

Pilotprojektet för massdigitalisering: slutrapport

Offentlig version

Innehåll

1	Utgångspunkter	2
2	Mål och slutrapportens struktur	2
3	Begränsningar	3
4	Material	5
5	Tidtabell och skeden	6
6	Myndighetsamarbete och -styrning	8
7	Anskaffning av utrustning och infrastrukturlösningar	8
8	Datasäkerhet och dataskyddslösningar	10
9	Pilotprojektets resurser	11
10	Finansiering	11
11	Genomförande av digitaliseringen	12
12	Resultat	13
	Kriterium 1: Digitaliseringens kvalitet	14
	Kriterium 2: Överföring av det digitaliserade materialet till förvaring	16
	Kriterium 3: Det digitaliserade materialets användning och tillgänglighet	16
	Kriterium 4: Materialet kan förstöras	17
	Kriterium 5: Uppnående av pilotprojektets produktionsmål	18
	Kriterium 6: Digitaliseringen av pilotmaterialet lyckas	21
	Kriterium 7. Myndighetsberedningens funktionalitet	21
	Kriterium 8. Den övergripande processens funktionalitet	22
	12.2 Övriga resultat och observationer	23
	Katalogisering av material: utfall	23
	Funktionen och tydligheten för Riksarkivets myndighetsstyrning	23
	De nyttjade programmens omfattning och funktion	24
	Dataöverförings- och datahanteringskapacitet	24
	Observationer om olika materialtyper samt om utrustningens lämplighet och kapacitet	25
13	Slutsatser och fortsatta åtgärder	26

1 Utgångspunkter

Målet för massdigitaliseringen är att digitalisera statliga myndigheters handlingar som ska förvaras varaktigt och under en lång tid så att de ursprungliga analoga handlingarna kan förstöras efter att deras elektroniska förvaring och tillgänglighet har säkerställts. Genomförandet av massdigitaliseringen har planerats vid Riksarkivet sedan 2017. Enligt den styrgrupp som tillsattes för projektet förutsätter inledningen av den egentliga massdigitaliseringen ett pilotprojekt. Beslutet om att genomföra ett pilotprojekt fattades av styrgruppen i februari 2018.

Utgångspunkten för pilotprojektet var de uppskattningar om digitaliseringens referensmetoder, hastighet och kostnader man fick under massdigitaliseringens planeringsprojekt 2017. Pilotprojektets mål preciserades gradvis. Den slutliga versionen utarbetades i augusti 2018. I detta skede hade beredningen av pilotprojektet redan inletts. Utvärderingskriterierna för pilotprojektet definierades av styrgruppen i april 2019.

2 Mål och slutrapportens struktur

Målet för pilotprojektet var att få konkreta bevis om tillförlitligheten hos uppskattningarna om digitaliseringsprocessens funktion och hastighet. Massdigitaliseringsprocessen, den teknik som används för den, genomloppstiderna och kostnaderna hade bedömts under massdigitaliseringens planeringsprojekt 2017. Målet för pilotprojektet var att utreda hur realistiska dessa uppskattningar var för den egentliga digitaliseringsproduktionen. De centrala målen för pilotprojektet var att utreda:

- Digitaliseringsprocessens tillförlitlighet, den tid som går åt till olika arbetskedan samt processens totala längd
- Digitaliseringens kvalitet utifrån den egentliga produktionen (filobjekt som följer kriterierna för kraven på digitalisering med avsikt att förstöra)
- Hur smidigt och enhetligt digitaliseringsprocessen fungerar och framskrider vid den egentliga produktionen
- På vilket sätt massdigitaliseringens produktionsarbete effektivast kan organiseras och ledas
- Tydligheten i och omfattningen av de instruktioner som digitaliseringen förutsätter

Målen, begränsningarna och tidtabellen definierades i detalj i en separat utarbetad plan som godkändes av projektets styrgrupp.

Genomförandet av pilotprojektet förutsatte att man byggde upp en engångsproduktionsmiljö som bestod av lokaliteter, utrustning och produktionslokaler. Denna miljö gav värdefull information som kunde utnyttjas då man planerade hur massdigitaliseringens produktion skulle byggas upp. Även under pilotprojektet byggde man upp en helhet som bestod av olika databaser, program, appar och serverlösningar. Denna helhet är en första version av den produktionsstyrningshelhet som behövs för massdigitaliseringen.

Under pilotprojektet testades i praktiken även skannerutrustningens hållbarhet, funktionsduglighet och hastigheter vid egentlig produktion samt olika referensmodeller. Motsvarande aspekter hade testats redan tidigare utan produktion i utredningen massdigitaliseringens *Proof of Concept 2018*.

Särskilda kriterier för pilotprojektet, dvs. kriterierna 1–8, fastställdes senare av projektets styrgrupp. Dessa kriterier framgår av tabell 3 och utvärderingen av pilotprojektet i enlighet med dessa kriterier beskrivs i kapitel 12.

Man har försökt hålla slutrapporten så koncis som möjligt och koncentrera sig på att presentera resultat, observationer och slutsatser. Slutrapporten fokuserar på pilotprojektets digitaliseringsskede.

3 Begränsningar

På grund pilotprojektets korta genomförandeperiod vad gäller planering och digitalisering utslöts en del av de komponenter och mål som har inkluderats i den egentliga produktionen ur pilotprojektet. Utgångspunkten var att pilotprojektet ska kunna genomföras med en snäv tidtabell, utan behov av systemändringar som kräver betydligt mer tid än man hade till förfogande för att förbereda pilotprojektet. Sådana var bland annat tjänstehelheten för arkivering SAPA, ett integrerat produktionsstyrningssystem samt produktion av metadata med hjälp av innehållsanalys. I pilotprojektet ville man ha information om funktionsdugligheten hos så många olika digitaliseringsprocesser som möjligt. Därför omfattade genomförandet allt från myndighetsberedning till myndigheternas användning av digitalt material. Begränsningarna framgår mer i detalj i tabell 1 nedan. Sammanfattningsvis kan man konstatera att pilotprojektet genomfördes enligt de ursprungliga begränsningarna.

Tabell 1. Pilotprojektets begränsningar

Produktion	Pilotprojekt (mål)	Pilotprojekt (utfall)
Processkede/datasystem	Ingår (ja/nej)	
Myndighetsberedning och inmatning av metadata i AHAA-tjänsten	Ja	Ja
Separat logistikprocess för att samla in handlingar för digitalisering	Nej (Riksarkivet ansvarar dock för att avhämta handlingarna)	Nej
Styrning av digitaliseringens skeden med hjälp av produktionsstyrning	Nej (sköts manuellt)	Ja (logistikprogrammet genomfördes av Riksarkivet)
Beredning av material för digitalisering i samband med digitaliseringen	Ja	Ja
Högpresterande och annan skanning i enlighet med digitalisering som avser att förstöra handlingar som ska förvaras varaktigt	Ja	Ja
Högpresterande och annan skanning i enlighet med digitalisering som avser att förstöra handlingar som ska förvaras under viss tid	Nej	Nej
OCR-textidentifiering	Delvis (i digitaliseringen ingår textidentifiering, men inte innehållssökning i material)	Ja
Validering, paketering och överföring av det digitala uttrycket till KP-PAS-tjänsten (förvaring)	Ja	Delvis (väntar på band på SAPA-tjänsten och den KP-PAS-överföring som kommer att genomföras i och med denna)
Förvaring av användningsexemplar i SAPA-tjänsten	Nej	Delvis (vid digitaliseringen producerades användningsexemplar som kommer att överföras till SAPA-tjänsten)
Kundanvändning av uttrycken	Ja	Nej (användargränssnittet har genomförts men används inte ännu av kunderna)
Automatiserad valideringsprocess för digitaliseringen och integrerat produktionsstyrningssystem	Nej	Nej
Automatisk produktion av metadata ur handlingar med hjälp av innehållsanalys	Nej	Nej
Förstöring av analoga dokument	Nej (förstöring av analoga dokument ingår inte i pilotprojektet)	Nej

4 Material

I pilotprojektet var utgångspunkten för valet av material dess lämplighet för pilotprojektet. För pilotprojektet valdes material ur gruppen Färdplan 1, dvs. det material som har schemalagts först för massdigitalisering. Då man samlade in materialet ville man även beakta olika typer av material i syfte att få konkret tilläggsinformation om genomloppshastigheterna och hur väl olika material lämpar sig för massdigitalisering. Förutom av materialtyperna påverkades det kvantitativa målet även av digitaliseringshastigheten samt den utrustning som användes och personalresursernas omfattning. Slutresultatet var att man i pilotprojektet förband sig att under den fem månader långa produktionsperioden digitalisera ca 450 hyllmeter material ur färdplanens grupp 1, från fem olika myndigheter, och som representerade olika materialtyper.

Materialet bestod av olika pappershandlingar, band och och kartotek. Den största delen av materialet var handlingar av storlek A4. Andelen band var ca 2 procent av den totala mängden. Materialen har presenterats i tabell 2 nedan. Materialets omfattning har angetts i hyllmeter.

Tabell 2. Pilotprojektets material, antal och omfattning

Organisation	Arkiv och materialtyp	Startår– slutår	Förvarings enheter (st.)	Förvaringsutr ymme (hyllmeter =hm)
Skatteförvaltni ngen	Skatteverket i Esbo A4-ark, gem	1994–2011	1119	93,65
Skatteförvaltni ngen	Skatteverket i huvudstadsregionen A4-ark, gem	2014–2015	144	13,75
Fimea	Fimea A4, bundna	1995–2016	1238	116,7
Institutet för hälsa och välfärd	Folkhälsoinstitutet A4-häfte, olika arkstorlekar, nitade	1966–2001	993	99,4
Arbets- och näringsministe riet	Arbetsministeriet A4-ark, gem	1999–2009	877	71,05
Västra Nyland tingsrätt	Esbo domsaga Kartotek, band, A4, velobind-bindning	1971–1994	286	30,4
Västra Nyland tingsrätt	Esbo tingsrätt A4	1993–1994	12	1,3
SUMMA		1966–2015	4669	426,25¹

¹ Materialet kontrollerades i samband med digitaliseringen och mängden befanns vara 386,17 hyllmeter.

5 Tidtabell och skeden

Pilotprojektet omfattar inte enbart digitalisering. Pilotprojektet inleddes genom att definiera målen och materialvalen. Sedan omfattade projektet samtliga skeden i digitaliseringsprocessen från myndighetsberedning till kundernas användning av materialet. Pilotprojektets skeden var delvis överlappande. På mer allmän nivå kan pilotprojektet indelas i planerings-, berednings-, genomförande- och analyseringsskeden. Av dessa genomfördes det förstnämnda under perioden 02/2018–03/2019 och de två sistnämnda under perioden 04–10/2019. Under planerings- och beredningsskedet definierades bl.a. pilotprojektets mål, genomfördes en myndighetsberedning samt instruktioner och styrning av denna och konkurrensutsattes pilotprojektets anskaffningar. Under genomförande- och analyseringsskedet byggdes bl.a. produktionsmiljön, digitaliserades materialet och analyserades resultaten.

Som tidpunkt för pilotprojektets genomförande- och analyseringsskede fastställdes 04–10/2019. Pilotprojektets digitalisering och resultat skulle alltså vara klara senast 31.10.2019.

Pilotprojektets planerade och förverkligade tidtabeller framgår av bild 1 på nästa sida.

Bild 1. Tidtabell för pilotprojektets framskridande enligt delområde

Pilotprojektets delområden och längd	År	2018					2019					
		Månad	03-04	05-06	07-08	09-10	11-12	01-02	03-04	05-06	07-08	09-10
Planering av och inledande budgetering för referensmodellen	Plan											
	Utfall											
Myndighetssamarbete (materialval, avtal)	Plan											
	Utfall											
Myndighetsstyrning (introduktion och myndighetsberedning av materialet)	Plan											
	Utfall											
Teknisk beredskap (applikationer och system)	Plan											
	Utfall											
Konkurrensutsättning och anskaffande av utrustning och program	Plan											
	Utfall											
Personalrekrytering	Plan											
	Utfall											
Beredning av och ändringsarbeten i infra och arbetslokaler	Plan											
	Utfall											
Materiallogistik (transportavtal och materialtransporter)	Plan											
	Utfall											
Personalintroduktion, organisering av arbetet, inledning av digitaliseringen	Plan											
	Utfall											
Pilotprojektets aktiva genomförandeskede (beredning och digitering) och slutrapportering	Plan											
	Utfall											

6 Myndighetssamarbete och -styrning

Avtal för pilotprojektet utarbetades med varje myndighet. I dessa definierades avtalsparternas ansvar, bl.a. vad gäller hanterings- och användarrättigheten till materialets digitala och analoga versioner, digitalisering, förvaring och förstöring, behandling av personuppgifter och personuppgiftsansvar, datatjänsten samt frågor som anknyter till datasäkerheten och dataskyddet.

Valet av material gjordes inom tidtabellen, men beredningen av avtal tog längre än planerat, särskilt på grund av att noteringarna om dataskyddet måste preciseras.

För pilotprojektet utarbetades nya versioner av Riksarkivets överföringsguide och överföringsanvisning, i vilka kriterierna för och kraven på massdigitalisering presenterades. Pilotprojektets myndighetsberedning genomfördes i enlighet med dessa program. I och med massdigitaliseringen underlättades myndighetsberedningen något i jämförelse med de tidigare överföringskriterierna.

Innan överföringen av materialet till pilotprojektet skulle myndigheterna katalogisera de katalogiseringsuppgifter som gällde materialet i Riksarkivets AHAA-katalogtjänst. AHAA:s produktionsversion togs i drift i samtidigt och därför utvecklades AHAA och handleddes myndigheterna till att använda AHAA som en del av pilotprojektet. För myndigheterna utarbetades separata katalogiseringsinstruktioner om användningen av AHAA-katalogtjänsten. Myndigheternas katalogiseringsarbete styrdes och arbetes framskridande följdes upp med hjälp av handledningsbesök och fjärrstyrning.

De myndigheter som deltog i pilotprojektet fick en separat enkät om den styrning de erbjöds och de instruktioner som hade utarbetats för dem. Utifrån enkätens resultat vidareutvecklas instruktionerna och styrningen efter pilotprojektets slut.

Riksarkivet ansvarade för transporten av materialet från myndigheterna till Riksarkivet samt för transportkostnaderna. Transporterna genomfördes av det transportföretag som valdes utifrån Hansels ramavtal genom förenklad konkurrensutsättning. Vid transporterna beaktades förutom normal transport av arkivmaterial och säkerställande av datasäkerheten även myndigheternas villkor för behandlingen av materialet.

7 Anskaffning av utrustning och infrastrukturlösningar

Vad gäller den tekniska beredskapen var alternativen att genomföra pilotprojektet med Riksarkivets egna resurser, helt genom extern anskaffning och inköpstjänster via konkurrensutsättning, eller som en mellanform av dessa. Som den bästa genomförandemodellen valdes en kombination där skannarna och programhelheten

skaffades som köptjänst, infrastrukturlösningarna skaffades av CSC Oy och de applikationer som anknyter till produktionsstyrning och materialöverföring utvecklades internt vid Riksarkivet.

Riksarkivet utarbetade en teknisk kravspecifikation för skanner- och programhelheten, utifrån vilken en upphandling genom öppet anbudsförfarande ordnades. Hansels konsulttjänster användes vid upphandlingen. Leveransprojektet för skannerhelheten avslutades med ett acceptanstest genom vilket man säkerställde att helheten uppfyllde kravspecifikationerna.

Skannrarnas förmåga att producera bilder i enlighet med kvalitetskraven och de fastställda maximumhastigheterna verifierades vid implementeringstestningen. Produktionens maximumhastigheter berodde på och fastställdes enligt materialets kvalitet. Den allmänna medelhastigheten för högpresterande enheter var 210 ark i minuten. Med dokumentskanner var den maximala hastigheten cirka 170 ark i minuten. Med mastskanner var medelhastigheten cirka 9 ark i minuten.

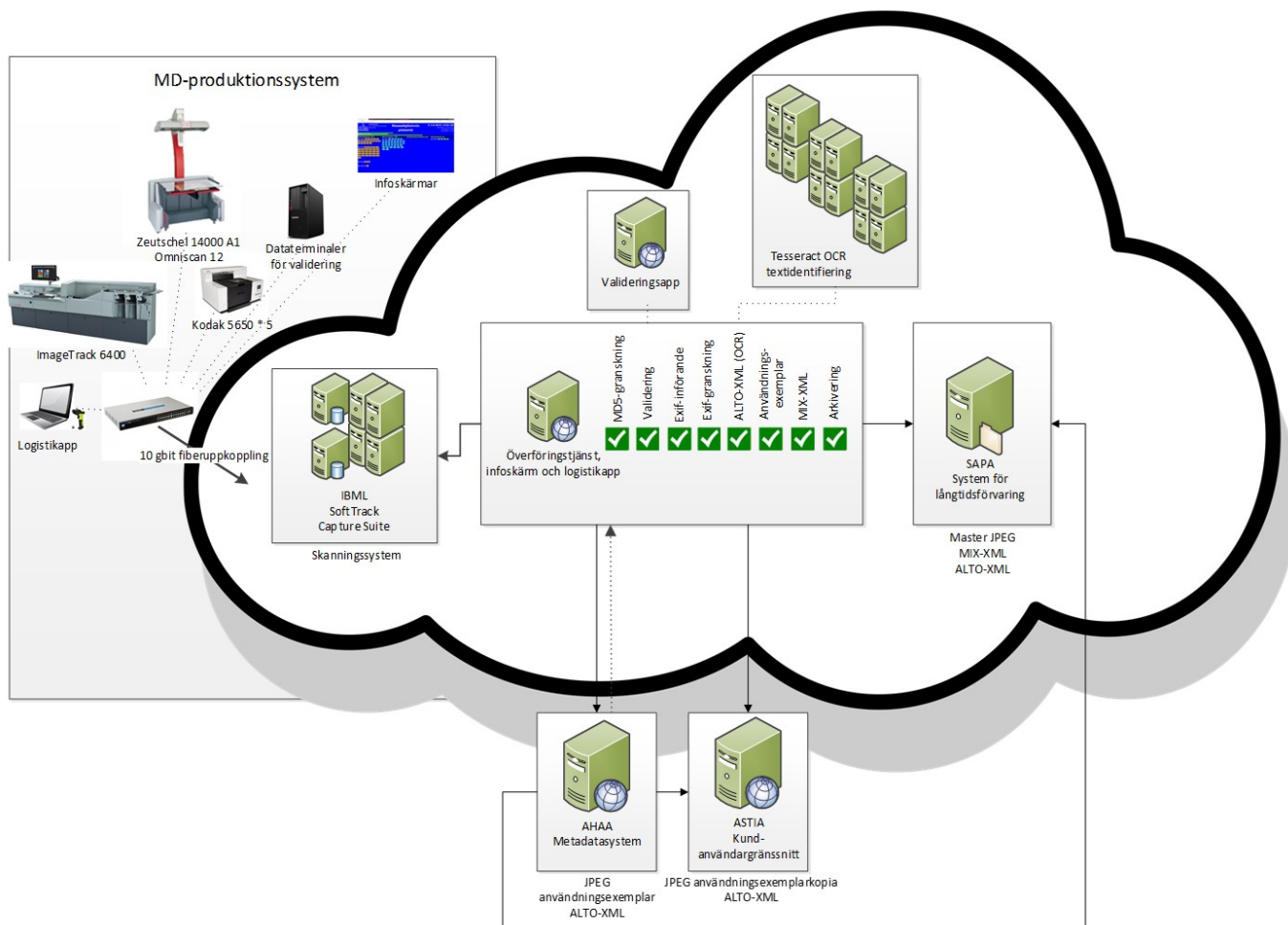
De infrastrukturlösningar som CSC genomförde grundade sig på ett virtuellt serversystem med hjälp av vilket man bäst kunde avgöra behoven i anknytning till servrarnas placering, antal och skalning. CSC:s genomförande grundade sig på ett ramavtal mellan Riksarkivet och CSC.

Riksarkivet tog fram en logistikapp, valideringsapp, överföringstjänstapp och en infoskärmapp för pilotprojektet.

Logistikappens uppgift var att säkerställa hanteringen av de analoga och digitala uttrycken i processens olika skeden. Med hjälp av logistikappen kunde man följa med överföringen av materialet till olika skeden i realtid. Logistikappens syfte var att skapa en smidig metod för att säkerställa att funktionerna i anknytning till kvalitetskontroll, produktionshastighet och de rapporteringsegenskaper som beskriver produktionens nyckeltal skulle fås till en del av pilotprojektet.

Det digitaliserade materialets bildkvalitet vad gäller datainnehållet granskades med hjälp av en konfigurerbar valideringsapp. Överföringstjänstappen utvecklades för efterbehandling och bearbetning av det digitaliserade materialet (bl.a. OCR-identifiering). Infoskärmappens uppgift var att dela ut olika former av aktuell information till produktionspersonalen, bl.a. om mängden material i olika skeden av processen, behandlingen av produktionstillstånd, tidtabellerna och diskkapaciteten. Pilotprojektets systemhelhet framgår på allmän nivå av bild 2 nedan.

Bild 2. Appar, program och gränssnittssystem i systemhelheten för pilotprojektets digitalisering



8 Datasäkerhet och dataskyddslösningar

Det material som skulle digitaliseras omfattade rikligt med sekretessbelagda handlingar. För att säkert sälla dataskyddet genomfördes flera olika datasäkerhetsförfaranden. De centrala lösningarna infördes i avtalen för pilotprojektet mellan myndigheterna och Riksarkivet. Tillträde till digitaliseringens förvarings- och övriga lokaler begränsades uteslutande till massdigitaliseringens personal. Digitaliseringens program- och servermiljö jämte applikationer genomfördes som differentierade lösningar från andra nätverk. Personalen fick information om och introduktion till dataskydds- och datasäkerhetskraven samt förfaringsätten. En egen datasäkerhetsplan utarbetades för pilotprojektet, utifrån vilken de nödvändiga datasäkerhetslösningarna genomfördes. Status för pilotprojektets datasäkerhet följdes upp regelbundet och separata instruktioner för hantering av avvikelser utarbetades.

9 Pilotprojektets resurser

Personalrekryteringen genomfördes i två huvudskeden. Efter att resursbehoven hade definierats, rekryterades expert- och styrpersonalen under det första skedet. Under det andra skedet rekryterades produktionspersonalen. Ett introduktionsprogram utarbetades för dessa personer. Programmet fokuserade på pilotprojektets mål, referensmetod, verktyg samt dataskyddskrav.

Förutom de nya rekryteringarn definierades och styrdes även massdigitaliseringens och Riksarkivets övriga personalresurser till pilotprojektets expertuppgifter. Dessa expertuppgifters andel av pilotprojektet var betydande. Under 2018–2019 anvisades pilotprojektet en expertinsats på sammanlagt över 14 årsverken.

Pilotprojektets produktionspersonal bestod av helt nya personer och personer som hade rekryterats för denna uppgift. En personalresurs på 7+2 personer hade planerats för pilotprojektets produktion. Eftersom man använde sju skannrar i pilotprojektet var detta personalantal det minsta möjliga med vilket man uppskattade sig klara av produktionen inom tidsfristen. En viss personalomsättning förekom. Det genomsnittliga antalet personal som var närvarande under en produktionsdag var 7,8. Förutom skanning skulle personalen även ansvara för bl.a. beredning av materialet, datatjänsten och underhåll av lokalerna. De två personer som ansvarade för logistiken deltog även efter behov i andra uppgifter. Hela personalen introducerades till samtliga delområden inom produktionen. År 2019 var produktionspersonalens totala resurs över 5 årsverken.

I pilotprojektet deltog sammanlagt 35 av Riksarkivets fastanställda eller visstidsanställda.

10 Finansiering

Pilotprojektets finansiering ingick i finansieringen av massdigitaliseringens utveckling, som bestod av olika delar och var sammansatt på olika sätt 2018 och 2019.

Inför pilotprojektet var man tvungen att genomföra ett ganska stort antal enskilda anskaffningar, eftersom både produktionslokalerna, -utrustningen samt till stora delar även programmen, apparna och den övriga infrastrukturen måste planeras och byggas upp från början. Kostnadsfördelarna med den finansiering som riktades till pilotprojektet kommer därför att omfatta en längre period efter pilotprojektet, eftersom pilotprojektets produktionsmiljö och de processer och verksamhetsmodeller som skapades kan utnyttjas för senare massdigitalisering.

Dessutom utvecklade man i samband med pilotprojektet mer omfattande förutsättningar för massdigitalisering, bl.a. genom ändringar i de anknyttande datasystemen (AHAA och AHJ) samt genom att utveckla bl.a. myndighetsstyrningen, precisera färdplanen och utveckla den automatiska produktionen av metadata med hjälp av innehållsanalys. Samtliga helheter som utvecklades har inte ännu kunnat utnyttjas i pilotprojektet och utvecklingsarbetet fortsätter delvis ännu under de kommande åren.

Pilotprojektets centrala kostnader bestod av personalkostnader (cirka 31 procent) och anskaffning av utrustning, program och tjänster (cirka 52 procent).

11 Genomförande av digitaliseringen

Pilotproduktionens centrala processer var materialets myndighetsberedning, beredning av materialet inför digitalisering, digitalisering (inklusive bl.a. skanningens och valideringens delprocesser), bearbetning av det digitaliserade materialet (inklusive OCR- och annan efterbehandling av det digitala materialet) samt behandling av av det analoga materialet efter digitaliseringen.

Genomförandet av digitaliseringens produktion förutsatte separat produktionsplanering, genom vilken man schemalade produktionen och delade ut material till olika enheter, riktade resurser samt styrde och handledde personalen inför uppgifterna. Utifrån den anskaffade utrustningen beslöt man sig i produktionsplaneringen för en modell med tre produktionslinjer. Materialet bereddes och matades in i dessa med hjälp av fyra olika referensmodeller. Med produktionslinjer avses de skannertyper som användes i pilotprojektet och med genomförandemodellen avses de olika förfaringssätt som utnyttjades för att bereda materialet samt för skanningen.

Produktionsplaneringen omfattade även lokalplanering och anskaffning av materiel. Betydande investeringar gjordes i pilotprojektets produktionslokaler genom att skaffa den nödvändiga inredningen, övrig utrustning och övriga redskap samt genom att uppdatera lokalens datakommunikationsförbindelser.

I sin helhet strävade man i pilotprojektet efter industriell linjeproduktion i den utsträckning detta var möjligt i de lokaler som stod till förfogande och utan helt automatiserade produktionsstyrningssystem.

Ett centralt redskap i produktionsplaneringen var pilotprojektets logistikapp, med hjälp av vilken man följde upp och styrde överföringen av material mellan processerna. Under

myndighetsberedningen hade man på materialet skrivit ut en streckkod utifrån de metadata som hade matats in i AHAA-systemet. Med hjälp av streckkoden lästes materialet in från ett processkede till följande. Materialet förvarades i produktionslokalerna endast under den tid som beredningen och digitaliseringen krävde. Genast efter digitaliseringens slut återbördades materialet till förvaringsmagasinet i väntan på förstöring och lästes in i detta behandlingsskede via logistikappen.

Formatet, kvalitetssäkringen och bearbetningen av de digitala uttryck som producerades i digitaliseringen beskrivs i kapitel 12.

12 Resultat

I detta kapitel utvärderas resultaten utifrån pilotprojektets kriterier samt de andra kvalitetsmått som projektets styrgrupp ställde upp. En sammanfattning av kriterierna, måtten och genomförandet visas i tabell 3 nedan. Uppfyllandet av de enskilda kriterierna utvärderas mer i detalj under tabellen.

Tabell 3. Pilotprojektets kriterier och deras uppfyllande

Huvudkriterium	Förklaring av kriteriet	Bedömningsgrund/skala (genomfört alternativ i fet stil)
K1. Digitaliseringens kvalitet	Bildkvaliteten och kvalitetsnivån för digitalisering som avser att förstöra handlingar uppnåddes i pilotprojektet med den hastighet som krävdes	Genomförs <i>Smärre avvikelser</i> <i>Betydande avvikelser</i> <i>Genomförs inte</i>
K2. Överföring av det digitaliserade materialet till förvaring	I digitaliseringen lyckades produktionen av ett definierat överföringspaket för långtidsförvaring med hjälp av metadata	Genomförs <i>Smärre avvikelser</i> <i>Betydande avvikelser</i> <i>Genomförs inte</i>
K3. Det digitaliserade materialets användning och tillgänglighet	Myndigheterna och Riksarkivet kan använda det digitaliserade materialet för de nödvändiga användningsändamålen. Materialets tillgänglighet förbättras märkbart.	<i>Genomförs</i> <i>Smärre avvikelser</i> Betydande avvikelser <i>Genomförs inte</i>
K4. Materialet kan förstöras	Kraven i kriterierna K1–K3 uppfylls i så hög grad att materialet kan förstöras efter karenstiden	<i>Ja/Nej</i>
K5. Uppnående av pilotprojektets produktionsmål	I pilotprojektet uppnår man en genomsnittlig produktionshastighet i enlighet med målen	<i>Genomförs</i> Smärre avvikelser <i>Betydande avvikelser</i> <i>Genomförs inte</i>
K6. Digitaliseringen av pilotmaterialet lyckas	Samtligt material kan digitaliseras under pilotprojektet, inget material förblir odigitaliserat	Genomförs <i>Smärre avvikelser</i> <i>Betydande avvikelser</i> <i>Genomförs inte</i>

K7. Myndighetsberedningens funktion	Myndigheterna har berett materialet i tid och enligt Riksarkivets instruktioner. Beredningen av materialet uppvisar inga brister som stör digitaliseringen eller användningen av det digitaliserade materialet.	Genomförs <i>Smärre avvikelser</i> <i>Betydande avvikelser</i> <i>Genomförs inte</i>
K8. Den övergripande processens funktion	Digitaliseringens övergripande process från myndighetsberedningen till användningen av materialet: enskilda processkedan behöver inte upprepas eller repareras i betydande utsträckning. Pilotprojektets delprocesser kan genomföras. Pilotprojektets övergripande tidtabell genomförs.	<i>Genomförs</i> Smärre avvikelser <i>Betydande avvikelser</i> <i>Genomförs inte</i>

Kriterium 1: Digitaliseringens kvalitet

Kriteriet var att den bildkvalitet och kvalitetsnivå som krävs vid digitalisering som avser att förstöra handlingar uppnås i pilotprojektet med den hastighet som krävs.²

- **Kriteriet uppfylls**

De krav som fungerar som kvalitetsmått för kriterium 1 har definierats som obligatoriska i den tekniska kravdefinition som gäller anskaffningen av skanner- och programhelheten. Vid systemets implementeringstestning säkerställdes uppfyllandet av dessa krav i utrustnings- och programkonfigurationen. Då systemet byggdes upp kalibrerade utrustningsleverantören utrustningen i enlighet med kraven. Då implementeringen godkändes kunde man konstatera att kraven hade uppfyllts.

Utifrån de installationsinställningar som hade kalibrerats definierades referensvärden för bildkvaliteten hos pilotprojektets produktion. Dessa följdes upp dagligen. På så sätt gavs dagsspecifika produktionstillstånd för varje apparat. Mastskannern var ett undantag: eftersom den användes så lite säkerställdes kvaliteten en gång per produktionsvecka. Universal Test Target (UTT)-bildkvalitetsobjekt och programmet iQ-Analyzer användes för de automatiska bildkvalitetsmätningarna. De vanligaste orsakerna till att produktionslovsmätningarna förkastades var slitage, snedhet eller dammighet hos objekten.

I samband med skanningen följdes skannrarnas bildkvalitet kontinuerligt upp även okulärt. Det vanligaste fenomenet som ledde till nedsatt bildkvalitet och som den automatiska analysen inte identifierade var skråmor eller streck på bilden. Dessa berodde vanligtvis på damm eller skräp.

² <https://arkisto.fi/fi/viranomaisille/Julkishallinnon-asiakirjahallinnon-ja-arkistotoimen-ohjaus/maeeraeykset/kansallisarkiston-vaatimukset-h%C3%A4vitt%C3%A4miseen-t%C3%A4ht%C3%A4v%C3%A4v%C3%A4n-digitointiin>

I samband med valideringen granskades bildkvaliteten en tredje gång. För detta genomfördes en särskild valideringsapp internt vid Riksarkivet. I pilotprojektet beslöt man validera 100 procent av det digitaliserade materialet, dvs. varje bild granskades visuellt. Vid valideringen säkerställdes integriteten hos bildfilens informationsinnehåll. I pilotprojektet definierades särskilda kriterier för förkastande av bilder.

Om man vid valideringen uppdagade felaktiga bilder, skannades hela arkivenheten huvudsakligen på nytt. I fall av enstaka fel avvägdes huruvida endast de felaktiga bilderna skulle skannas på nytt. Cirka 9 procent av pilotprojektets material förkastades vid valideringen och skannades sedan på nytt.

Valideringens kriterier för förkastande och fördelningen av förkastanden bland dessa framgår av tabell 4 nedan.

Tabell 4. Valideringens kriterier för förkastande

Kriterier för förkastande av bilder	Fördelningen av fel (i procent)
Dubbelmatning	5,28
Information syns inte	61,80
Bilden är upp och ner, gäller över 1 procent av de bilder som ingår i arkivenheten (gäller endast maskinskriven text)	18,94
Dålig bildkvalitet	5,90
Trasig handling	0,31
Fel innehåll	0,00
Ark/sida saknas	5,59
Annan orsak	2,17
Fel som inte ledde till förkastande av bilder	
Tomt formulär som är upp och ner.	
Ett litet område fattas ur formulärmallen, t.ex. ett hörn har vikts, men de övriga bilderna har samma informationsinnehåll.	
Helt handskrivna handlingar som är upp och ner (räknas inte med i samplet av handlingar som är upp och ner).	
Felaktigt skannat. Detta har uppdagats under skanningsskedet och handlingen har därför skannats på nytt. Till följd av detta syns bägge bilder i valideringen (dvs. den dåliga bilden och den bra bilden efter varandra).	
Formulär som är upp och ner, som hela tiden återkommer i enheterna, och i vilket informationen har fyllts i för hand.	
Handling utan textinnehåll och som är upp och ner.	

Kriterium 2: Överföring av det digitaliserade materialet till förvaring

Kriteriet var att man i digitaliseringen kan producera ett överföringspaket som uppfyller kravdefinitionerna för digitalisering med avsikt att förstöra och som innehåller de metadata för långtidsförvaring som krävs.

- **Kriteriet uppfylls**

Vid konkurrensutsättningen för skanner- och programhelheten förutsattes att de metadata som krävs för långtidsförvaring produceras under skanningen. Vid godkännandetestningen av den anskaffade utrustningen säkerställde man att dessa metadata bildas i samband med digitaliseringen.

Med hjälp av Riksarkivets överföringstjänsteapp genomfördes en filstruktur som följer skanningsapparnas (Softack Capture Suite, Omnican 12) kravdefinitioner samt kontinuerligt säkerställande av metadata. Appen granskade att filstrukturens kvantitativa motsvarighet till kravdefinitionen samt räknade ut MD5-kontrollsummor. Därefter styrde överföringstjänsten materialet till valideringsappen som har beskrivits under kriterium 1. Sedan överfördes materialet till optisk textidentifiering (OCR), där man för varje sida med informationsinnehåll producerade textidentifieringsfiler (XML) som följde ALTO 3.0-standarden med hjälp av redskap med öppen källkod (Tesseract). För bildfiler producerades i överföringstjänsteappen JPEG (90 %, 300 ppi) som förvaringsexemplar av de digitala uttrycken, och för de metadata som beskriver digitaliseringsprocessen producerades en XML-fil som följer MIX-schemat i enlighet med Riksarkivets kravdefinition för digitalisering med avsikt att förstöra som nämns under kriterium 1.

Förfarandet garanterade att de metadata för långtidsförvaring som krävs i kriterium 2 bildades och förvarades.

Kriterium 3: Det digitaliserade materialets användning och tillgänglighet

Ett kriterium var att myndigheterna och Riksarkivet kan använda det digitaliserade materialet för nödvändiga användningsändamål och att materialets tillgänglighet förbättras märkbart.

- **Tillsvidare omfattar kriteriet märkbara avvikelser**

Enligt planen skulle de digitala bilder som producerades i pilotprojektet visas, laddas och sökas, kataloguppgifter sökas och bläddras samt innehållsidentifierad text sökas av kunderna i Astia-användargränssnittet. Tjänsten utvecklades i ett separat projekt som

ett tjänstearbete vid Riksarkivet samtidigt som pilotprojektet planerades och genomfördes.

Astia-användargränssnittet har en produktionsversion i vilken det massdigitaliserade materialet kan granskas med beaktande av dess användningsbegränsningar. På grund av två olika orsaker har man dock inte kunnat ta Astia-användargränssnittet i myndighetsanvändning som planerat. För det första blev det under pilotprojektet klart att alla pilotmyndigheter inte använder överföring av rollinformation via VIRTU-identifiering. Inloggning i Astia kräver VIRTU-rollinformation. När problemet kom till Riksarkivets kännedom, beslöt man ändra Astias identifieringslösning så att myndigheterna inte behöver uppge rollinformation. Detta ändringsarbete pågår fortfarande. Ett annat problem uppdagades då systemet testades: samtidig uppladdning av stora mängder bilder i användargränssnittet orsakade fel i bildvisningen. Reparationen av detta problem pågår.

Trots de funktionella bristerna i tjänsten har myndighetersanvändarna introducerats till användningen av Astia på ett särskilt utbildningsevenemang. Då man efter tillräcklig testning har kunnat lägga sista handen vid de icke slutförda lösningarna finns beredskapen att starta upp tjänsten.

Kriterium 4: Materialet kan förstöras

Här var kriteriet att kriterier 1–3 har uppfyllts i så hög grad att materialet kan förstöras efter karenstiden.

- ***Kriteriet uppfylls – materialet kan förstöras efter karenstiden***

I praktiken var avsikten med pilotprojektet att genomföra högpresterande och annan skanning i enlighet med kraven för digitalisering med avsikt att förstöra. Detta bygger på de kriterier för digitalisering med avsikt att förstöra som definieras i arkivlagens 14 a §: en handling får förstöras om detta kan ske utan att bevarandet av, integriteten för eller konstaterandet av autenticiteten av handlingen eller av uppgifterna i den äventyras och utan att handlingens kulturhistoriska värde eller juridiska bevisvärde minskar.

Avsikten var dock inte att förstöra de analoga exemplaren under pilotprojektets genomförandeperiod, utan att digitalisera det analoga materialet så att det kan förstöras senare efter en karenstid. Detta förutsätter att det digitala materialets långtidsförvaring har säkerställts genom övergöring till SAPA-tjänsten och därifrån vidare till KP-PAS-förvaring.

Kvalitetsmått för pilotprojektets framgång vad gäller bildkvalitet och metadata i enlighet med kriterier 1–2 uppfylldes för det material som digitaliserades under

pilotprojektet. En del av det förkastade materialet kunde skannas på nytt först efter den egentliga pilotperioden. (Se kriterium 6 och anknyttande observationer.) Kundanvändningen av digitalt material kunde inte heller genomföras inom tidsfristen.

Vid valideringen uppfylldes kraven i kriterium 4 för det material som redan hade godkänts. Sammanfattningsvis kan det konstateras materialet kan förstöras när man har fått erfarenhet av kundanvändningen av det digitala materialet och materialet framgångsrikt har kunnat överföras till systemet för långtidsförvaring.

Kriterium 5: Uppnående av pilotprojektets produktionsmål

Kriteriet var att man i pilotprojektet uppnår en genomsnittlig produktionshastighet på 4,5 hm/dag i enlighet med målen.

- **Kriteriet omfattar smärre avvikelser**

Den dagsspecifika produktionshastighet som hade fastställts som kriteriets kvalitetsmått hade beräknats utifrån den kalkylmässiga materialmängd som användes vid planeringen av pilotprojektet, dvs. 450 hyllmeter, och utifrån den förväntade produktionen för fem månader. I samband med pilotprojektet kontrollräknades materialet, till följd av vilket den faktiska mängden material som skulle skannas preciserades till 386,17 hm (se tabell 2). Den dagsspecifika produktionshastigheten som motsvarar denna mängd skulle vara 3,86 hm. Den dagsspecifika takt som uppnåddes i pilotprojektet var 3,91 hm. Det ursprungliga och förverkligade målet är inte helt jämförbara på grund av den förutnämnda skillnaden i mängden material samt antalet förväntade och genomförda produktionsdagar.³

Kriteriet omfattar smärre avvikelser, eftersom produktionshastigheten inte stabiliserades och inte heller blev kumulativ mot slutet på det sätt som hade planerats. Materialets varierande karaktär och den utrustning som användes påverkade detta väsentligt. I juli digitaliserades det material som går fortast att skanna med den effektivaste högpresterande skannern, vilket syns på talen. Dessutom påverkades produktionshastigheten av antalet personalresurser. I de uppskattade minimiresurserna hade man inte beaktat frånvaro i tillräcklig utsträckning. På grund av frånvaro kunde apparaterna inte användas med full kapacitet. Personalen belastades även av den relativt stora mängden material (9 procent) som måste skannas på nytt och som inte hade beaktats i planeringen av pilotprojektets tidtabell. Detta beror på att avsikten i planeringsskedet var att genomföra valideringen sampelspecifikt. I början av

³ I hastigheten har man beaktat de faktiska produktionsdagarna.

skanningsskedet beslöt man dock genomföra 100-procentig validering, eftersom man ville säkerställa bildkvaliteten. Därför gick arbetstid som hade planerats för skanning åt till validering. Det större behovet av att skanna material på nytt påverkade i det närmaste arbetets totala längd.

Med pilotprojektets takt och de tillgängliga resurserna skulle man kunna digitalisera cirka 978 hyllmeter per år. Det finns dock skäl att betona att takten först och främst påverkas av formatet och skicket hos det material som digitaliseras.

De absoluta och genomsnittliga produktionshastigheterna för olika referensmodeller samt datamängderna framgår av tabell 5 nedan.

Tabell 5. Pilotprojektets produktionsmängder och -hastigheter samt datamängder

Referensmodell	Linjenät	April-maj		Juni		Juli		Augusti		September		Oktober		Totalt	
		Dag (medel-värde)	Totalt (hm)	Dag (medel-värde)	Totalt (hm)	Dag (medel-värde)	Totalt (hm)	Dag (medel-värde)	Totalt (hm)	Dag (medel-värde)	Totalt (hm)	Dag (medel-värde)	Totalt (hm)	Dag (medel-värde)	Totalt (hm)
Referensmodell 1 (högpresterande)	Beredning (på förhand)	2,27	61,28	1,13	18,00	1,83	42,20	1,34	28,04	1,92	34,54	1,27	24,08	1,68	208,14
	Skanning (separat)	2,61	18,30	3,81	39,96	3,90	52,70	3,54	33,62	2,98	25,33	2,95	26,51	3,39	196,42
Referensmodell 2 (5 dokumentskannrar)	Beredning (på förhand)	1,19	32,20	0,38	6,14	0,02	0,39	0,05	0,99	0,04	0,70	0,34	6,46	0,38	46,88
	Skanning (separat)	1,12	30,18	0,55	8,81	0,02	0,46	0,01	1,14	0,25	4,57	0,56	10,64	0,45	55,80
Referensmodell 3 (5 högpresterande skannrar)	Beredning och skanning (samtidig)	0,18	4,96	1,05	16,87	1,80	41,31	1,80	37,88	2,11	38,06	1,52	28,83	1,35	167,91
Referensmodell 4 (mastskanner)	Beredning	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Skanning (samtidig)		0,00	0,02	0,11	0,04	0,80	0,04	0,68	0,01	0,10	0,00	0,00	0,02	1,69
Totalt	Beredning	3,46	93,48	1,51	24,14	1,85	42,59	1,38	29,03	8,81	35,24	1,61	30,54	2,06	255,02
	Skanning (skanning, beredning+skanning)	4,45	53,44	4,11	65,75	4,14	95,27	3,49	73,32	3,78	68,06	3,47	65,98	3,91	421,82
Antal (st., Gb)		Dag	Mån.	Dag	Mån.	Dag	Mån.	Dag	Mån.	Dag	Mån.	Dag	Mån.	Dag medel-värde	Maj-oktober
Bildantal (st.)		23 808	642 809	61 294	980 702	64 607	1 485 959	46 425	974 935	46 545	837 804	62 472	1 436 851	51 283	6 359 060
Datamängd (Gb)		50,29	1 357,80	148,29	2 372,61	152,88	3 516,23	113,25	2 378,26	95,97	1 727,46	161,93	3 076,68	116,36	14 429,04

Kriterium 6: Digitaliseringen av pilotmaterialet lyckas

Det fastställda kriteriet var att allt material som ingår i pilotprojektet kan digitaliseras under projektet och att inget material förblir odigitaliserat.

- **Kriteriet uppfylls**

Allt material i pilotprojektet kunde digitaliseras åtminstone en gång inom tidsfristen. Detta arbete färdigställdes 28.10.2019.

Alla olika typer av material kunde beredas och digitaliseras, dvs. inget material förblev odigitaliserat på grund av dess egenskaper. Bortsett från bundet material motsvarade digitaliseringshastigheterna huvudsakligen det som hade planerats.

Dessutom uppdagades ett tekniskt problem i en del av digitaliseringen, till följd av vilket enstaka bilder fattades ur en del av de digitaliserade enheterna. Felet kunde lokaliseras till funktionen för en drivrutin i en av skannertyperna och man fick ett verktyg för att identifiera felet. Förekomsten av enskilda fel, särskilt utan tillgång till ett gränssnitt för det digitaliserade materialet, skulle dock ha förutsatt manuell genomgång av en del av materialet, vilket inte ansågs arbetsekoniskt lönsamt. Som det bästa alternativet ansågs att skanna en del av materialet på nytt. Denna skanning färdigställdes 18.11.2019

Kriterium 7. Myndighetsberedningens funktionalitet

Kriteriets kvalitetsmått var att myndigheterna kunde bereda materialet i tid och enligt Riksarkivets instruktioner. Beredningen av materialet uppvisar inga brister som skulle störa digitaliseringen eller användningen av det digitaliserade materialet.

- **Kriteriet uppfylls**

Riksarkivet utarbetade instruktioner för att bereda pilotprojektets material för överföring (Överföringsguide) samt myndighetsstyrning för AHAA-katalogisering, dvs. den mest centrala uppgiften att katalogisera materialet i AHAA-tjänsten. Även ett utbildningsevenemang om användningen av AHAA ordnades. Dessutom styrdes myndigheternas katalogiseringsarbete genom uppföljning av den information som myndigheterna matade in i systemet och genom att ge tilläggsinstruktioner vid ämbetsverksbesök, per telefon och via den e-postadress för respons som hade grundats för detta.

Innan myndighetsberedningen inleddes bekräftades materialets motsvarighet till utgångsmaterialet på plats under ämbetsverksbesöken. Riksarkivet ansvarade också för

transporten av materialet för digitalisering samt anknyttande kostnader. Myndigheterna gav positiv respons om pilotprojektets styrning.

Myndighetsberedningen lyckades väl och inga fel hittades i materialet eller de kataloguppgifter som har införts i AHAA-tjänsten. Myndigheternas engagemang och sakkunskap var en central faktor för att materialet kunde fås till digitalisering i tid och utan brister.

Kriterium 8. Den övergripande processens funktionalitet

Ett mått för kriteriet var att digitaliseringens övergripande process eller dess enskilda skeden från myndighetsberedningen till användningen av materialet fungerar, att den inte behöver upprepas eller repareras märkbart, att pilotprojektets delprocesser kan genomföras och att pilotprojektets övergripande tidtabell genomförs.

- **Kriteriet innefattar smärre avvikelser**

Pilotprojektet förutsatte hantering av en omfattande helhet, där myndigheternas beredningsarbete, transporter av material samt digitaliseringens produktion måste samordnas för att materialet skulle fås till digitaliseringen vid rätt tid och berett på rätt sätt. I pilotprojektet omfattade denna helhet dessutom samtidigt uppbyggande av hela produktionsmiljön från början samt rekrytering och introduktion av personal till sina uppgifter.

På det stora hela höll pilotprojektets tidtabell och produktionskedet förlöpte till och med bättre än väntat. Myndighetsberedningen under hösten 2018 samt skanner- och programanskaffningen och genomförandet av servermiljön under våren 2019 genomfördes med en mycket snäv tidtabell. Särskilt det sistnämnda momentet omfattade stora risker för uppbyggandet av systemet och implementeringsgodkännandet inom den planerade tidtabellen. Trots detta inleddes digitaliseringsskedet mindre än två veckor senare än vid den planerade tidpunkten. Den manuella beredningen av materialet inför skanning inleddes redan i april.

Utrustningen fungerade väl, beredningen av materialet och skanningen framskred huvudsakligen på planerat sätt och inom tidtabellen. Produktionsplaneringen fungerade effektivt och flexibelt, utrustningen och personalresurserna användes effektivt och inga betydande flaskhalsar uppstod i produktionen. Den omsorgsfulla introduktion till materialet som genomfördes under planeringen var betydelsefull. Tack vare den gick myndighetsberedningen snabbare än väntat, den produktionsplanering som gällde materialet, dvs. styrningen av materialet till linjenät i enlighet med olika beredningssätt, gick huvudsakligen på det sätt som hade planerats. Pilotprojektets delprocesser

fungerade huvudsakligen på det planerade sättet och inga betydande ändringar behövde göras i dem. Processerna utvecklades och finjusterades i liten skala vid behov.

De tekniska störningar som förekom i servermiljön och de åtföljande produktionsavbrotten fördröjde dock digitaliseringen något. En hög andel material förkastades i valideringen, det fanns ett behov av att skanna material på nytt på grund av bilder som fattas efter enskilda programfel och bristen på ett användargränssnitt för digitala bilder medförde utmaningar. För att en smidig övergripande process som framskrider väl ska kunna uppnås finns det i sin helhet mycket att finslipa i delprocesserna, särskilt vad gäller återhämtningen från undantagssituationer.

12.2 Övriga resultat och observationer

Katalogisering av material: utfall

Katalogiseringen av material som en del av myndighetsberedningen lyckades väl. Situationen var krävande, eftersom AHAA-tjänsten togs i produktionsbruk på samma gång som myndighetsberedningen inleddes. En del av materialet matades in manuellt i programmet, en del exporterades via gränssnitt. I det senare fallet var man tvungen att planera överföringen från fall till fall, eftersom färdiga kravspecifikationer inte fanns tillgängliga. Med beaktande av hur ny tjänsten är lyckades katalogiseringen mycket väl.

I AHAA-tjänsten förekom det fel i hur noteringarna om materialets användningsbegränsningar visades, trots att de hade matats in korrekt i samband med katalogiseringen. Därför ska tjänsten ännu utvecklas. Målet med den digitalisering som sker efter pilotprojektet är att ta i bruk den myndighetsversion av AHAA-tjänsten som är under beredning och som är enklare att använda än den nuvarande.

Funktionen och tydligheten för Riksarkivets myndighetsstyrning

Den styrning som har utarbetats för myndighetsberedningen fungerar väl. En enkät om detta utarbetades i pilotprojektets digitaliseringsskede, dvs. efter att beredningen hade tagit slut. Utifrån enkätens resultat kommer styrningen att utvecklas ytterligare. Enligt enkäten uppskattades speciellt den styrning som genomfördes vid ämbetsverksbesök och via e-post. AHAA-tjänstens tillgänglighet sågs som ett centralt utvecklingsmål. Medeltalet för bedömningarna av instruktionerna och styrningen var 4 (På skalan 1–5, Dåligt–Bra).

De nyttjade programmens omfattning och funktion

De skanningsprogram som användes vid pilotprojektet fungerade som en helhet. Med programmen kunde man tillförlitligt producera bilder i enlighet med kravspecifikationerna (kriterierna för bildkvalitet och långtidsförvaring). Små korrigeringar av programkoderna under produktionen kunde göras smidigt och genom gott samarbete med underleverantören, huvudsakligen via distansförbindelse.

De största utmaningarna berodde på fel som inte i sig producerade en bildkvalitet som inte uppfyllde kriterierna, utan på grund av vilka en del av bilderna inte bildades överhuvudtaget eller deras läsriktning var fel ur innehållsidentifieringens synvinkel. Felen gällde endast dokumentskannrar. Den automatiska vändningens pålitlighet ledde framför allt till att man var tvungen att bereda materialet manuellt i rätt läsriktning för skannrarna, vilket gjorde jobbet långsammare. Att kunna identifiera enstaka bilder som fattas i efterhand kräver tidskrävande granskningar eller att man skannar materialet inom vissa intervall på nytt.

På alla apparatmodeller vändes bilderna i huvudläsriktningen, för en del automatiskt, för en del i samband med den manuella beredningen. På en av apparattyperna gjordes den automatiska vändningen genom en efterhandsprocess vilket ledde till att bilden måste packas om en gång.

Även en del utmaningar med servermiljöns funktion förekom. I värsta fall ledde detta till och med till avbrott på hela dagar i produktionen, eftersom bilderna inte kunde exporteras till överföringstjänsten. I disksystemen uppdagades även ett fel på grund av vilket samtliga filtyper inte gick vidare i överföringstjänsten. I värsta fall kan detta leda till att man måste skanna på nytt.

En stor del av de ovan beskrivna felen skulle troligtvis ha kunnat hindras om man hade genomfört en längre testperiod före digitaliseringen och om digitaliseringens produktionspersonal skulle ha haft direkt åtkomst att granska det digitala materialet efter valideringsskedet.

Dataöverförings- och datahanteringskapacitet

Vid pilotprojektet användes en dataöverföringskapacitet på 10 gbit, vilket inte ledde till någon som helst fördröjning i dataöverföringen. Exakta överföringshastigheter kan fås endast av en nätleverantör.

Infrastrukturen genomfördes som virtualserverlösning, vilket allmänt taget var en lyckad lösning vad gäller systemets funktion. I praktiken överfördes data från skannrarna direkt till den virtuella servermiljön. Antalet servrar och diskutrymmet har ökats så småningom

efter behov genom att granska kapacitetens tillräcklighet. Uppskattningen av behovet av diskutrymme (23–25 TB) har räckt väl för utfallet (14,4 TB). Vad gäller OCR beslöt man sig för en helhet på 12 servrar. Den nuvarande lösningen kan även skalas för större mängder. De största utmaningarna har haft att göra med uppföljning och hantering, vilket innebär att en mer exakt tjänstebeskrivning och definiering av rollerna än de som nu genomfördes skulle ha förutsatts.

Observationer om olika materialtyper samt om utrustningens lämplighet och kapacitet

Precis som väntat och planerat kunde digitaliseringen av A4-material av jämn kvalitet genomföras allra snabbast. Under den tid det tar att digitalisera material bör man inte enbart följa med skanningshastigheterna, utan koncentrera sig mer på den arbetstid som täcker hela digitaliseringsprocessen. Största delen av tiden går åt till att bereda materialet, särskilt till att avlägsna gem, dekalark och tejp, medan hastighetsskillnaderna mellan de viktigaste skannertyperna inte avviker stort från varandra. Detta beror speciellt på att inga av skannrarna kan köras med full effekt, vilket innebär att de snabbaste apparaternas hastighetsfördel i förhållande till de andra apparaterna jämnas ut. I framtida anskaffningsbeslut, när man väger högrepresterande och dokumentskannrar mot varandra, är det avgörande hur väl man kan gå igenom materialet på förhand, identifiera det och klassificera det för rätt typ av apparat.

Under pilotprojektet lyckades skanningen av så kallade "separerade ark" bland materialet. På grund av egenskaperna hos dessa var man tvungen att temporärt separera dem från det övriga materialet och skanna dem separat med specialskanner (mastskaner). Till pilotprojektet försökte man få material som innehöll rikligt med ark för specialskanner, men under myndighetsberedningen sållades den största delen av detta material bort, eftersom det ska förvaras på viss tid. På grund av specialmaterialets begränsade mängd kan man utifrån pilotprojektet inte dra heltäckande slutsatser om hur mycket material som behöver specialbehandling en digital helhet maximalt får innehålla, innan det på grund av den tid som behövs för arbetet utesluts från massdigitaliseringen.

Materialtyper vars behandling kräver rikligt med avvikande åtgärder, specialåtgärder eller åtgärder som kräver särskild försiktighet hade uteslutits ur massdigitaliseringen och således även ur pilotprojektet. I pilotprojektet ingick 60 band som digitaliserades antingen med mastskaner genom att i samband med digitaliseringen vid behov öppna upp bandstrukturen, eller genom att ta isär banden och digitalisera materialet som lösa ark. Utifrån pilotprojektets resultat lämpar sig band som måste tas isär helt för hand samt band som är svåra att klippas med elektrisk sax inte för massdigitalisering på grund

av att arbetet är för långsamt. I fortsättningen kommer man att bedöma huruvida band bör uteslutas ur massdigitaliseringen. Om ett band huvudsakligen kan tas isär med elektrisk sax, eller om eventuell informationsförlust ur marginalen på grund av klippning kan godkännas, kan bandet inkluderas i massdigitaliseringen.

Vid skanningen av band är det även avgörande huruvida man använder en traditionell mastskanner, för vilken hastigheterna varierar mellan olika apparater, eller mer skräddarsydda apparater som lämpar sig bättre för detta slags specialmaterial. Hanteringen av dessa skannrar och flyttningen av materialet tar märkbart med tid, vilket har en betydelse då man skalar för stora mängder.

13 Slutsatser och fortsatta åtgärder

Massdigitaliseringens pilotprojekt genomfördes 2018–2019 i form av en planerings- och produktionshelhet på cirka ett och ett halvt år. Digitaliserings-, dvs. produktionskedet tog cirka sex månader.

Pilotprojektet uppnådde sitt primära mål, dvs. bevisade att den planerade massdigitaliseringsprocessen och genomförandemodellen fungerar och att digitaliseringen kan genomföras med den planerade effektiviteten. Den digitalisering som genomfördes i projektet följde kriterierna för digitalisering med avsikt att förstöra, vilket möjliggör förstöring av de analoga handlingarna efter karenstiden.

Pilotprojektet visade att samordnandet av massdigitaliseringens helhet fungerade i praktiken. Genom myndighetsberedningen säkerställde man att materialet överförs vid rätt tidpunkt och är rätt förberett för digitaliseringen. Man har tillförlitligt kunnat följa med hela processen och överföringen av materialet från ett skede till nästa, trots att man inte ännu har ett system för produktionsstyrningen som täcker samtliga skeden.

Massdigitaliseringens pilotprojekt uppfyllde de centrala kriterier som hade fastställts och visade att den referensmetod som har planerats för massdigitaliseringen fungerar i praktiken. I pilotprojektet lyckades man uppnå den eftersträvade kvalitetsnivån för hanteringen av helheten, digitaliseringsproduktionens effektivitet samt den kvalitetsnivå som krävdes. Utifrån pilotprojektet kan massdigitaliseringen framskrida till produktion med den nuvarande referensmetoden som grund.

Pilotprojektet kunde i huvuddrag genomföras på det sätt som hade planerats, även om planeringen och genomförandet av vissa delområden genomfördes med en mycket snäv tidtabell. Efter pilotprojektets slut ska delprocessernas processbeskrivningar och arkitekturbeskrivningen uppdateras och de delar av dessa som bör utvecklas utvärderas.

Pilotprojektets dagsspecifika framskridande motsvarade relativt väl de uppställda målen. Vid planeringen och resursförsörjningen av framtida digitalisering ska effekten av frånvaro och beräkningen av den faktiska arbetstid som används för digitaliseringen beaktas mer noggrant då man gör resurser tillgängliga. De referenssiffror som behövs för detta fås från pilotprojektets resultat.

Pilotprojektet visade att kännedomen om materialet, dvs. färdplansplaneringen, återbetalar sig i digitaliseringsskedet. Myndighetsstyrningen och hörandet av myndigheternas synpunkter och respons lägger en grund för uppbyggandet av en fungerande massdigitalisering. Ju bättre man känner till materialet, desto enklare är det att styra myndighetsberedningen och att planera och genomföra digitaliseringen. Efter pilotprojektet är ett viktigt utvecklingsmål att precisera materialklassificeringen utifrån pilotprojektets observationer. Detta har även direkta effekter på kostnads kalkylens tillförlitlighet. Materialklasserna bör utvecklas och i sista hand är ett av målen att dra upp linjerna för vilka material som på grund av sina fysiska egenskaper och den åtföljande långsamma beredningen och skanningen inte lämpar sig för att digitaliseras genom massdigitalisering.

Följande utvecklingsmål observerades i pilotprojektet:

1. I pilotprojektet gjordes en del externa anskaffningar, en del anskaffningar från CSC Oy och en del genomfördes internt vid Riksarkivet. Helheten var fungerande, men förutsatte omfattande planering. Om digitaliseringen genomförs som ett gemensamt projekt mellan olika leverantörer ska den grunda sig på omsorgsfulla tjänstebeskrivningar, roller och ansvarsdefinitioner samt plattformar på vilka man delar dokumentation och material, och på beslut om vilka personer som samordnar helheten.
2. Digitaliseringsprocessen ska så långt som möjligt grunda sig på automatiserade funktioner. Medan man i pilotprojektet och den tidigare Proof of Concept-testningen fokuserade på skanner- och programlösningar och klassificering av materialet, bör man i den framtida planeringen dessutom inkludera en tätare integrering av produktionsstyrningen i systemet samt utvidga produktionsstyrningen till massdigitaliseringens samtliga skeden.
3. Från första början var avsikten att genomföra pilotprojektet så att man först skapar en utvecklingsmiljö eller ett laboratorium för produktionen, på vilken den egentliga produktionen byggs upp. På grund av tidtabellen genomfördes produktionsmiljön i pilotprojektet. I framtiden ska produktionen alltid vila på en fullskalig testmiljö som används parallellt och omfattar digitaliserings samtliga delprocesser från början till slut. Ett centralt utvecklingsmål är byggandet av en fullskalig testmiljö för pilotprojektet.

4. Styrningen för pilotprojektets digitaliseringsarbete samlades i ett heltäckande instruktionspaket. I fortsättningen bör den helhetsmässiga dokumentationen främjas ytterligare så att den täcker även den dagliga produktionsstyrningen bättre, och så att man kan säkerställa och enkelt tillgängliggöra resultat- och personalstyrningens funktion samt en aktuell lägesbild i alla skeden. I all styrning ska rollfördelningen, uppdragen och ansvaren beskrivas omsorgsfullt. De mått som gäller produktionens centrala mål ska definieras på förhand och i programmen eller de appar som stöder dem ska man bygga rapporteringsegenskaper, ur vilka de centrala parametrarna kan samlas automatiskt i batchform vid överenskomna tidsintervall.