfattat, om ens över huvud taget. Som modell för massdigitalisering står den som utförs vid Nasjonalbiblioteket och kvalitativ digitalisering såsom den utförs vid Kungl. biblioteket.

6.1 Besvarande av frågeställning 1

I uppsatsen har jag arbetat utifrån två frågeställningar. Jag väljer att belysa en fråga i taget. Den första frågan löd:

- Hur skiljer digitaliseringsprocessen sig åt när den sker med mass- respektive kvalitativ digitalisering?

Frågan syftar till att granska hur digitaliseringen skiljer sig åt mellan de två metoderna, underförstått även hur de liknar varandra. Uppdelat på delprocesser redogör jag institutionernas arbete, därefter besvaras frågeställningen fördelat på kategorier.

Sammanfattning av NB respektive KB digitalisering.

Massdigitalisering enligt NB	Kvalitativ dig. enligt KB
Syfte	
Tillgängliggörande av material, även att bevara.	Bevarande av material, men också tillgängliggöra.
Urval	
Allt ska på sikt digitaliseras. Börjar med det äldsta. Särskilt prioriterade har förtur. Bl.a. har dokument om nordområdet digitaliserats som prioriterat.	Har än så länge ingen utvecklad urvalsordning. Främst digitaliseras material i dåligt skick, och intresse finns för studier eller som är aktuellt vid ex. särskilda händelser, ex. att <i>Codex Gigas</i> digitaliserades för Tjeckienbesöket. Huvudsakligen har bildmaterial digitaliserats.
Förberedelse inför bildfångst	
Om NB har böcker i flera exemplar än tre plockas ryggen bort och lösgör sidorna, för att automatiskt skannas, vilket beräknas ske med ca. 1/4 av samlingen. Konservator medverkar om det finns risk för skador på dokumentet.	Alla objekt som KB digitaliseras ska granskas av en konservator, som beslutar hur bildfångsten bäst utförs med tanke på att dokumentet inte ska ta skada. Konservatorn rengör materialet, och bl.a. plockar bort tejp och jämnar ut sidor.
Bildfångst	
Använder automatisk skanner för dokument som har demolerats. Övrigt material sker med manuell skanner, då en operatör vänder sidbladen.	I exemplet <i>Codex Gigas</i> skedde bildfångsten med kamera.

Bearbetning	
Skjer i bildfångstprocessen. Vid automatiska skanningen görs bearbetningar för varje blad, med syfte att underlätta framvisningen. Övrigt skannat material för att varje lojal mot originalet.	En obehandlad och en behandlad master sparas. Bearbetningen görs efter bildfångsten och ska återspegla originalet i färger och nyanser.
Metadata	
Använder METS och XML. Bibliografisk metadata hämtas från katalogposten. Strukturell metadata observeras när objektet OCR:as, med ex. innehållsförteckning, kapitel, sektioner, sidnummer. Teknisk/administrativ metadata som registreras är bl.a. produktionssätt och process.	Använder TEI och XML. Bibliografisk metadata hämtas från katalogposten. Strukturell metadata som registreras är bl.a. kapitel, stycken, radbrytning, sidnummer, sidhuvud. Teknisk/administrativ metadata anskaffas automatisk vid bildfångstutrustning, därefter manuella tillägg, ex. programvara, färgprofil, kontrollsumma.
Uppmärkning/OCR	
Skjer automatiskt, med viss subjektiv kontroll av strukturell metadata. Se ovan vad som registreras.	Skjer automatiskt och manuellt. Först provskannas ett par sidor för att se tecken den har svårt för och gör särlösningar av dessa. Text och struktur ska överensstämma med originalet. Uppmärkning kan ske på två nivåer. Första nivån: struktur (beskrivs ovan), andra nivån: sådan som är relevant för sammanhanget och innehållet, ex. ortnamn.
Kvalitetsbedömning	
Ingen kontroll av att bäst verkexemplar bildfångas. Automatiskt skannade dokument kontrolleras av operatörer under bildfångsten. Sidorna som färgkorrigeras kontrolleras med stickprov. Böcker som skannas manuellt kontrolleras osystematiskt, operatörerna förväntas upptäcka felen. OCR-texten kontrolleras inte. Strukturell metadata kontrolleras automatiskt och till viss del subjektivt. Kontrollen varierar delvis utifrån materialet och dess ålder. Kontroll görs även vid användning av NBdigital.	Manuell kontroll av alla moment: bildfiler, metadata, OCR-text och framvisningen. Bild och text ska återspegla originalet, och minsta tecken ska vara läsbart. Kontrolleras med verket bredvid. OCR-texten ska läsa av minst två personer. Presentationen kontrolleras för alla verk, att visningen är korrekt, senare förväntas stickprov räcka.
Lagring	
Ingår som generell del av datahanteringen. NB hanterar en masterbild, och lagrar den i tre exemplar: två på band, en på disk, två skilda områden. Räknar med att automatisk migrera med 2-3 års mellanrum. Accessbilden produceras ur mastern - "on-demand". Använder kontrollsummor.	Ingår som generell del av datahanteringen. Har två olika lagringslösningar, för masterbilder och accessbilder. På sikt är lösningen: en kopia på disk för tillhandahållande, en på band i samma geografiska område, en tredje på band, men annan plats. Beräknas migrera innehållet med 7-8 års mellanrum. Använder kontrollsummor.

Framvisning	
Plattstrukturerad: bara länk till verket, möjligen visas innehållsförteckning, ej kapitel. Vanligtvis inga beskrivande texter, endast vissa fotografier. Visar materialet i 2-3 olika kvalitetsnivåer. Sökning försiggår i OCR-texten och metadata. Visning sker utav bildreproduktionen.	Hierarkiskt strukturerad: grupperar innehåll i kategorier, verk, kapitel, avsnitt och vissa fall – enskilda sidor. Beskrivande texter till verken och, särskilda motiv, ex. djävulen i <i>Codex Gigas</i> . Visningsgränssnittet använder zoom. Ingen speciell sökning i OCR-texten. Nerladdningsbara lågupplösta bilder.

Utifrån analystabellen och resultatet av NB och KB och deras digitalisering framträder ett par skillnader i deras metoder. De mest framträdande skillnaderna som påverkar de enskilda digitaliseringsmomenten kan sammanfattas i kategorier.

Automatisering och manuellt arbete

Tydligast skillnad är att NB vill automatisera så långt det är möjligt över alla led. Manuella operationer ses som fördröjande processer. De manuella moment som finns är att objekten som ska digitaliseras hämtas upp till bildfångsten och i vilken ordning digitaliseringen sker. Om objekten anses kunna ta skada vid digitaliseringen så bedömer en konservator hur den ska utföras. Bildfångas objektet automatiskt så ska sidorna först lösgöras, sker det med manuell skanning måste sidbladen vändas av en operatör. När texterna ska OCR:as och strukturanalyseras så läggs de manuellt in i programmet docWorks. Övriga processer med urval, OCR och strukturbehandling, kvalitetsgranskning, inhämtning av metadata, lagring och publicering är automatiska. Det ska dock tilläggas att manuella processer krävs för att hålla igång utrustning och se till att allt fungerar.

KB:s arbete är övervägande manuellt. I motsättning till NB är de automatiska processerna få. De som finns är OCR-inläsningen med strukturanalys och att bildfångstutrustningen registrerar en viss andel data. Manuell hantering är urval, förberedelse, bildfångst, bearbetning, tillägg till metadata, textuppmärkning, kvalitetskontroll och framvisning.

Förberedelse och bearbetning av analoga och digitala originalen

I vilket mån förberedelserna och bearbetningarna görs skiljer mellan institutionerna. KB genomför noga undersökningar inför alla moment och av alla objekt: vilket objekt som ska digitaliseras när flera exemplar finns, hur bildfångsten bäst utförs med tanke på objektet, enskild rengöring av boksidorna, ett par sidor OCR:as för att se vilka tecken som är svåra att avläsa som därefter justeras. Bearbetningarna av det digitala materialet görs jämförande med det analoga verket för att de ska stämma överens i färgnyanser, text och struktur.

NB:s förberedelse är att objektet granskas utifrån vilken skanningsmetod de ska använda. Blir det med automatisk skanner så tas boken isär med hjälp av specialsaxar. Om inte så utförs det manuellt med bladvändning. NB bearbetar automatiskt det digitala materialet i bildfångstprocessen, utifrån vilken skanner som använts. Sker det med automatisk skanner så bearbetas bilden till god visningskvalité, är det med manuell skanner så bearbetas den för att likna originalet. OCR-texten struktureras automatiskt, textinnehållet bearbetas inte med särskild uppmärkning.

Kvalitetskontroll

NB har automatiska kontroller av allt digitaliserat material, undantag OCR-texten. Korrigeras bilden i samband med den automatiska skanningen, så görs stickprovskontroller av dessa. Vid manuell skanning övervakar operatörer processen, både manuella och automatiska operatörer och gör även osystematiska stickprov.

KB kontrollerar genomgående manuellt sina moment och digitala dokument för att se att det håller rätt kvalitets nivå. Speciellt nämns den digitala bilden och texten, metadata samt presentationen. Det digitala materialet kontrolleras med det analoga verket till hands för att textinnehåll, struktur och bildnyanser ska stämma överens.

Stora samlingar och enskilda dokument

Digitaliseringen KB gör sker med hänsyn till enskilda dokument. De tar beslut hur originaldokumentet bäst gynnas i och med den digitala produktionen. Verket granskas och korrigeras en sida i taget. KB:s andra uppmärkningsnivå fokuserar på textens innehåll, där viktiga avsnitt för sammanhanget uppmärksammas, vilket kräver detaljstudie av enskilda texter. När verket publiceras på webbplatsen skrivs tilläggstexter som beskriver dokumentets innehåll och vad det representerar och placerar innehållet i en kontext. Texterna kan som vid exemplet *Codex Gigas* ta upp sammanhanget i vilket verket skapades. Webbplatsen med *Codes Gigas* har dessutom översatts till engelska och tjeckiska.

NB använder samma enhetliga procedur vid digitaliseringen av böcker, detta oavsett dokumentstorlek eller innehållet som färger, text, tabeller. Digitaliseringen skiftar beroende på om det t.ex. är böcker, fotografier, tidningar. Det förekommer i vissa fall organisering i samlingar, som ex. dokumenten om nordområdet.

Resultat visar även på likheter, som exempelvis vid lagring. När KB fått upp sitt lagringssystem med kopior så har de samma lösning som NB. Däremot skiljer uppfattning om hur ofta materialet ska migrera.

Sammanfattningsvis uppvisar exemplen Nasjonalbiblioteket och Kungl. biblioteket och deras mass- respektive kvalitativ digitalisering ett par tydliga skillnader. Den tydligaste är att massdigitaliseringen strävar efter att automatisera alla delar av processen. Detta medför i sin tur att fokus på enskilda dokument undviks och istället används mer heltäckande lösningar för en större mängd samlingar. Att automatiska inställningar används medför att kvalitetsgranskningarna inte kan vara lika omfattande och noggranna. Fel som görs i den automatiska processen måste företrädesvis rättas till manuellt, något de vill undvika. Kvalitativ digitalisering skapar individuella lösningar för dokumenten, med förberedelse för vart och ett av dem. De kontrollerar delprocessen varje gång så att den blir lyckad. Vid framvisningen presenteras verken av texter om dess utformning men sätter verket även i en historisk kontext.

I kapitel 3.4 beskrevs vilka förfaranden forskare förknippar med massdigitalisering och kvalitativ digitalisering. Kännetecknen som Coyle ansåg beteckna massdigitalisering var att det sker i industriell skala, med enhetliga metoder för stora kvantiteter material. Arbetsprocessen baseras på automatik och manuella procedurer undviks. Enligt Coyle och Rieger kvalitetskontrolleras materialet sällan, och sker då med stickprov. Den digitala text som skapas är sällan tillrättalagd och bearbetad och strukturerad i samlingar. Om någon kontroll görs så är den maskinell. Kvalitativ digitalisering som det beskrivs av Dahlström och Hansson involverar djup textuppmärkning, kritisk bild- och textredi-

gering, och rik informationstilldelning vid urvalskriterier och en strategi tillämpad för enskilda dokument. Coyle skriver att urvalet ska vara noga övervägt, med syfte att skapa representativa kopior utav originalet. OCR-texten ska vara rik på uppmärkning så att den kan användas i många sammanhang. Utifrån forskarnas beskrivning överensstämmer Nasjonalbibliotekets massdigitalisering med många av kännetecknen, och Kungl. bibliotekets med vad som anges vara kvalitativ digitalisering. Båda i avseende på utförande med att massdigitaliseringen hos NB bl.a. sker i industriell skala och automatiskt och att kraven därmed sänkts. KB överensstämmer bl.a. med kvalitativ digitalisering och vilka mål och riktning de har på arbetet med att skapa en kopia trogen originalet.

6.2 Besvarande av frågeställning 2

Den andra frågeställningen löd:

- Hur skiljer de digitala dokumenten tekniskt och bildmässigt sig åt när digitaliseringsprocessen sker med mass- respektive kvalitativ digitalisering?

Frågan fokuserar på produkten som uppkommer av digitaliseringsmetoden – ”det digitala dokumentet”, i vilket jag innefattar dels den digitala bildfilen och dels OCR-texten. Hur dokumentet blir är självklart starkt sammankopplat med den första frågeställningen, och således har antydningar redan gjorts.

Sammanställning av teknisk och bildmässig information.

	Mass. dig enligt NB	Kval. dig. enligt KB
Syfte med det digitala materialet	Vill återskapa egenskaper på ett tillfredställande sätt, vidare betonas att innehållet ska bevaras.	Den digitala kopian ska ersätta användningen av originalverket och fungera som substitut, och ska kunna tillgodose samtliga behov som kan komma efterfrågas.
Filformat – masterbild	JPEG2000	TIFF
Bildfångstupplösning	400dpi/ppi	800ppi/dpi för dokument mindre än A5 400pp/dpi för dokument större än A5 300ppi/dpi för <i>Codex Gigas</i>
Färgdjup	24-bit	48-bit
Filstorlek – masterbild	20 Mb	650 Mb
Bearbetad masterbild	Ja, beskuren. Automatiskt skannade bilder - färgkorrigerade efter god visning. Manuellt skannade bilder – korrigerade efter originalet.	Nej/Ja. Har två masterbilder, ena obehandlade efter bildfångst, den andra korrigerad efter originalet.
Komprimerad	Ja, icke-förstörande	Nej
Filformat – accessbild	JPEG2000	Pyramidtiff
Filstorlek – accessbild	20 Mb	50 Mb

Bearbetad accessbild	Ja.	Ja. Skärpa.
Uppmärkning – struktur	Ja, automatiskt.	Ja, automatisk & manuell.
OCR-text – bearbetad	Nej.	Ja, manuellt för att likna originalet.
Textuppmärkning	Nej.	Ja, i nivå två.

Frågeställningen gällde vad som skiljer de digitala dokumenten åt när digitaliseringen sker med massdigitalisering eller kvalitativ digitalisering. Frågan skulle även kunna formuleras: vad är resultatet, vad konverteras från analogt till digitalt, vad kan de digitala bilderna användas till?

Utifrån avsnitt 4.2 och bildmässiga rekommendationer som där ges så uppfyller båda digitaliseringsmetoderna vad Lee ansåg vara minimikravet: 300 dpi skannerupplösningen och 24-bits färgdjup, även om han ansåg 600 dpi vara mest idealiskt. Besser ansåg att färgdjupet för den digitala mastern borde ligga på 36-bit. Om institutionernas inställningar jämförs så är KB:s färgdjup dubbelt så stort som NB:s. Det betyder att antal möjliga färger KB:s bild kan inneha är flera miljarder. Vid bildfångsten kan därför originalets alla färger fångas upp och bevaras i digitaliseringen, trots att människor endast uppfattar ett par miljoner av dem. NB:s färgdjup bevarar de flesta färger och är materialet monokromt bevaras troligen alla färger. Att KB använder högre upplösning på mindre objekt betyder att mindre detaljer urskiljs som kan missas i NB:s digitalisering. Det bör dock tilläggas att NB:s uppgifter gäller böcker, men då färgfotografier förekommer i böcker så finns en risk för att detaljer i dessa inte urskiljs. NB saknar i sin bildfångst anpassning till objektets storlek vilket medför att små och stora objekt blir behandlade och bearbetade på samma sätt.

NB och KB lagrar bilderna i två olika filformat, JPEG2000 och TIFF. Den senare är vanligt förekommande och rekommenderas som bevarande filformat. KB:s val av TIFF är alltså i bevarandenaspekter att betrakta som säkert val för de digitala bilderna. Eftersom formatet är inarbetat i många digitaliseringsprojekt är det troligt att stödet för TIFF kvarstår länge, även om ett nytt standardformat utvecklas. Troligen utvecklas då snabba konverteringsprogram mellan formaten. Av samma anledning finns det fog till oro för JPEG2000. Formatet är än så länge inte inarbetat och i jämförelse med TIFF är spridningen liten. Hur väl formatet står sig som bevarandeformat beror delvis på hur mycket det använts, om antalet är få riskerar formatet att hamna i skymundan. Att NB uppger att de utan informationsförlust kan konvertera tillbaka till TIFF, innebär dock att deras digitalisering inte kommer att vara förgäves, men att de då blir tvungna att lägga resurser på konverteringen, istället för på fortsatt digitalisering. Ifall JPEG2000 ökar i popularitet och användning så betyder det att NB:s djärva beslut varit rätt och att deras slutsatser besparat dem stora kostnader med datahantering och lagring. Tack vare NB:s filformatet och dess egenskap att ”on-demand” skapa en accessbild, medför det att de slipper hanteringen av flera filvarianter av samma bild, något KB behöver konfrontera. Formatet gör också att de kan vara flexibla och byta ut algoritmen som skapar accessbilden om en mer fördelaktig utvecklas.

Institutionerna får efter digitaliseringen bilder vars filstorlek skiljer, där NB:s är 20Mb och KB:s *Codex Gigas* 650Mb. KB:s storlek innebär att deras material tar stort lagringsutrymme och kräver mer datakraft vid hanteringen. Speciellt när KB sparar en bearbetad och obearbetad version. Den stora masterfilen underlättar troligen framvisning-

en då användaren kan zooma och se detaljer. Att KB sparar den obearbetade bilden gör det möjligt att vid behov återvända till filen med mest bildinformation. NB:s fil är liten i jämförelse och tar därmed mindre plats och är lättare att hantera i datasystemet. Att deras masterbild är kantbeskuren innebär att hela analoga originalet inte är representerat, även om endast kanterna saknas. Att filen är komprimerad i ickeförstörande medför att bilden kan bli obrukbar om en pixel förvanskas, något de själva tagit upp med att de ser risken som liten. Bilden som är master och digitalt original är bearbetad, vilket innebär att de inte kan gå tillbaka till en obearbetad version.

Den uppmärkta texten som generas av OCR-skanningen visar ett par betydande skillnader. Den viktigaste är att KB:s text är kontrollerad av flera personer för att stämma överens med originalet i både innehåll och utformning. Det betyder att den digitala texten kan ses som trogen återgivning av analoga verket. Texten kan därför fungera på egna ben, utan bilden. NB:s text innehar inte samma tillförlitlighet. Konsekvensen är att texten i både innehåll och strukturering är mindre flexibel. NB:s text kommer med nuvarande OCR-behandling att endast fungera som sökunderlag eftersom NB inte kan vara säkra på att texten är helt korrekt mot originalet. En användare som vill läsa in-skannat material måste därför alltid hänvisas till bildmaterialet.

Sammanfattningsvis så visar KB:s val av det digitala materialet på aspekter som försiktighet och säkerhet. Formatet är välprövat och deras bildfångstmetod skapar en representation som får med alla färger och nyanser, och bildfångstupplösningen når troligen de allra flesta detaljer. Den obearbetade mastern sparas så att de kan göra om beslut om de i framtiden upptäcker att fel gjorts. Att studera digitala kopior av analoga verk underlättas av inställningarna KB använder, då detaljer och färger är med i framställningen. KB måste tillskillnad från NB använda multipla bildversioner, som är väldigt stora och belastar datasystemet i större omfattning. NB:s beslut för materialet visar på risktagande och djärvhet när de använder relativt obeprövat filformat och lösningar där de inte kan gå tillbaka till ett icke-behandlat original av hela bildrepresentationen. Deras val av skannerinställningar hänger naturligtvis ihop med massdigitaliseringen och att de endast avser att vilja återskapa egenskaperna tillfredställande, men innebär även att en del bildinformation aldrig kommer fram i den digitala versionen. Valet av filformat gör att NB lätt kan anpassa bilden till olika visningsgränssnitt, med hjälp av att byta ut den använda algoritmen. De kan därför delvis vara flexibla med materialet och förändra accessbilden utifrån krav och behov.

7. Diskussion och slutsatser

Uppsatsens frågeställningar har varit vad skillnaden är om digitaliseringsprocessen är del av massdigitalisering respektive kvalitativ digitalisering, samt vad skillnaden blir på deras resultat – det digitala materialet. I uppsatsen har jag utgått ifrån Nasjonalbiblioteket (NB) i Norge och Kungl. biblioteket (KB) i Sverige. I detta kapitel diskuterar jag först uppsatsen och hur metoder och distinktioner fungerat, därefter tar jag i ett kritiskt perspektiv upp uppsatsens resultat och analys.

I uppsatsen har jag som metod använt dokumentskrifter och e-postkontakt med institutionerna. Om deras uppgifter är korrekta så stämmer det förhoppningsvis med hur de faktiskt går tillväga. Eftersom uppgifterna är deras egna finns en risk att de har beskrivit det optimalaste arbetet men att metoden i verkligheten inte alltid fungerar likadant.

Dessutom finns risken att jag missat eller misstolkat avsnitt av deras verksamhet, och läst in mer i texten än vad som egentligen varit textens avsikt.

Det är viktigt att påminna att institutioners digitaliseringsbeslut inte nödvändigtvis hänger samman med att de ämnar ha just massdigitalisering eller kvalitativ digitalisering, utan för att beslutet är det mest realistiska i förhållande till tid, utrymme, syfte, ekonomi, personal, m.m. Digitaliseringsarbeten kan därför omedvetet eller medvetet hämta inspiration ifrån andra metoder. Distinktionen mellan vad som är kvalitativ- och massdigitalisering kan därmed vara svår att dra, eftersom gränsdragningen i viss mån är konstruerad och inte helt passar faktiska förfaranden. Förhoppningsvis finns dock likheter mellan kvalitativ digitalisering så som den utförs av KB med andra liknande projekt, och NB:s massdigitaliseringsarbete med andra. Det finns också en föreliggande risk att jag i bearbetningen av resultatet fokuserat på att framhäva de typiska dragen som förknippas med metoderna och (omedvetet) missat de icke-typiska dragen. Det bör även upprepas att varken NB eller KB endast ägnar sig åt den typ av digitalisering som de i uppsatsen har sammankopplat med.

Ytterligare en aspekt av undersökningen är hur omfattade slutsatser man kan dra, och om de även är giltiga för andra digitaliseringsarbeten som betecknas mass- eller kvalitativ digitalisering. Uppsatsen har i många avseenden fokuserat digitaliseringsprocessen, dess utförande och vilka tekniska lösningar institutionerna valt, med exempelvis filformat, upplösning, hur mycket metadata som används. Hur applicerbart slutsatserna är anser jag skiljer mellan frågeställningarna. Den första frågan kan troligen användas på liknande arbeten, även om variationer i utförandet skulle förekomma. Den andra frågan är i större grad styrd av tekniska val institutionerna gjort, resultat kan därför vara mer begränsad i användning, speciellt angående filformatet som erbjuder olika möjligheter.

Resultatet och analysen bör betraktas utifrån vetenskapen om frågeformuleringarna och urvalet. Uppsatsens första frågeställning var att utifrån arbetet hos NB och KB undersöka skillnaden mellan digitaliseringsprocessen då den utförs i mass- respektive kvalitativ digitalisering. Analysen och resultatet visade samtidigt just de skillnader av mass- samt kvalitativa digitalisering som urvalet ifrån början valt att utgå ifrån. Jag bör därför i viss mån vara kritisk till min frågeställning och mitt urval. Men samtidigt åskådliggör resultatet i analystabellen skillnaderna i deras digitaliseringsprocess på ett mer konkret sätt, nämligen fördelat på delmomenten i arbetet, moment som kan skilja sig från fall till fall. Att den första frågeställningen och urvalet i de generella kategorierna ”bevisar” urvalspremissor som legat till grund för undersökningen borttar därmed inte betydelsen av att åskådliggöra digitaliseringsmomenten var för sig.

KB och deras kvalitativa digitalisering fokuserar på att det digitala materialet ska vara originalet troget så långt det är möjligt; bevarande och trogenhet mot originalet i både utseende och innehåll anses vara av stor betydelse. Det finns som jag ser det en föreliggande risk vid kvalitativ digitalisering att originalets innehåll överbearbetas, med högre kvalitet än vad som egentligen är nödvändig. För en del material är den höga upplösningen och färgkorrigeringen inte nödvändigt, och när det gäller användargrupper är det kanske inte ens relevant. KB digitaliserar inte allt i form av kvalitativ digitalisering, vilket medför att de måste ta beslut om vilka objekt som ska användas med vilken digitaliseringsmetod. För att KB ska kunna digitalisera all den mängd material de vill digitalisera så krävs det från deras sida mer bedömning av hur omfattande arbetet ska ske, för vart och ett av materialen – vad de ska prioritera. Det är då viktigt att materialet inte

överarbetas. KB:s säkerhetstänkande med bevarandekvalité borde möjligen bli något djärvare. Allt material behöver kanske inte återspegla originalet helt fullständigt med ex. färger. Jag menar dock inte att det mindre viktiga ska negligeras, men om resurser inte räcker för att digitalisera allt på mycket hög nivå där allt material ligger över för vad som krävs, så anser jag det kan vara värt att tänka över vad som är prioriterat, att fördela var arbetsenergin ska läggas.

Som en konsekvens av KB:s högkvalitativa material är att det underlättar materialets framtid och tillgängliggörande, det blir mer flexibelt och användbart, ex. mobiltelefoner, e-bok, att kopiera, m.m. En stor del av denna möjlighet beror på att allt material kvalitetsgranskas. Här ligger samtidigt en nackdel då materialet blir omfattande, med hantering av flera bildversioner, alla mycket stora. Den omfattande kontrollen av OCR-texten kan KB troligen inte förväntas använda i någon större omfattning, endast i speciella fall. Vilken säkert är vad de tänkt.

Kvalitativ digitalisering löser inte problemet med att det tar lång tid att utföra och att övrigt material fortsätter försämrats utan att ha en digital kopia. Vilket kan innebära att KB alltid ligger efter med att tillgängliggöra material digitalt, och befolkningen fortsätter att ha svårt att nå samlingarna. Det går ifrågasätta för vem det högkvalitativa materialet är relevant. En vanlig medborgare har troligtvis inte samma höga krav som KB på det digitala materialet, så att det måste vara i bevarandenivå. Kvalitativ digitalisering och KB:s digitalisering ligger troligtvis långt över mångas behov. I egenskap av nationalbibliotek så bör KB kunna delge objekt och material, inte bara för forskare utan också för allmänheten. Nuvarande digitalisering verkar i större grad vara för den först nämnda gruppen. Gemene man har troligen ingen kunskap om samlingarna KB hantlar, där digitalisering av dem skulle öppna samlingarna och nationalbiblioteket skulle på så sätt komma närmare vanliga personer.

Det är också anmärkningsvärt att KB hittills inte haft någon särskild plan för urvalsmaterial. Något som kan vittna om oförståelse till hur digitalisering av material kan användas, och att det saknats visioner för det. Att inte veta vad för material man i framtiden tänkt digitalisera kan ses som oprofessionellt. Därför är det bra att KB nu börjat utreda frågan.

Massdigitaliseringen såsom den görs av NB uppvisar ett tvetydligt förhållande mot originalet, och trogenheten ligger generellt på lägre nivå än vid KB:s kvalitativa digitalisering. Vid bildfångsten används två olika synsätt; automatiskt skannat material bearbetas för god visning, manuellt skannat efter originalet, och OCR-texten lämnas orörd. Det visar på olika uppfattningar mellan originalet och det digitala materialet - vad som är dess syfte. Där det ibland återspeglar originalet, ibland eftersträvas bättre framvisning. I NB:s digitalisering förefaller bildmaterialet vara viktigare än texten, även om en textbok digitaliseras, med tanke på att texten inte kontrolleras hur korrekt den är. Förståeligt med tanke på att textmaterial är svårare och mer tidsödande att hantera och rätta upp än bildmaterial. Även om NB främst digitaliserar för att tillgängliggöra och inte för att ha ett bevarandeexemplar, så anser jag ändå att de i egenskap av nationalbibliotek ska sätta högre kvalité på arbetet och resultatet, där främst text försummas och borde åtminstone kontrolleras. Om det innebär att digitaliseringen tar längre tid, så tror jag ändå de i längden tjänar på fler manuella kontroller, då tillförliten mellan det digitala och analoga materialet ökar. Alternativt att de istället försöker anpassa materialet helt till framvisningen.

Ett förhållande som togs upp i problemformuleringen är bildfångsten kontra originalverket och digitala originalet. Att ex. skanna ett textdokument i 800ppi/dpi med 48-bits färgdjup gör inte den digitala reproduktionen bättre, det ökar endast filstorleken. Inställningarna på bildfångsten kan därmed bero på vad som digitaliseras, är materialet endast text, text och bild i svart/vit och färg, eller är objektet litet eller stort? Alla dokument blir hos NB betraktade på samma sätt, stöpta i samma form. Konsekvens blir att bildfilerna ibland ligger på nivå av vad som krävs för att återge analoga dokumentet, ibland över och ibland under. Digitaliseringen representerar med andra ord originalobjektet olika bra. Om NB önskar minska ojämnheten men ändå ha trogenhet mellan digitalt och analogt så anser jag att de måste bestämma vad som är betydelsefullt att ha med i den digitala produktionen, och varierar bildfångsten utifrån det. Är allt innehåll viktigt, eller är bildkvalitén mer betydelsefull än texten? I en textbok är kanske texten viktigare än bildkvalitén, och då är det bättre att texten kontrolleras mer än bilden.

För att veta vad i ett objekt som är viktigt bör det tänka syftet och användningen redogöras. Om NB digitaliserar för att vilja tillgängliggöra samlingar så borde de också satsa på att få det tillgängligt, flexibelt och användbart. Att bara tillgänggöra en bildrepresentation av boksidan och möjligheten att söka i texten tror jag i längden kan visa sig vara en väldigt kortsiktig tillgänglighet. När mer av världens analoga material blir digitalt så är det troligt att anta att människors krav på materialet de vill åt ökar, både på dess kvalitet men också hur de kan interagera med materialet. I nuläget uppvisar NB:s material vissa tveksamheter, och kvalitén på det digitaliserade material skiftar avsevärt både i bild och text. Framvisningen är främst fokuserad på förmedling med begränsade möjligheter till dialog och hänvisningar till externa informationsplattformar.

Frågan är hur mycket som blir kvar av massdigitaliseringen om NB skulle praktisera fler manuella moment som kontroller och övervägningar. ”Snabbhet” som är ett av nyckelorden vid massdigitalisering minskas troligen. Men när tempot blir långsammare så får de i gengällt material som är något mer kontrollerat och flexiblere för olika användningsområden. Användare kommer därmed också kunna lite mer på materialets autenticitet.

En viktig fråga som tidigare berörts är om nationalbibliotek ska massdigitalisera, speciellt om det innebär att tillförlitligheten mellan det analoga och digitala materialet kan ifrågasättas. Borde inte ett visst mått av omdöme finnas om tillräcklig kvalitet och långsiktighet? Jag tycker också att det går att ifrågasätta hur mycket vinning det är i att digitalisera allt. Hur stor blir förtjänsten av att digitalisera ett verk, som kanske redan finns tillgängligt på de flesta bibliotek? Eller att digitalisera böcker vars innehåll av rättighetsskäl inte får visas? Ytterligare en fråga är om allt material är intressant att digitalt ta del av, ex. att läsa en skönlitterär bok på nätet? Det är frågor jag anser värda att beakta, speciellt vid massdigitalisering, när stora samlingar digitaliseras. Å andra sidan kan man resonera att mycket material i vart fall blir nåbart även om man befinner sig på andra sidan jordklotet.

Det föreligger en stor skillnad mellan KB och NB och hur stort ekonomiskt stöd de får. NB har stort stöd av politiker ända upp till regeringen, som bidrar ekonomiskt och har politiskt intresse av deras arbete, vilket troligen varit nödvändigt för NB:s möjlighet till massdigitalisering. En bidragande orsak kan vara NB:s entusiasm och vilja till stora satsningar, och att det har imponerat på den norska regeringen. Ett eventuellt regerings-

skifte skulle kunna påverka detta och att de inte ser samma betydelse av NB:s digitaliseringsarbete och att den ekonomiska uppbackningen då minskar. Vilka möjligheter får NB då med att ”digitalisera allt”?

KB har i sina skrifter ett flertal gånger påpekat bristen på politiskt intresse och ekonomiskt bidrag. Antagligen är det många faktorer som spelar in, men en möjlighet kan vara att KB till stor del ägnat sig åt kvalitativ digitalisering och att regeringen inte ser så stor värde av det. KB behöver troligen visa på större visioner med sitt digitaliseringsarbete, vad de vill uppnå, och då något som i större grad kan attrahera och komma den vanlige medborgaren till del.

Slutsatser som kan dras av uppsatsen om kvalitativ digitalisering och massdigitalisering är att metoderna skiljer sig åt och att metoden har stor inverkan på hur resultatet blir. Massdigitalisering går snabbt, men innebär att resultatet inte kan sägas vara tillförlitligt och överensstämmande mellan originalet och det digitala materialet, vilket i arbete också varit av underordnad betydelse. Massdigitalisering bör därför användas när kraven på trogenhet och bevarande av originalet är mindre prioriterat. Däremot kan det vara nödvändigt att öka flexibiliteten i materialet. JPEG2000 är i grunden ett flexibelt filformat, men bara för bildrepresentation. Större fokus på den digitala texten och att den blir mer lik originalet skulle innebära fler möjligheter för texten.

Kvalitativ digitalisering är tidskrävande i deras vilja om absolut trogenhet till originalet, och bör därför endast användas för objekt av stor betydelse, ex. unika verk eller material där trogen färgåtergivning och hög bildkvalité är nödvändig. Hur uppdelningen mellan vad som är aktuellt för kvalitativ digitalisering och vad som inte är det, måste däremot då utredas i större grad än vad som i dag görs. Då materialet efter digitaliseringen blir flexibelt betyder det att det kan fungera i många sammanhang och användningsområden.

Båda metoderna innebär fördelar respektive nackdelar, och ingen av metoderna bör användas genomgående för alla objekt. Som uppsatsen visar fyller de olika funktioner och bör så vara. Däremot behöver båda metoderna utvecklas och bli bättre i deras utförande och syfte. I massdigitalisering borde en större ansträngning göras för att få materialet mer flexibelt, vilket skulle för text kunna ske med större trogenhet mot originaltexten, med ex. bättre avläsningsalgoritmer och fler manuella kontroller. De digitala bilderna skulle kunna anpassas mer för webbvisningen, något som är möjligt med JPEG2000 utan att den ursprungliga digitala bilden förändras. I kvalitativ digitalisering blir materialet av mycket hög kvalité, men det bör ifrågasättas hur mycket kvalité som egentligen är nödvändig, dels utifrån vad användarna egentligen har behov av och dels vad som egentligen ska accepteras som tillräcklig kvalité för bevarande av analoga objekt.

Vad tekniken presterar och vilket material som ska digitaliseras fortsätter vara viktigt för hur digitaliseringen utförs. Institutioner behöver troligen använda sig av både kvalitativ digitalisering och massdigitalisering, även om den sistnämnda behöver förbättras för att på ett effektivt sätt uppfylla kraven som kan tänkas krävas av materialet.

8. Sammanfattning

Digitalisering är i dag ansedd av många som betydelsefullt, då världen blir allt mer digital och olika digitaliseringsmetoder och moment utvecklas. Det har de senaste åren startats många digitaliseringsarbeten, som utgår ifrån skilda syften och förutsättningar. Två av dessa är massdigitalisering och kvalitativ digitalisering. Det är metoder som skiljer sig i utförande och hur dess resultat blir. Massdigitalisering har blivit ansett som en metod där material snabbt digitaliseras för att tillgängliggöras på internet, men att dess kvalitet kan ifrågasättas. Kvalitativ digitalisering uppfattas som en metod där hög kvalitet och bevarandenivå prioriteras på resultatet, men att arbetet därmed tar längre tid.

Syftet med denna magisteruppsats är att undersöka vilka skillnaderna blir i digitaliseringsprocessen när den sker i form av massdigitalisering och kvalitativ digitalisering, samt hur resultatet – det digitala materialet blir.

För att undersöka syftet och uppnå svar har två frågeställningar formulerats:

- Hur skiljer digitaliseringsprocessen sig åt när den sker med mass- respektive kvalitativ digitalisering?
- Hur skiljer de digitala dokumenten tekniskt och bildmässigt sig åt när digitaliseringsprocessen sker med mass- respektive kvalitativ digitalisering?

För att kunna besvara frågorna så har jag utgått ifrån två verksamma digitaliseringsprojekt: massdigitaliseringen vid Nasjonalbiblioteket (NB) i Norge och kvalitativ digitalisering vid Kungl. biblioteket (KB) i Sverige. Det ska dock tilläggas att institutionerna använder fler digitaliseringsmetoder än bara dem.

Metoden som används för att få fram information bygger främst på dokument och texter, både interna och officiella. Då digitaliseringsprojekten hela tiden utvecklas och förändras så har det inneburit att vissa dokument är mindre aktuella än andra och att alla moment inte berörs. För att få klarhet om det aktuella arbetet så har e-postkontakt förts med institutionerna, en från NB och två ifrån KB.

Resultatet presenteras uppdelat på institution, men samlas även kortfattat i analysstabeller. Redogörandet för första frågeställningen listar i analysstabellen institutionernas arbeten moment för moment i digitaliseringsprocessen. Vidare åskådliggörs skillnaderna tydligare med att indela i fyra kategorier. Kategorierna och deras resultat är:

Automatisering och manuellt arbete: Massdigitalisering automatiserar så långt det är möjligt alla moment, och manuellt arbete hålls till absolut nödvändigaste. Kvalitativ digitalisering har mycket få automatiska moment, de flesta är manuella.

Förberedelse och bearbetning av analoga och digitala originalen: Kvalitativ digitalisering gör noga förberedningar för varje moment och resultat av digitaliseringen. Vid massdigitalisering görs dem främst om det innebär att resterande digitalisering går snabbare.

Kvalitetskontroll: Massdigitalisering förlitar sig på automatiska kontroller av kvalitén, med systematiska och osystematiska stickprov. OCR-texten kontrolleras dock inte. Kvalitativ digitalisering kontrollerar manuellt alla moment och resultat, där det digitala materialet ska stämma överens med analoga verket i textinnehåll, struktur och bildnyanser.

Stora samlingar och enskilda dokument: Kvalitativ digitalisering tar hänsyn till enskilda dokument och kan bestämma hur digitaliseringen bäst görs för att gynna det, med bl.a. särlösningar. Massdigitaliseringen använder företrädesvis samma lösning för många dokument, utan att hänsyn tas till objektstorleken eller innehåll.

Resultatet av den andra frågeställningen sammanfattas med att det digitala materialet som skapas vid kvalitativ digitalisering visar tecken på försiktighet och säkerhet, då digitaliseringen använder filformatet TIFF, format vanligt för bevarande. Deras bildfångstmetod skapar en representation som får med alla färger och nyanser, och bildfångstens inställningar når troligen alla detaljer. Obearbetade mastern sparas, vilket underlättare i fall fel upptäcks på den bearbetade mastern, och de vill göra om.

Massdigitaliseringen visar på mer risktagande och djärvhet då de som bevarandeformat använder JPEG2000, som är relativt obeprövat. Deras ursprungliga masterbild är redan bearbetad, med bl.a. ickeförstörande komprimering, beskärning av kanter och bildjusteringar, och de kan därför inte återvända till en första obearbetad version som innehåller mest bildinformation. Vidare så finns en föreliggande risk om att de i bildfångsten inte fångar upp alla färgnyanser och mindre detaljer i bildmaterialet.

I avslutande kapitlet, diskussion och slutsatser, konstateras att massdigitalisering och kvalitativ digitalisering överensstämmer väl med vad forskare sagt om respektive digitaliseringsprocess. Vidare diskuteras metodernas förhållande mellan digitala materialet och originalet. Då resultatet i studien visade att massdigitalisering hade vid bildfångsten två förhållningssätt beroende på vilken skanner som användes, antingen skulle det digitala materialet stämma överens med originalet eller bearbetas till bra framvisning, och att OCR-texten helt lämnas som den skannades. Uppsatsresultat visar också att NB:s massdigitalisering inte lägger mycket energi på texten, med t.ex. kontroller, även om mycket textmaterial digitaliseras, vilket antas vara för att det är tidsödande att få det överensstamma med originalet. Beslutet ifrågasätta med tankar om att ett nationalbibliotek borde ha högre krav på kvalitet i arbetet och resultatet. KB:s kvalitativa digitalisering kritiserar då det tenderar rikta sig mer till forskare än allmänheten. Det digitala materialet är av mycket hög kvalitet, vilket en vanlig medborgare troligen inte kräver och att de har lika stor rätt att ta del av objekten som nationalbiblioteket innehar.

Diskussionen lyfter också upp tanken om vad som är viktigt att få med från det analoga objektet i den digitala kopian. Och att man vid digitaliseringen bör reflektera dels om syftet men också användningen. Där återigen NB:s OCR-text står som exempel och att en kontrollerad text kan ha fler användningsområden än bara sökunderlag. Att material då tillgängliggörs bättre och blir flexiblare. Något som konstateras att KB:s material är.

Avslutningsvis konstateras att material från en kvalitativ digitaliseringsprocess är i bra bevarandekvalité men att processen är tidsödande, och att de därmed ständigt kommer att ligga efter med att tillgängliggöra material. Att massdigitalisering möjligen borde höja sina krav på trogenhet mot original, men att mycket av snabbheten då kan gå förlorad, men att det i längden troligen är värt det p.g.a. att de då får flexiblare material.

Källförteckning

Publicerade källor

Acharya, Tinku (2005). *JPEG2000 standard for image compression: Concepts, algorithms and VLSI architectures*. Hoboken: John Wiley & Sons.

Anderson, Cokie G. (2006). *Ethical decision making for digital libraries*. Oxford: Chandos Pub.

Andersson, Therese & Nilsson, Ann-Katrin (2006). *Digitalisering av bilder vid två museer*. Magisteruppsats BHS 2006:107. Borås: Högskolan i Borås.

Bakken, Frode (2006-10-19). *Nasjonalbiblioteket: Mange bøker snart på Nett*. Norsk biblioteksforening. <http://www.norskbibliotekforening.no/article.php?id=1445> [2009-03-05]

Banks, Paul N. & Pilette, Roberta, red. (2000). *Preservation: Issues and planning*. Chicago: American Library Association.

Besser, Howard (2003). *Introduction to imaging*. Rev. ed. Los Angeles: Getty Pub.

Codex Gigas. <http://www.kb.se/samlingarna/digitala/codex-gigas> [2009-03-05]

Coyle, Karen (2006). Mass digitization of books. *Journal of Academic Librarianship*, vol. 32, nr. 6, S. 641-645.

Dahlström, Mats & Hansson, Joacim (2008). "On the relation between qualitative digitization and library institutional identity." In *Proceedings of the International Society for Knowledge Organization 12*. (Advances in knowledge organization ; Vol. 13). S. 112-118.

Deegan, Marilyn, & Tanner, Simon (2002). *Digital Futures: Strategies for the information age*. London: Library Association Pub.

Deegan, Marilyn, & Tanner, Simon (2004). Conversion of primary sources. Ingår i Schreibman, Susan & Siemans, Ray & Unsworth, John, red. *A companion to digital humanities*. Oxford: Blackwell. S. 488-504

de Stefano, Paula (2000). Digitization for preservation and access. Ingår i Banks, Paul N. & Pilette, Roberta, red. *Preservation: Issues and planning*. Chicago: American Library Association. S. 307-322.

de Stefano, Paula (2002). Selection for digital conversion. Ingår i Kenney, Anne R. & Rieger, Oya Y., red. *Moving theory into practice: Digital imaging for libraries and archives*. Mountain View, Calif.: Research Libraries Group. S. 11-23.

DeWitt, Donald L., red. (1998). *Going digital: Strategies for access, preservation, and conversion of collections to a digital format*. New York: The Haworth Press.

Emanuelsson, Charlotte (2006). *Digitalisering av kulturarvet: En studie av digitalisering vid två museer*. Magisteruppsats BHS 2006:65. Borås: Högskolan i Borås.

Erway, Ricky L. (1998). The technology context. Ingår i DeWitt, Donald L., red. *Going digital: strategies for access, preservation, and conversion of collections to a digital format*. New York: The Haworth Press. S. 161-168.

Flemming, Gösta (2007). Ett djävulskt uppdrag. *F: Fotografisk tidskrift*, årg. 119, nr 1, S. 30-32.

Fotoguiden. *Ordlistan – CMYK*. <http://www.fotoguiden.se/ordlista/cmyk-4.html> [2009-03-05]

Google. *Google biblioteksprojekt – Ett utvidgat kortregister över världen böcker*. <http://books.google.com/intl/sv/googlebooks/library.html> [2009-03-05]

Graham, Peter S. (1998). Long-term intellectual preservation. Ingår i DeWitt, Donald L., red. *Going digital: strategies for access, preservation, and conversion of collections to a digital format*. New York: The Haworth Press. S. 81-98.

Gram, Magdalena & Kjellman (2000). *Plattform för bilddatabaser*. Stockholm: Kungl. biblioteket. (Kungl. biblioteket, Rapport, 27). http://www.kb.se/Dokument/Om/publikationer/rapport_bilddatabaser.pdf [2009-03-05]

Grave, Tonje, red. (2007a). *NB21*. Nr. 2. Oslo: Rolf Ottesen trykkeri. Även tillgänglig: <http://www.nb.no/pressebilder/NB21.pdf> [2009-03-05]

Grave, Tonje (2007b). Hvis det ikke finnes på nett Ingår i Grave, Tonje, red. (2007). *NB21*. Nr. 2. Oslo: Rolf Ottesen trykkeri. S. 3.

Grave, Tonje (2008). Banebrytende avtale om digital avlevering og bevaring http://www.nb.no/aktuelt/banebrytende_avtale_om_digital_avlevering_og_bevaring [2009-03-05]

Hart, Michael (1992). Gutenberg: The history and philosophy of Project Gutenberg. http://www.gutenberg.org/wiki/Gutenberg:The_History_and_Philosophy_of_Project_Gutenberg_by_Michael_Hart [2009-03-05]

Hirtle, Peter B. (2000). Image management system and web delivery. Ingår i Kenney, Anne R. & Rieger, Oya Y., red. *Moving theory into practice: Digital imaging for libraries and archives*. Mountain View, Calif.: Research Libraries Group. S. 119-134.

Hughes, Lorna M. (2004). *Digitizing collections: Strategic issues for the information manager*. London: Facet Pub.

Lagoze, Carl & Payette, Sandra (2000). Metadata: Principles, practices, and challenges. Ingår i Kenney, Anne R. & Rieger, Oya Y., red. *Moving theory into practice: Digital*

imaging for libraries and archives. Mountain View, Calif.: Research Libraries Group. S. 84-100.

Lee, Stuart D. (2001). *Digital imaging: A practical handbook*. New York: Neal-Schuman Pub.

Lervik, John M. & Brygfjeld, Svein Arne (2006). Search engine technology applied in digital libraries. ERCIM News. Nr. 66, juli 2006. S. 18-19. Även tillgänglig: http://www.ercim.org/publication/Ercim_News/enw66/EN66.pdf [2009-03-05]

LOCKSS. Lots Of Copies Keep Stuff Safe. <http://www.lockss.org/lockss/Home> [2009-03-05]

Jones, Maggie & Beagrie, Neil (2001). *Preservation management of digital materials: A handbook*. London: The British Library.

KB – ett nav i kunskapssamhället (2004). Statens offentliga utredningar (SOU), 2003:129). <http://www.sweden.gov.se/sb/d/108/a/669> [2009-03-05]

Kenney, Anne R. (2000). Digital benchmarking for conversion and access. Ingår i Kenney, Anne R. & Rieger, Oya Y., red. *Moving theory into practice: Digital imaging for libraries and archives*. Mountain View, Calif.: Research Libraries Group. S. 24-60.

Kenney, Anne R. & Rieger, Oya Y., red. (2000a). *Moving theory into practice: Digital imaging for libraries and archives*. Mountain View, Calif.: Research Libraries Group.

Kenney, Anne R. & Rieger, Oya Y. (2000b). Introduction: Moving theory into practice. Ingår i Kenney, Anne R. & Rieger, Oya Y., red. *Moving theory into practice: Digital imaging for libraries and archives*. Mountain View, Calif.: Research Libraries Group. S. 1-10.

Kungl. biblioteket (2008a-10-21). *Upptäck KB:s digitala magasin*. <http://www.kb.se/aktuellt/nyheter/2008/Det-digitala-magasinet-oppet/> [2009-03-05]

Kungl. biblioteket (2008b-12-02). *Kvalitet och standarder. Bilddatabaser och digitalisering – plattform för ABM-samverkan*. http://abm.kb.se/akt4cd/dok_formattyper.htm [2009-03-05]

Kungl. biblioteket (2008c-12-16). *8,3 miljoner extra för digitalisering*. <http://kb.se/aktuellt/nyheter/2008/83-miljoner-extra-for-digitalisering/> [2009-03-05]

Manuels Web. *Convert inches & centimeters*. http://www.manuelsweb.com/in_cm.htm [2009-03-05]

Metadata Engine Project (METAe). <http://meta-e.aib.uni-linz.ac.at> [2009-03-05]

Myrvang, Merethe (2006-03-29). *Nå er vi i gang!* http://www.nb.no/aktuelt/naa_er_vi_i_gang [2009-03-05]

Nasjonalbiblioteket (2007). *Digitalisering av bøker i NB: Metodikk og erfaringer*. http://www.nb.no/content/download/2325/18195/version/1/file/bokdigitalisering_sep07.pdf [2009-03-05]

Nasjonalbiblioteket, (2008a-12-02). *Digital bevaring / bevaring / kompetansesenter*. http://www.nb.no/fag/kompetansesenter/bevaring/digital_bevaring [2009-03-05]

Nasjonalbiblioteket, (2008b-12-02). *Søk i NB eller Om Nasjonalbibliotekets søketjeneste*. <http://www.nb.no/sok/about.jsf> [2009-03-05]

Official website of the Czech Republic (2007-11-01). *Enormous interest in the Devil's Bible!*. <http://www.czech.cz/en/news/culture/enormous-interest-in-the-devils-bible/> [2008-12-02]

Olstad, Bjørn & Seres, Silvija (2005). "What is Contextual Search?". *KMWorld*. November/December. S.10-11. Även tillgänglig: [http://www.fastsearch.com/What is Contextual Search_sbqQ5_nruLv.pdf](http://www.fastsearch.com/What%20is%20Contextual%20Search_sbqQ5_nruLv.pdf).file [2009-03-05]

Persson, Catrin & Tångemar, Annevie (2006). *Varför digitalisera?: En studie av tillkomsten av Kungl. Bibliotekets digitaliserade samlingar*. Magisteruppsats BHS 2006:119. Borås: Högskolan i Borås.

Price-Wilkin, John (2000). Access to digital image collections: System building and image processing. Ingår i Kenney, Anne R. & Rieger, Oya Y., red. *Moving theory into practice: Digital imaging for libraries and archives*. Mountain View, Calif.: Research Libraries Group. S. 101-118.

Prismas IT-ordbok. *Magnet-optiska skivor*. <http://www.pagina.se/itord/default.asp?iD=2206> [2009-03-05]

Ray, Erik T. (2003). *Learning XML*. 2. uppl. Sebastopol, CA: O'Reilly Media.

Rieger, Oya Y. (2000a). Establishing a quality control program. Ingår i Kenney, Anne R. & Rieger, Oya Y., red. *Moving theory into practice: Digital imaging for libraries and archives*. Mountain View, Calif.: Research Libraries Group. S. 61-83.

Rieger, Oya Y. (2000b). Project to programs: Developing a digital preservation policy. Ingår i Kenney, Anne R. & Rieger, Oya Y., red. *Moving theory into practice: Digital imaging for libraries and archives*. Mountain View, Calif.: Research Libraries Group. S. 135-152.

Rieger, Oya Y. (2008). Preservation in the age of large-scale digitization: A white paper. Washington: Council on Library and Information Resources. <http://www.clir.org/pubs/reports/pub141/pub141.pdf> [2009-03-05]

Renear, Allen H. (2004). Text encoding. Ingår i Schreibman, Susan & Siemens, Ray & Unsworth, John, red. *A companion to digital humanities*. Oxford: Blackwell. S. 218-239

Scherman, Anne., red. (2005). *DIGSAM: digitalisering och dess samordning inom Kungl. biblioteket*. Stockholm: Kungl. biblioteket. (Kungl. biblioteket, Rapport, 28). <http://www.kb.se/Dokument/Om/projekt/digsam.pdf> [2009-03-05]

Schreibman, Susan & Siemans, Ray & Unsworth, John, red. (2004) *A companion to digital humanities*. Oxford: Blackwell. Även tillgänglig: <http://www.digitalhumanities.org/companion> [2009-03-05]

Shape, Robert (2007). Digital Preservation: Solving archive challenges. *Research Information*, June/July 2007. http://www.researchinformation.info/features/feature.php?feature_id=134 [2009-03-05]

Skarstein, Vigdis Moe (2006). Europas første med alt på data. *Aftenposten*, 2006-03-29. <http://www.aftenposten.no/meninger/kronikker/article1261628.ece> [2009-03-05]

SOU 2003:129 (2004) Se: *KB ett nav i kunskapsamhället*.

Støre, Jonas Gahr (2007). Nordområdenes betydning. Ingår i Grave, Tonje, red. (2007). *NB21*. Nr. 2. Oslo: Rolf Ottesen trykkeri. S. 83-84

Svärd, Anna (2006). *Google digitaliserar bibliotekssamlingar: En analys av hur biblioteksvärlden reagerar på Google Book Search*. Magisteruppsats BHS 2006:78. Borås: Högskolan i Borås.

Taylor, Arlene G. (2004). *The organization of information*. 2. uppl. Westport, Con.: Libraries Unlimited.

Wikipedia. *Checksum*. <http://en.wikipedia.org/wiki/Checksum> [2009-03-05]

Wikipedia. *Color Depth*. http://en.wikipedia.org/wiki/Color_depth [2009-03-05]

Wikipedia. *Pixel*. <http://sv.wikipedia.org/wiki/Fil:Pixel-example.png> [2009-03-05]

Wikipedia. *RGB Color model*. http://en.wikipedia.org/wiki/RGB_color_model [2009-03-05]

Wikipedia. *XML Schema (W3C)*. [http://en.wikipedia.org/wiki/XML_Schema_\(W3C\)](http://en.wikipedia.org/wiki/XML_Schema_(W3C)) [2009-03-05]

Willett, Perry (2004). Electronic texts: audiences and purposes. Ingår i Schreibman, Susan & Siemans, Ray & Unsworth, John, red. *A companion to digital humanities*. Oxford: Blackwell. S. 240-253.

Witten, Ian H. & Bainbridge, David (2003). *How to build a digital library*. San Francisco: Morgan Kaufmann pub.

Opublicerade källor

Digitaliseringsmanual, v2. Kommentar: sammanfattningen är undertecknad Viktoria Enmark den 10 april 2007. (Finns i uppsatsförfattarens ägo).

Kungl. biblioteket (2008c). *Metadataprofil för digitaliserade dagstidningar – enstaka nummer*. ver. 2008-10-15. (Finns i uppsatsförfattarens ägo).

E-postfrågor

Informant A, 2008a-08-01 (Nasjonalbiblioteket)

Informant A, 2008b-09-22 (Nasjonalbiblioteket)

Informant A, 2008c-10-11 (Nasjonalbiblioteket)

Informant B, 2008a-10-15 (Kungl. biblioteket)

Informant B, 2008b-10-15 (Kungl. biblioteket)

Informant B, 2008c-10-22 (Kungl. biblioteket)

Informant B, 2008d-11-05 (Kungl. biblioteket)

Informant B, 2008e-11-12 (Kungl. biblioteket)

Informant B, 2008f-11-13 (Kungl. biblioteket)

Informant B, 2008g-11-19 (Kungl. biblioteket)

Informant C, 2008a-11-14 (Kungl. biblioteket)

Bilaga 1

Träfflistan vid sökning på "noreg" i NBdigital.

The screenshot shows the NBdigital search interface. At the top, there is a navigation bar with the logo 'nb.no' and the text 'Søk i Nasjonalbiblioteket'. Below the navigation bar, there are links for 'Forside', 'Om Nasjonalbiblioteket', 'Fag', 'Opplevelse', 'Kontakt', 'Logg inn', and 'English'. On the left side, there is a sidebar with a menu for 'Digitalt innhold' and 'Materialtyper'. The 'Materialtyper' section is expanded, showing a list of categories with their respective counts: Artikler (293), Aviser (424), Bøker (699), Film (1), Kart (4), Musikk (20), Nettsider (53), Noter (85), Tidsskrift (36), and ukjent (4). Below the sidebar, there are two blue buttons: 'Vil du ha hjelp?' and 'Har du kommentarer til våre nettsider?'. The main content area shows the search results for 'noreg'. At the top of the main content area, there is a search bar with the text 'noreg' and a 'Søk' button. Below the search bar, there are links for 'Søketips' and 'Om søketjenesten', and a 'Nullstill søket' button. To the right of the search bar, there are three checkboxes: 'Digitalt' (checked), 'Ikke digitalt', and 'Nettsider'. Below the search bar, there is a section for 'Du søkte på: noreg'. The 'Resultat:' section shows the search results. At the top right of the results section, there is a summary: 'Resultat: 1 - 10 av 1619 treff' and a dropdown menu set to '10' with a 'Neste->' link. The first result is a book: 'Bøker - Den første norske loge av Odd Fellow ordenen (I.O.O.F.), loge Noreg/Norvegia 100 år : 1898-26. april-1998, Norvegia : beretningen om logens første 100 år'. Below the title, there is a description: 'Omslagstittel: Norvegia : beretningen om logens første 100 år'. The second result is a note: 'Noter - Aukrust-songar. Fjell-Noreg; arr., Fjell Noreg, Norwegian mountain poem, Aukrust-songar. Fjell-Noreg, 3 Aukrust-songar. Fjell-Noreg, Tre Aukrust-songar. Fjell-Noreg'. Below the title, there is a description: 'For janitsjarkorps. - Originalens tittel: 3 Aukrust-songar. Fjell-Noreg. - Vanskelighetsgrad: 3. - Platenummer: N.M.O.9977 av: Olsen, Sparre; Hurum, Helge'. The third result is a book: 'Bøker - Ferietur til sola sitt land Italia, Hovding-garden i Øvre Vats, Forklaring til katekisma etter Martin Luther, Noreg Budapest i høvet 40-års jubileet til Nordea radio, Strand frå gamle dagar, Blegrebygdens mandforening, Hovdinggarden Eike i Øvre Vats, Hovinggarden Eike i Øvre Vats, Noreg-Budapest i høvet 40-års jubileet for Norea, Blegrebygdens Mandforening'. Below the title, there is a description: '...ROM-spiller Tittel fra tittel skjerm bildet Innholder også: Noreg-Budapest i høvet 40-års jubileet for Norea radio ; Blegrebygdens...ROM-spiller Tittel fra tittel skjerm bildet Innholder også: Noreg-Budapest i høvet 40-års jubileet for Norea radio ; Blegrebygdens... av: Landa, Jørgen'. The fourth result is a book: 'Bøker - Stat[t]haldarinstitusjonen i Noreg 1722-1739, Statthaldarinstitusjonen i Noreg 1722-1739, Stathaldarinstitusjonen i Noreg 1722-1739'. Below the title, there is a description: 'av Steinar Supphellen "Med tillegg: 1. Opprettinga av ein norsk stathaldarinstitusjon i 1572 : ny samling av Noreg? ; 2. Supplikken som institusjon i norsk historie : framvokster og bruk særleg først på 17.h.talet." Avhandling (doktorgrad... av: Supphellen, Steinar'.

Bilaga 2

Bildvisning på webbplatsen för *Codex Gigas*.

