



Sveriges geologiska undersökning

Pilotprojekt Geodigitalia

för digitalisering av analoga geodata vid
SGUs arkiv vid Mineralinformationskontoret i Malå

Helge Reginiussen & Lisbeth Hildebrand

Pilotprojekt Geodigitalia för digitalisering av analoga geodata
vid SGUs arkiv vid Mineralinformationskontoret i Malå

SGU-rapport
2006:2

SGU-rapport
2006:2

SVERIGES GEOLOGISKA UNDERSÖKNING
Mineralinformationskontoret

Projektets namn: Geodigitalia	Datum: 2006-02-20	Dokumentnr: SGU rapport 2006:2
Projektbeställare: Malå kommun	Projektledare: H. Reginiussen & L. Hildebrand	Forsknings- och utvecklingschef: L. Persson

Digitalisering av analoga geodata vid SGUs arkiv vid Mineralinformationskontoret i Malå



Sammanfattning	5
Allmänt	6
Bakgrund och syfte.....	6
Bemanning.....	6
Praktiskt genomförande.....	6
Utvärdering av skannrar	8
Skanner i pilotprojektet	8
Skannrar och datorer till ett framtida digitaliseringsprojekt	8
Utvärdering av programvaror	10
Adobe Acrobat (PDF).....	10
Document Express with DjVu.....	10
Omnipage Professional 15.....	11
Photoshop Elements.....	12
MS Office.....	12
Kartex.....	12
Material för digitalisering	13
Prospekteringsrapporter	14
Översikt rapportslag och tidsperioder.....	15
Inbindning.....	15
Resultat och datalagringsbehov.....	18
OCR-konvertering.....	19
Tabellstruktur och sökmöjligheter i befintligt rapportregister.....	20
Rekommendation.....	21
Markgeofysiska kartor och data	22
Beskrivning.....	22
Testresultat.....	26
Utvärdering av alternativ.....	27
Förslag till metadatastruktur.....	28
Markgeofysiska data i pärmar.....	28
Datalagringsbehov.....	28
Rekommendation.....	29
Borrhålsprotokoll och kemiska analyser av borrhårnor	30
Beskrivning och skanning.....	30
Resultat av OCR-konvertering.....	31
Rekommendation.....	34
Geologiska kartor	35
Resultat av skanning.....	36
Förslag till metadatastruktur.....	36
Datalagringsbehov.....	36
Utvärdering av alternativ.....	36
Rekommendation.....	38
Borrhålskartor	39

Resultat av skanning	39
Metadata	39
Datalagringsbehov	39
Geokemiska data av olika typer samt kartor	40
Beskrivning och resultat av skanning	40
Resultat av OCR-konvertering	40
Datalagringsbehov	41
Metadata	41
Rekommendation	41
Block- och hälldata samt fältdagböcker	42
Beskrivning och skanning	42
Rekommendation	42
Uranarkivsmaterial	44
Beskrivning	44
Resultat av skanning	44
Datalagringsbehov	45
Rekommendation	45
Gruvkartor samt arkivakter med handlingar vid Bergsstaten	46
Övrig prospekteringsintressant information	47
Sammanfattning av arbetstid och kostnader	48
Tidsuppskattning av olika arbetsmoment	48
Kostnader	50
Sammanfattning av datalagringsbehov	52
Sammanfattande rekommendationer för digitalisering av hela arkivet	53
Personal	53
Prioriteringsordning och arbetsplan	53
Beroenden	54
Risker	54
OCR	54
Bilaga 1: Testskanning av markgeofysikmaterial	55
Resultat	56
Bilaga 2: Metadatastruktur för markgeofysikmaterial	57
Bilaga 3: Programvara	58
Document Express with DjVu	58
Beskrivning av Document Express with DjVu Enterprise Edition	58
Bilaga 4: Skannrar och annan hårdvara	61
A3 skannrar	61
Bilaga 5: Digitaliserat material i Geodigitalia	62

Prospekteringsrapporter	62
Geokemianalyser	64
Block-och hållpärmar	64
Uranpärmar	64
Projektpärmar	64
Geologi och borrhålskartor	64
Borrhålsprotokoll.....	64

Sammanfattning

Pilotprojektet Geodigitalia har utvärderat möjligheten för digitalisering av olika typer av SGUs analoga data vid Mineralinformationskontoret i Malå. De viktigaste slutsatserna av projektet kan sammanfattas i följande punkter:

- Huvuddelen av digitaliseringen kan göras på plats i Malå, medan digitalisering av svårdigitaliserat material bör utföras av extern digitaliseringsbyrå.
- Tidsförbrukningen för digitalisering av hela arkivet anslås till ca 41 500 timmar, exkl. tid för extern digitaliseringsbyrå och utbildning av personal.
- Totalkostnaden för digitalisering av hela arkivet anslås till ca 19 200 000 kr.
- Filtypen PDF bör användas som standardformat för skrivna dokument. För att underlätta dokumentdistribution kan man dessutom använda filformatet DjVu.
- Datalagringsbehovet för de tio huvudtyperna av data anslås till att vara i storleksordningen 14 Terabyte.
- Det rekommenderas att en personalstyrka på ca sex personer arbetar heltid i projektet samt projektledare 75 %. Projektet kan då genomföras på ca 4 år.

Allmänt

Bakgrund och syfte

En av SGUs viktigaste uppgifter är att möta efterfrågan på geologisk information från samhället. SGU skall också enligt sin instruktion marknadsföra sin geologiska information och verka för att den snabbt tillgängliggörs.

SGUs Mineralinformationskontor i Malå är ett riksarkiv för prospekteringsinformation och kontoret tillhandahåller geoinformation till prospekterings- och gruvföretag aktiva i Sverige.

De dokument som nu finns i arkivet är till största delen analoga d.v.s. rapporter, analysresultat och kartor är tryckta på papper. Det har länge funnits ett önskemål, både från internt och externt håll om att denna information skall digitaliseras. Detta för att bl.a. bättre tillgängliggöra SGUs data för moderna prospektörer som använder GIS-teknologi och databashanterare. Dessutom representerar ett digitalt arkiv en ”back-up” av det analoga materialet och är därmed viktig ur säkerhetssynpunkt.

Målet med förstudien Geodigitalia har varit att tids- och kostnadsberäkna digitalisering av hela arkivet i Malå. Projektet ger även utifrån genomgång av arkivets olika material förslag på hur detta prospekteringsintressanta analoga material i sin helhet bör digitaliseras i ett framtida projekt med hänseende till programvaror, kompetenser, prioriteringsordningar etc.

Projektet har även påbörjat en viss del av digitaliseringen.

Bemanning

Statsgeolog Helge Reginiussen, SGU Malå har varit projektledare från projektstart 2005-08-15 till 2005-11-14 då han blev föräldraledig.

Vikarierande projektledare har varit statsgeolog Lisbeth Hildebrand, SGU Malå fr.o.m. 2005-11-14 t.o.m. 2006-02-28.

Josefin Perdahl och Linn Ulwebäck har varit projektanställda medarbetare fr.o.m. 2005-08-15 t.o.m. 2005-12-31.

Statsgeofysiker Olof Forslund och Jerry Hedström (borrkärneansvarig) har arbetat ett mindre antal timmar i projektet.

Praktiskt genomförande

Materialet i arkivet omfattar prospekteringsrelevant information som är insamlat över en period om mer än hundra år. Det är mycket heterogent och består av en rad olika typer av dokument, format, kvaliteter osv. Det finns dessutom dubletter och material som måste sorteras bort innan digitalisering kan påbörjas. En genomgång av delar av materialet har varit nödvändig för att få en översikt över vad som finns och i vilket skick materialet är. Detta arbete, som är mycket tidskrävande, har prioriterats i

pilotprojektet därför att det är en absolut förutsättning för att kunna göra en rimlig tids- och kostnadsuppskattning av digitalisering av hela arkivet.

Det har utförts testskanning av olika dokumenttyper för att estimeras tidsförbrukning och datalagringsbehov. Resultaten från testskanningen har därefter extrapolerats för att uppskatta totalbehovet för hela arkivet.

Testskanningen har också haft som mål att hitta optimala skanningsrutiner, inställningar, skanningsupplösning och kvalitet. Programvaror för textigenkänning (OCR) har utvärderats.

Det blev på ett tidigt stadium genomfört ett möte med representanter från IT avdelningen och ECM förstudien vid huvudkontoret i Uppsala för att presentera pilotprojekt Geodigitala och visa exempel på analoga data från pappersarkivet i Malå. Det blev på detta möte i Uppsala 2005-09-07 diskuterat möjliga strategier för filformat, metadata etc. En konklusion från mötet var att filtypen PDF eller liknande kan vara basis för textdokument i ett framtida digitalt arkiv. Om digitala dokument finns i detta format finns det möjlighet till konvertering till andra filformat om behov skulle uppstå vid en senare tidpunkt.

Utvärdering av skannrar

Skanning är grunden för digitaliseringen av arkivet. På marknaden finns specialiserade dokumentskannrar som är kapabla att snabbt hantera stora volymer med heterogen papperskvalitet och storlek. SGU kan upphandla skanningsutrustning till fördelaktiga priser via VHS avtalet.

Skanner i pilotprojektet

Huvuddelen av materialet som skall digitaliseras är svart-vita dokument i A4 format. Det finns emellertid också färgmaterial (foton, bilagor, kartor m.m.) samt större format som gör att en skanner till pilotprojektet måste kunna klara av A3 format och färg. Till testskanning av A1-A0 material har en A0 skanner, Colortrac Smart LF 4080 använts, som finns vid huvudkontoret i Uppsala.

Efter att ha studerat ett urval av skannrar från leverantörer som är med i VHS avtalet blev skannern Ricoh Aficio IS 300e (fig.1) vald.



Figur 1. Dokumentskanner som används i Geodigitalia.

Denna skanner har de funktioner som vi efterfrågar i pilotprojektet: nätverksanslutning, A3, färg, gråskala, svart-vitt, automatisk skanning (ADF), manuell skanning samt möjlighet för att skanna ett flertal papperskvaliteter. Beställning blev gjord via inköpsansvarig vid SGU. Skannern blev inköpt av SGUs filial i Malå och har belastat Geodigitalias budget med endast en internhyra på 830 kr/mån.

Ett problem som medförde att progressionen i skanningen blev något försenad var att leveransen av skannern tog längre tid än planerat. Den blev beställd från NRG Scandinavia den 16 september och blev levererad till filialen i Malå den 11 oktober. Skannern var operativ den 19 oktober.

Skannrar och datorer till ett framtida digitaliseringsprojekt

När det gäller val av hårdvara till ett fullskaligt digitaliseringsprojekt så ställs det följande krav på utrustningen:

Antal	Hårdvara	Användningsområde
1	A0 skanner (Färg).	<ul style="list-style-type: none"> • Geologiska kartor • Geofysiska kartor • Geokemiska kartor • Uranrelaterade kartor
2	Dokumentskanner (Svart-vitt eller färg A4/A3). Hög hastighet. Bör klara av blandad papperskvalitet.	<ul style="list-style-type: none"> • Prospekteringsrapporter • Borrhålsparmar • Geokemiska analyser
1	Dokumentskanner (Färg- A3). Hastighet ej avgörande.	<ul style="list-style-type: none"> • Bilagor till prospekteringsrapporter • Kartor i A3 format etc.
6	Datorer med utökad lagringskapacitet.	<ul style="list-style-type: none"> • Kontroll av skannade dokument • Överföring av filer • Produktion av dokument • Metadatasättning av kartor

Tabell 1. Hårdvarubehov till ett fullskaligt digitaliseringsprojekt.

Med dagens prisnivå så representerar denna utrustning en total kostnad på ca 435 000 kr (se sid. 50 för specificerad översikt). Hårdvarupriser är något som ständigt sjunker så denna uppskattning är sannolikt ett maximum-estimat. Det bör dock inberäknas att det måste företas uppgraderingar av t.ex. datorutrustningen under projektets levnadstid som antas att bli flera år.

SGU har genom VHS avtalen åtta avtalsleverantörer av skannrar att välja från. En förteckning över dessa finns på:

http://www.avropa.nu/templates/ramavtalsomrade___86.aspx

När det gäller datalagringskapacitet så kommer det vid ett fullskaligt digitaliseringsprojekt bli behov för att utöka diskutrymmet till en nivå som kan hantera de stora datavolymer som projektet kommer att generera (se sid. 52).

Kostnaden för denna utrustning är inte inberäknat. Det måste dessutom utarbetas en strategi för att ta back up av alla producerade filer.

Filialen i Malå har för 2006 fått finansiering för en datalinje mellan kontorsbyggnaden och pappersarkivet. Det betyder att skanningsutrustningen rent fysiskt kan placeras i pappersarkivet där de analoga data finns, vilket gör att det kan etableras en produktionslinje som inte kräver transport av materialet.

En fördel med att behålla materialet under samma tak är att pappersarkivet är funktionellt under skanningsfasen. Dessutom så riskerar man inte att materialet förkommer, förstörs eller skadas – något som kan inträffa om materialet skall skickas iväg till annan ort för digitalisering.

Utvärdering av programvaror

Det är utfört testskanning med olika programvaror och en utvärdering av olika typer av filformat. När det gäller rena textsidor så vill ett optimalt resultat vara produktion av sökbara filer. Detta gör att Adobe Acrobat (PDF) och Lizardtech Document Express with DjVu är möjliga kandidater.

Vad gällande att ta ut kemiska analyser i tabellform ur skannade dokument och OCR-konvertera samt sammanställa dessa i databaser har vi testat programmen Omnipage Professional 15, Adobe Acrobat 7 och Document Express with DjVu. Utfallet av testerna går att läsa under rubriken ”Material för digitalisering”.

OCR (Optical Character Recognition eller optisk teckentolkning) används för att omvandla text från skannade sidor eller bilder till redigerbar text som sedan kan användas i vanliga datorprogram.

Adobe Acrobat (PDF)

PDF formatet (Portable Document Format) är en öppen standard som har stor utbredning när det gäller elektroniskt dokumentutbyte. Adobe® PDF används av företag, högskolor och myndigheter över hela världen. PDF-filer ser likadana ut som originaldokumenten, med samma teckensnitt, grafik, bilder och layout som i källfilen – oavsett vilket program och vilken plattform som användes för att skapa den. PDF dokument kan utväxlas, visas och skrivas ut av alla som har gratisprogramvaran Adobe Reader® installerad. PDF har innebyggda funktioner för textsökning.

För att kunna OCR-konvertera skannade PDF-dokument till sökbara PDF-filer krävs att man har minst Adobe Acrobat 6.0 installerat. Projektet beställde den 2005-12-20 Adobe Acrobat 7.0 och fick tillgång till programmet den 2006-02-01, något som har medfört att test av OCR funktionaliteten inte har kunnat göras så noggrant som planerat.

Document Express with DjVu

Document Express with DjVu är en applikation som kan användas för att göra kraftigt komprimerade dokument (inklusive färgdokument) i ett format som kallas DjVu. DjVu-formatet har mycket mindre utbredning än PDF, men formatet börjar få en viss utbredning, speciellt till dokumentdistribution via Internet.

Typiskt kompressionsförhållande för DjVu är mellan 300:1 och 1000:1. Ett A4 dokument i DjVu-format med 300-400 dpi upplösning vill typiskt ha en storlek på endast 50 KB. DjVu-dokumentbilder är uppemot 1000 gånger mindre än TIFF-filer, och mellan 10-100 gånger mindre än JPEG och PDF filer (tabell 2).

DjVu möjliggör goda OCR möjligheter för gråskaliga dokument och kan hantera text i färg och text som är tryckt på grafik eller bilder.

Testobjekt	TIFF	PDF	DjVu (inkl. OCR)
Geologisk karta över Svanaträsk	78 MB	1,68 MB	0,155 MB
Prospekteringsrapport brap 79056	-	12,07 MB	1,78 MB
Prospekteringsrapport brap 79008	-	9,31 MB	1,54 MB
Karta över Stekenjokkmalmen	153 MB	-	0,088 MB

Tabell 2. Exempel på filstorlekar i olika filformat.

För att kunna öppna DjVu-filer behövs en webläsare och en gratis plugin som kan laddas ned från LizardTech's hemsida (<http://www.lizardtech.com>). Denna plugin kan köras på flera webläsare bl.a. Microsoft Internet Explorer, Firefox och Safari. De Operativsystem som stöds är MS Windows, Mac-OS och Linux-Unix. För MS Windows är storleken på pluginen endast 900 KB något som gör att det går snabbt att ladda ner den.

Document Express with DjVu finns i två varianter: Professional Edition och Enterprise Edition. I pilotprojektet blev det laddat ner en testversion av DjVu Enterprise edition från LizardTech's hemsida. Testversionen kunde användas i 30 dagar och hade en begränsning på komprimering av 500 sidor. En utvärdering av programmet blev gjort på befintliga digitala data från arkivet i Malå samt på data som är producerat av Geodigitalia.

I bilaga 3 finns en detaljerad beskrivning av programmet samt uppgifter om pris och svensk återförsäljare.

Efter att ha provat programmet under testperioden så är konklusionen att DjVu är ett mycket intressant filformat som har många fördelar. Framför allt så är komprimeringsförmågan imponerande. Detta möjliggör snabb och effektiv dokumentdistribution kombinerat med att DjVu filerna tar liten diskplats. En liten nackdel är möjligtvis att själva konverteringsprocessen är tidskrävande. Det finns dock möjlighet för "batch" processering så att man kan låta programmet köra stora konverteringsvolymmer t.ex. på natten.

Tillsammans med PDF så vill filformatet kunna utgöra en mycket stark kombination. Ett arkiv baserat på PDF i botten och med DjVu som alternativt distributionsformat vill tillgodose både kompatibilitetskrav ("alla" kan använda PDF filer) och behov för snabb dokumentdistribution via DjVu. Till distribution av dokument via web är filformatet mycket användbart.

Omnipage Professional 15

Programmet är ett textigenkänningsprogram som har många användbara funktioner. I Omnipage Professional 15 kan man t.ex. OCR-konvertera PDF-filer innehållande

tabeller och spara dessa som MS Excel, något som är aktuellt för projektet. Efter att ha testat programmet en kortare period kan man sammanfatta att funktionerna för att förbättra bilden (sidor i PDF-dokumentet) är mycket användbara när det gäller bl.a. analystabeller tryckta på äldre datautskriftspapper, där man vill filtrera bort en randig bakgrund. Programmet känner automatiskt av var kolumn- och radavskiljarna skall vara under konverteringen och avskiljarna går även att manuellt placeras om, läggas till eller tas bort.

Photoshop Elements

För att kunna beskära och rotera kartor som skannas i en A0-skanner kan Adobe Photoshop eller Adobe Photoshop Elements användas.

MS Office

MS Excel är nödvändig för att kunna spara och kontrollera konverterade analysfiler.

Kartex

Programmet Kartex underlättar vid metadatasättning av kartor, när kartorna inte alltid innehåller all behövlig geografisk information.

Material för digitalisering

I arkivet finns material som har tagits fram under den statligt finansierade prospekteringen de senaste hundra åren samt material från privat prospektering av nyare dato. Informationen täcker alla aspekter av modern prospektering och är insamlad av bl.a. SGU, SGAB, NSG, LKAB och andra prospekterande bolag.

Det analoga materialet som skall digitaliseras är av minst tio olika typer. Materialet kan indelas i:

1. Prospekteringsrapporter samt bilagor i A3- till A0-format; kartor m.m.
2. Markgeofysiska kartor och data
3. Borrhålsprotokoll och kemiska analyser av borrhållar
4. Geologiska kartor
5. Borrhålskartor
6. Geokemiska data av olika typer samt kartor
7. Block- och hålldata samt dagböcker
8. Uranarkivsmaterial
9. Gruvkartor samt tillhörande arkivakter med handlingar från Bergstaten
10. Övrig prospekteringsintressant information

Papperskvaliteten och tryckkvaliteten på dokument och kartor är ojämn. Den stora variationen med hänsyn till format och kvalitet ställer stora krav på den tekniska utrustningen. Dessutom så medför heterogeniteten att det måste utarbetas olika strategier för varje typ av data. Det krävs också mycket förarbete i form av genomgång och uppsortering före en digitalisering äger rum.

Prospekteringsrapporter

Material: Prospekteringsrapporter gjorda av bl.a. SGU, SGAB, NSG, LKAB. Antalet är ca 7000 st. och finns i pappersformat.

Totalt antal A4 sidor som skall digitaliseras: ca 250.000.

Totalt antal bilagor > A4 format: ca 30.000.

Målsättning: Att digitalisera alla rapporter med tillhörande bilagor för att öka tillgängligheten och användningen av materialet.

Antalet prospekteringsrapporter som finns vid arkivet i Malå (fig. 2) är f.n. 7012 st. Arkivet utökas fortlöpande allt efter som nytt material tillkommer. Under projektperioden tillkom till exempel flera rapporter från Outokumpus verksamhet i Sverige.



Figur 2. Prospekteringsrapporter i pappersarkivet i Malå.

Översikt rapportslag och tidsperioder

Prospekteringsföretag

ID code/no	från	till
Sveriges Geologiska Undersökning (SGU) – prospekteringsrapporter		
brap 1–949	före 1975	
brap 75001–82616	1975	1982
fm 7801–8232	1978	1982
geok 79005–82028	1979	1982
bsrp 78001–80002	1978	1980
Sveriges Geologiska AB (SGAB) – prospekteringsrapporter		
prap 83001–92513	1982	1992
Nämnden för Statens Gruvegendom (NSG) – prospekteringsrapporter		
NSG 86109–93009	1986	1993
Sveriges Geologiska Undersökning (SGU) – geologiska rapporter		
brap 83401–98032	1983	1998
brap 80001–80010	1980	1980
LKAB-Prospektering Stockholm		
pro 7801–8342	1978	1983
s 8401–8708	1978	1983
Bergslagen		
grb 1–300	1977	1983
bsg 300–380	1983	1984
b 8501–9110	1985	1992
Kiruna		
ki 7701–8609	1977	1986
k 8610–9004	1986	1990
Övriga rapporter vid Mineralinformationskontoret i Malå mink 95001–99010	1995–	

Rapporterna finns även i ett sökbart register (dBase-format) på filialens server med sökväg:

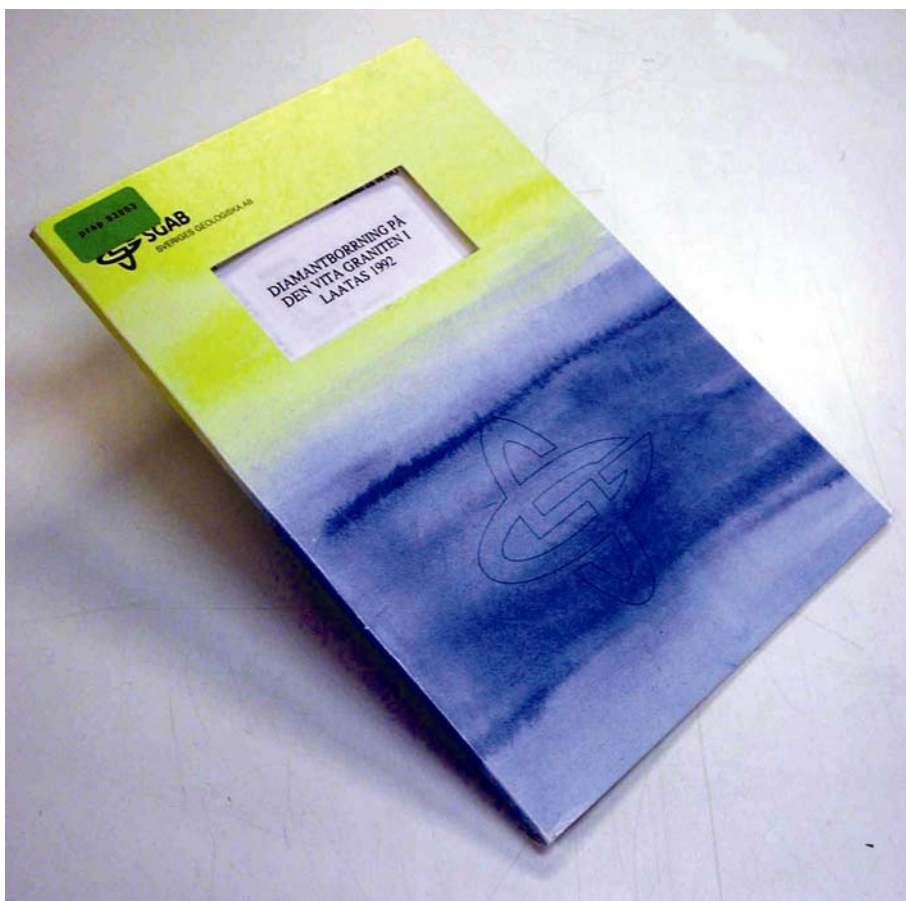
Malå/Server/Register/Reports/Rapreg21.DBF

Inbindning

Rapporternas fysiska skick och inbindning är ett moment som måste beaktas i ett kommande storskaligt digitaliseringsprojekt. Det är ett absolut krav att A4 och A3 sidor kan skannas i en dokumentkannare med ett automatiskt arkmatningssystem. Sidorna i inbundna rapporter måste därför tas ifrån varandra innan skanning kan påbörjas.

Prospekteringsrapporternas inbindning är av tre huvudtyper:

- Limmad kant (fig. 3). Rapporter av denna typ måste först tas isär med en inbindare (fig. 4). Limmet i rapportens rygg värms upp så att sidorna lossar och kan tas isär. Denna procedur tar ca 1 minut. Dessutom måste kanten på sidorna beskäras och detta tar 1-5 min beroende på rapportens tjocklek.

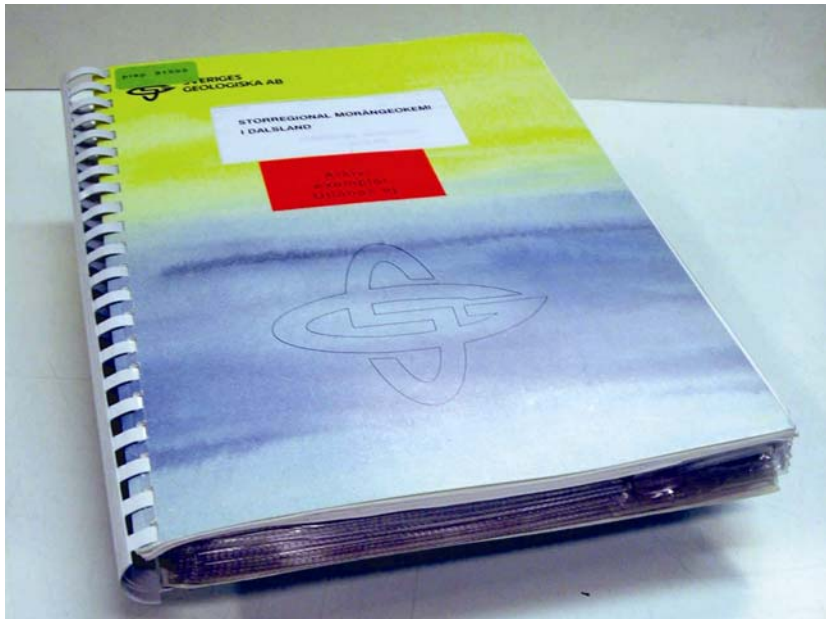


Figur 3. Exempel på prospekteringsrapport med limmad rygg från arkivet i Malå.



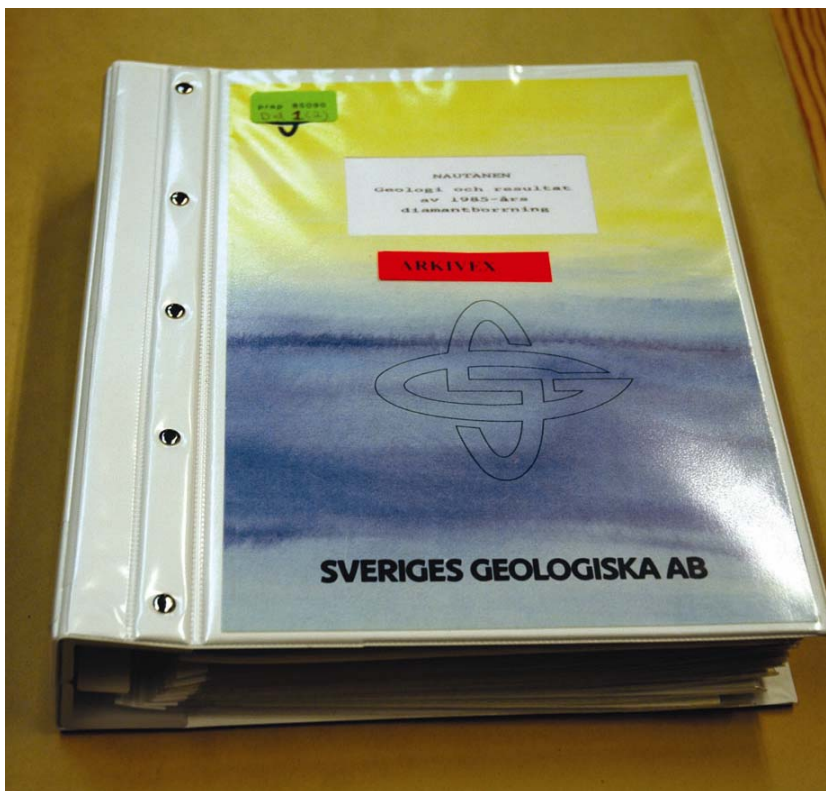
Figur 4. Inbindare (vänster) som används till att ta isär en rapport innan ADF-skanning. Till höger en öppnare för spiralinbindning.

- Spiralinbundna rapporter (fig.5). Dessa måste först tas isär med en speciell öppnare som viker upp spiralryggen (fig. 4). Detta tar ca en minut per rapport.



Figur 5. Prospekteringsrapport med spiralinbindning.

- Ringpärmsinbundna prospekteringsrapporter (fig. 7). Ofta innehåller dessa rapporter många bilagor med varierande storlek.



Figur 7. Ringpärmsinbunden rapport.

Efter skanning av rapporten så måste rapporten monteras ihop igen och detta tar 1-5 min. beroende på typ av inbindning och rapportens tjocklek.

För ett antal rapporter (ca 2 100 st.) finns det oinbundna original eller kopior av originalet (fig. 8). Dessa ägnar sig till automatiserad skanning i en dokumentskannare med arkinmatningssystem. Ett problem med dessa är emellertid att många inte är kompletta jämfört med de inbundna rapporterna, bl.a. så saknas bilagor. Detta medför att bilagorna måste skannas separat och infogas i de producerade digitala dokumenten.

Ett begränsat antal av dessa oinbundna rapporter är redan skannade i Geodigitalia.



Figur 8. Oinbunden prospekteringsrapport.

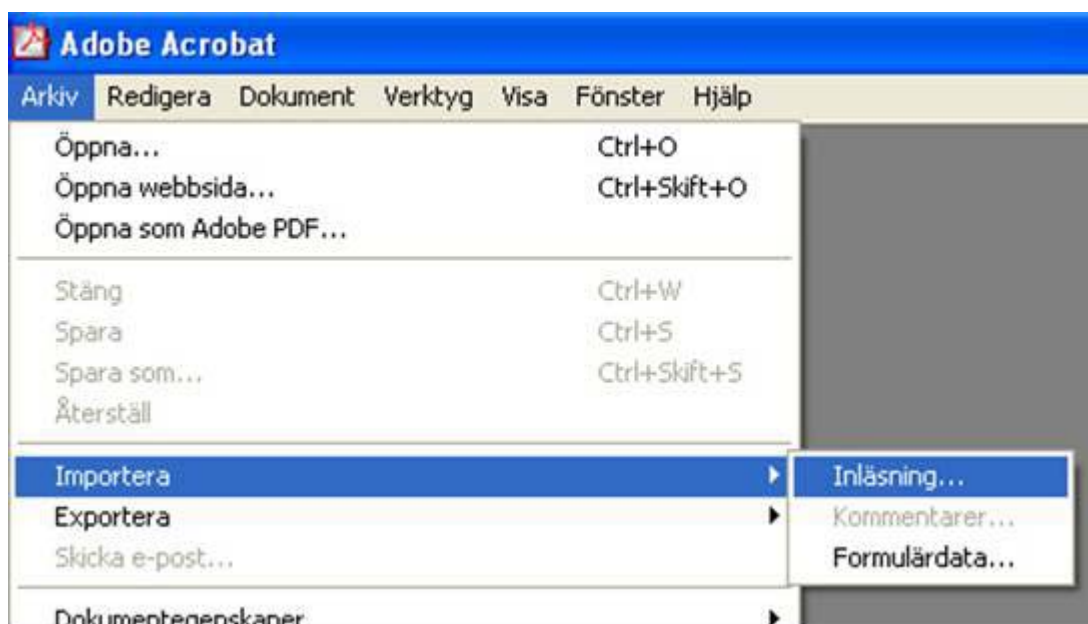
Resultat och datalagringsbehov

Det är utfört tester med olika upplösning och gråskala eller svart-vitt. Optimal kontrast på rapporttext för OCR och utskrift uppnås med svart-vitt och med en upplösning på 300 dpi. Skanning i svart-vitt går dessutom mycket snabbare än skanning i gråskala och filstorleken blir mindre. Figurer, kartor osv. måste dock i de flesta fall skannas i gråskala.

För att ytterligare effektivisera skanningsprocessen så kan man skanna direkt in i Adobe Acrobat via Arkiv - Importera - Inläsning (fig. 9). Efter inläsningen väljer man källa och skannerinställningar. Sidorna skannas nu fortlöpande med ADF och läggs automatisk in i PDF-dokumentet allt efter som de passerar igenom skannern.

Detta sker med en hastighet på många sidor i minuten (exakt antal sidor beror på vilken upplösning man väljer samt storlek, färg, svart-vitt, gråskala m.m.).

Snabb dokumentskanning betyder att flaskhalsen i PDF-filproduktionen av prospekteringsrapporter ligger i förberedelse och efterbehandling (separering av text och eventuella figurer, hoplimning osv.) av rapporterna. Dessutom så tar behandlingen av stora bilagor (A1-A0) tid.



Figur 9. Inläsning direkt i Adobe Acrobat från dokumentskannare.

Hittills i Geodigitalia så är 421 rapporter med alla bilagor skannade till PDF-filer (se bilaga 5). Detta utgör ca 6 % av alla rapporter i arkivet. Total filstorlek för dessa 421 rapporter är 9,18 GB. Genomsnittlig filstorlek är ca 21,8 MB.

Om man extrapolerar detta genomsnittsvärde så indikeras ett totalt diskutrymmebehov på 153 GB för alla 7000 prospekteringsrapporterna i arkivet (inklusive bilagor). Se också sammanfattning av datalagringsbehov på sidan 52.

OCR-konvertering

Prospekteringsrapporterna är en av de mest efterfrågade informationsslagen på Mineralinformationskontoret. Det kan vara till hjälp att det finns möjlighet för textsök, alltså möjlighet för att kunna skriva in ett sökord och få fram var någonstans i rapporten just ett specifikt ord finns.

I projektet har vi testat att göra de skannade PDF-rapporterna sökbara med programvarorna Adobe Acrobat 7.0 och Omnipage Professional 15. Resultatet, som är baserat på ett begränsat antal tester, indikerar att Omnipage har bättre textigenkänningsförmåga än Acrobat.

Omnipage separerar ut text ur bildobjekt i en högre omfattning än vad Acrobat klarar av. Dock har Adobe Acrobat en klar fördel eftersom rapporterna redan är i PDF-format vilket betyder att det är tidseffektivare när man ej behöver hämta in filen i ett

annat format. Dessutom går OCR-konverteringen nästan dubbelt så snabbt i Acrobat mot i Omnipage. Däremot blir den färdiga OCR-lästa PDF-filen sex gånger så stor i Acrobat mot PDF-filen sparad i Omnipage. Omnipage ändrar lite på utseendet på PDF-filen men detta är inte störande.

Även DjVu- formatet har ett mycket bra stöd för OCR och textsök inbyggt som kan inkapslas direkt i dokumentet när man konverterar från PDF till DjVu format.

Tabellstruktur och sökmöjligheter i befintligt rapportregister

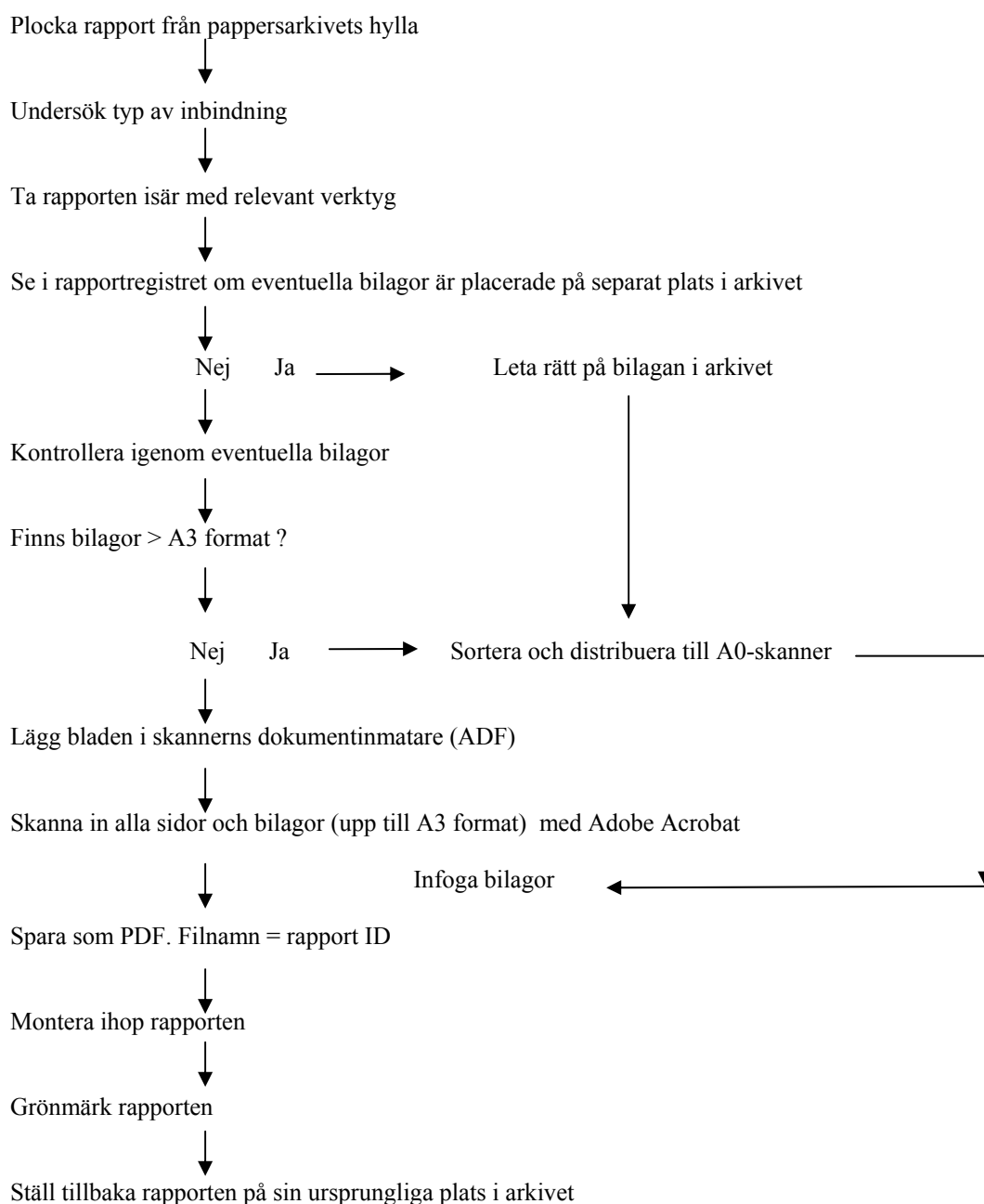
Rapportregistret som finns på filialens server har följande tabellstruktur:

ID_CODE	Indikerar företag eller organisation
ID_NO	Serienummer i varje rapportserie
ORDERER	Organisation som beställt arbetet
PROJECT	Projektets namn
TITLE	Rapporttitel
AUTHORS	Författare till rapporten
YEAR	År rapporten färdigställdes
MAP_SHEET	Visar kartblad berörda i rapporten
LOCALITY	Lokaler nämnda i rapporten
KEYWORDS	Nyckelord från rapporten
MAPS	Visar om kartor är inkluderade i rapporten
COUNTY	Visar län berörda i rapporten

Denna struktur kan användas som metadata till rapporterna i ett framtida digitalt arkiv. Om man återanvänder dessa data i stället för att göra nya metadata från grunden så kan stora besparingar göras. Ovanförnämnda databas representerar ca 85 000 datafält och om man antar att det tar ca 15 min. att göra metadata för varje rapport så vill det ta ca 1750 effektiva timmar att göra detta. Detta medför att man kan spara in mer än ett manår på att använda det existerande rapportregistrets tabellstruktur som metadata.

Rekommendation

Ett flödesschema för digitalisering av prospekteringsrapporter kan t.ex. se ut på följande sätt:



Digitalisering av prospekteringsrapporterna anses att vara förknippat med få problem. Materialet är i gott skick och det finns ett mycket bra sökbart register över alla rapporter i arkivet. Detta material är bland det mest efterfrågade vid Mineralkontoret och bör digitaliseras i ett tidigt skede av ett eventuellt storskaligt digitaliseringsprojekt.

Mer än 6% av alla rapporter är redan digitaliserade till PDF-filer i Geodigitalia. Ett mindre antal av dessa är dessutom också konverterade till DjVu-format (se bilaga 5). De olika textigenkänningsprogramvarorna har olika styrkor och svagheter för ändamålet att göra rapporterna sökbara.

Markgeofysiska kartor och data

Material: Analoga geofysikkartor med markmätningar (huvudsakligen A1) tryckta på papper och transparenta folier (translater). Materialet är heterogent både med hänsyn till metod, tryckkvalitet, typ av papper eller plastfilm, årtal för insamling osv.

Mätningarna är huvudsakligen av följande typer:

- 1) Magnetiska
- 2) VLF
- 3) Slingram
- 4) IP
- 5) Gravimetri

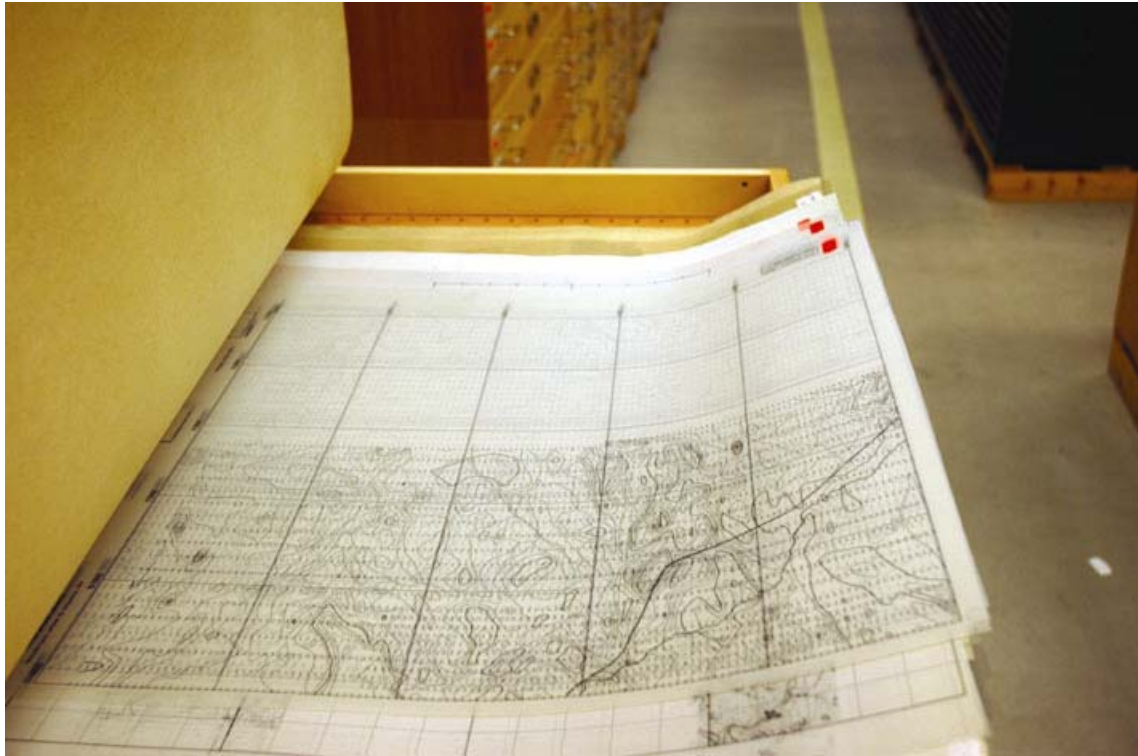
Målsättning: Att digitalisera mätvärden från kartor till digitalt format som kan sparas i en databas tillsammans med koordinatinformation. Dessa digitala data kan sedan användas i geofysiska applikationer och GIS-program. Det skall dessutom göras digitala kopior av alla kartor som skall sparas som bildfiler. Metadata skall antecknas för varje karta.

Beskrivning

Kartorna finns sorterade efter kartbladsnummer i 320 kartlådor i pappersarkivet i Malå (fig. 10, 11). Det finns i nuläget ingen exakt siffra på antalet markgeofysiska kartor i arkivet. I samband med pilotprojektet är det därför gjort en grov uppskattning baserat på räkning av polygoner och vid en extrapolation av innehållet i 20 slumpmässigt utvalda lådor.



Figur 10. Kartlådor med analog markgeofysik.



Figur 11. Kartlåda med analoga markgeofysiska kartor.

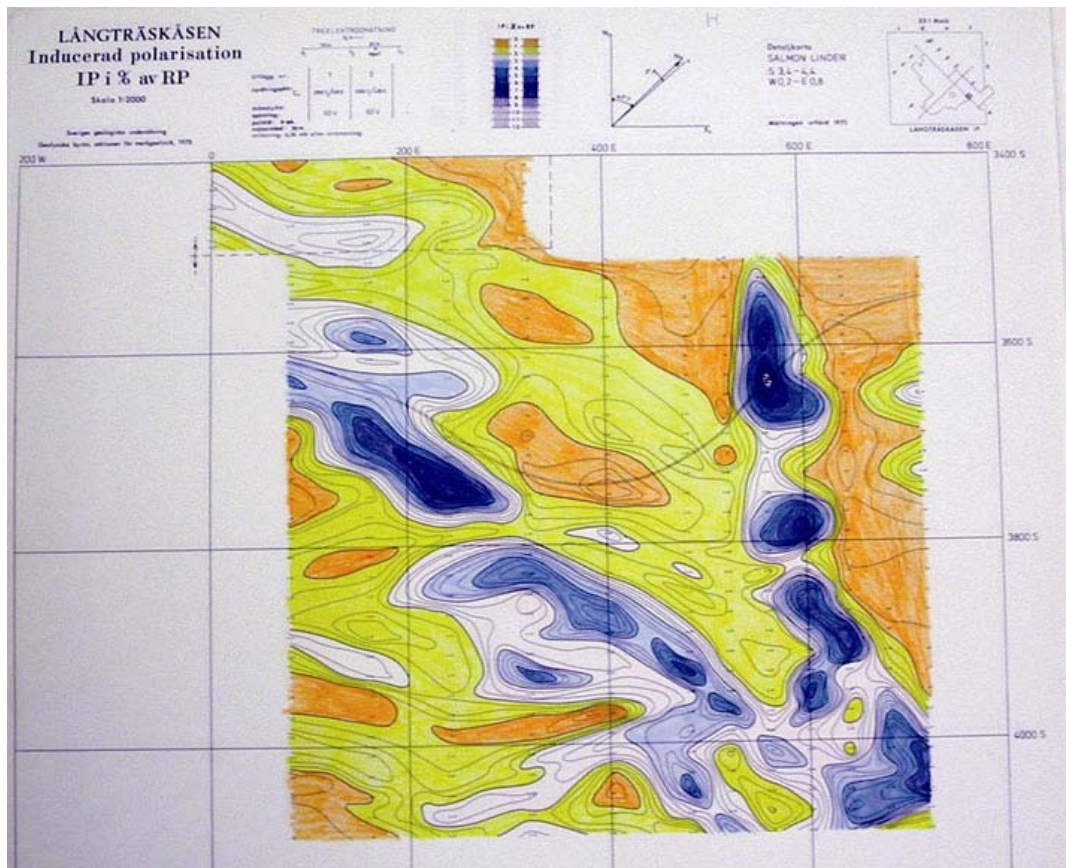
Denna uppskattning indikerar att det finns ca 19 000 kartor (huvudsakligen i A1 format) med analog geofysisk information. Utav dessa har ca 65 % mätvärden. Dessutom finns 61 stycken A1-pärmar med i genomsnitt 95 kartor (baserat på genomräkning av 23 av pärmarna med geofysikkartor med mätningar). Det är ca 5800 kartor med mätvärden i pärmarna. Även detta material är unikt.

Det betyder att det finns totalt ca 18 000 kartor med mätvärden. Efter en undersökning av redan digitalt geofysikmaterial (ca 4 400 kartor) kan vi estimerar att ca 75 % av dessa också finns som analoga kartor i kartlådorna. Det betyder att ca 3 300 kartor kan plockas ut ur det analoga materialet och det återstår då ca **15 000 kartor med mätvärden** som skall digitaliseras.

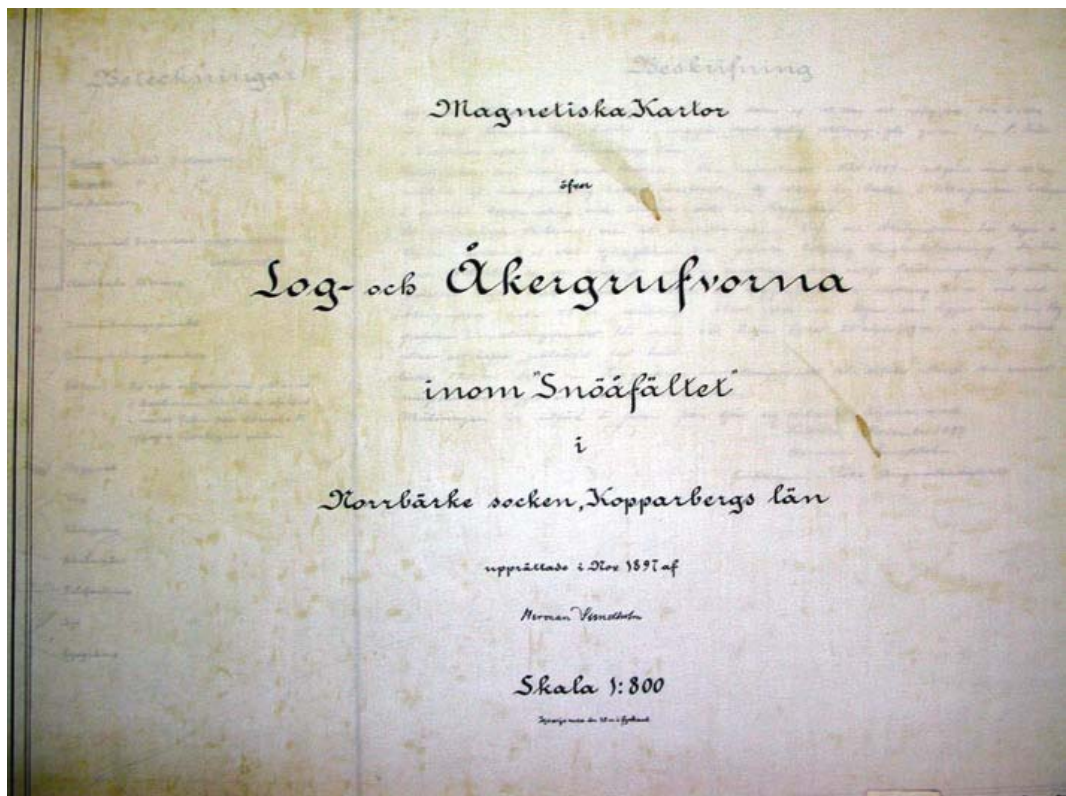
Det är önskvärd, bl.a. ur säkerhetssynpunkt att skanna samtliga kartor till digitala bildfiler. Det totala antal kartor (efter utsortering av dubletter och dyl.) för skanning uppgår till ca 18 000 st.

Geofysikmaterialet är heterogent och består fysiskt av en rad olika tryckkvaliteter och format. Det finns bland annat handritat material som är mer än 100 år gammalt. Följande format finns representerat:

1. Papperskartor (A3-A1 format).
Dessa kartor finns i en rad olika kvaliteter; bl.a. handritade, tryckta, färgade, svartvita, konturerade med mätvärden, konturerade utan mätvärden (fig. 12). Det är stor variation med hänsyn till kvalitet och läsbarhet. Många av dessa kartor är emellertid kopior, och original kan finnas som translarer.

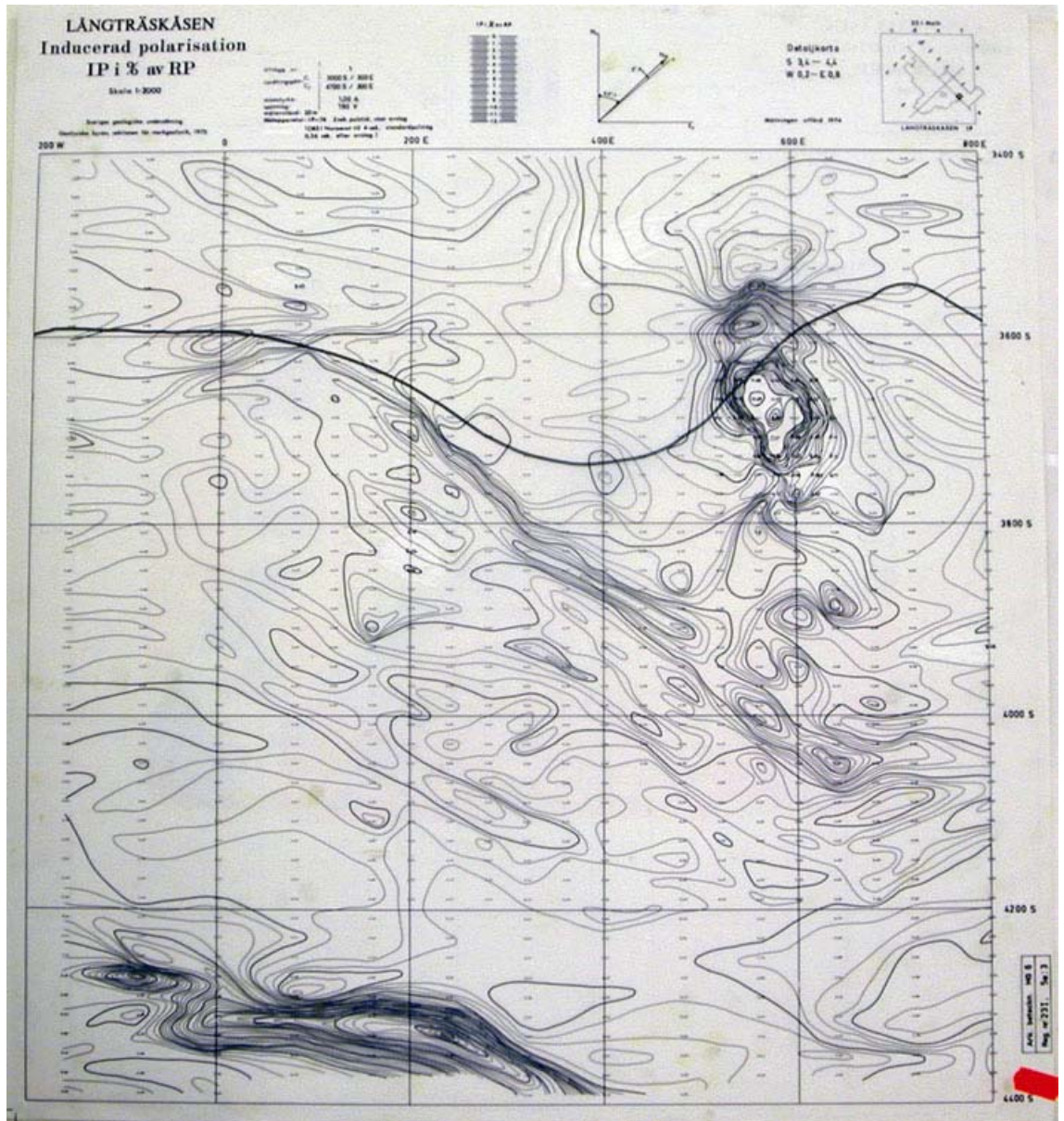


Figur 12. IP mätning tryckt på papper.



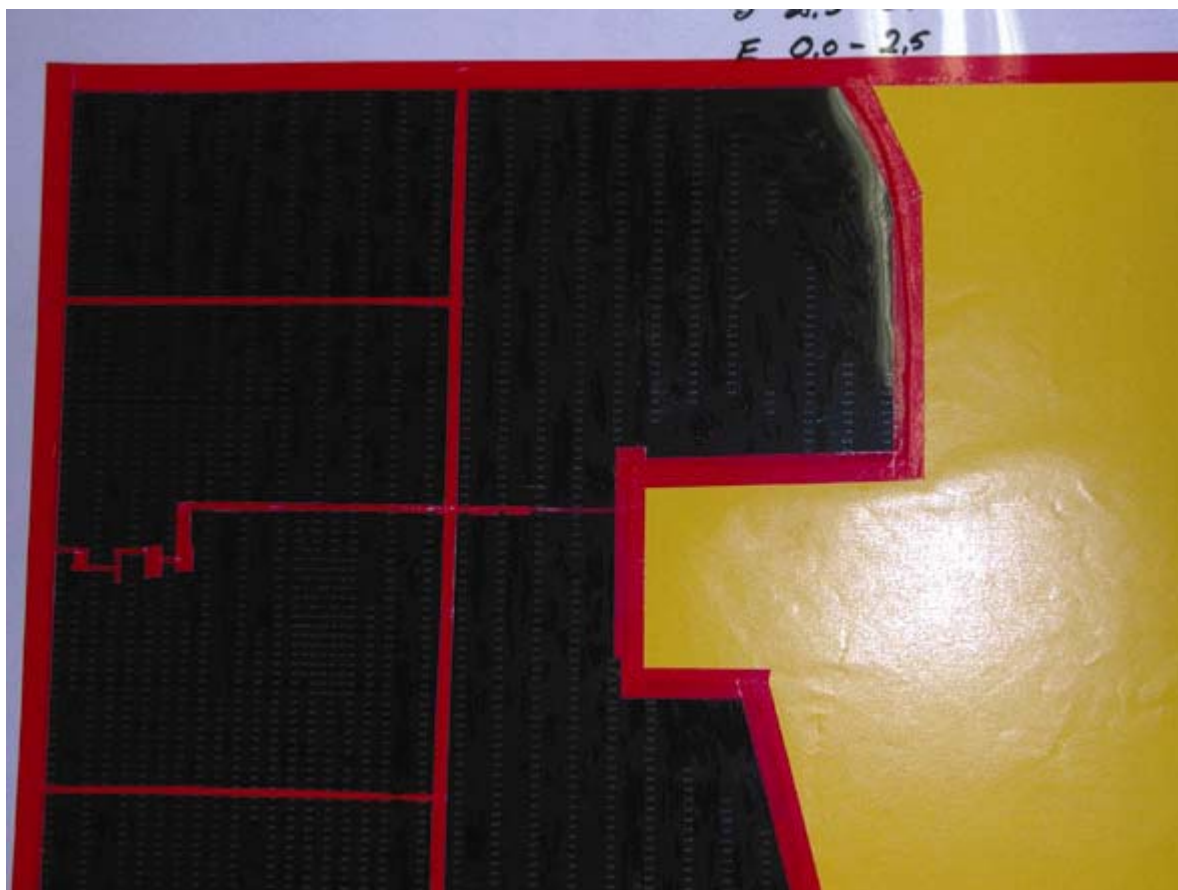
Figur 13. Magnetisk mätning från 1897.

2. Transparenta folier (translarer i A1 format) med svart tryck (fig. 14). Dessa är i huvudsak av mycket god kvalitet. Det finns kartor både med och utan konturlinjer.



Figur 14. Translar i storlek A1 med konturerad geofysikmätning (IP).

3. Repronegativ (varierande format). Dessa har svart bakgrund och genomskinliga siffror (Fig. 15). På repronegativ finns endast siffror och inte konturlinjer. Kvaliteten är god.



Figur 15. Repronegativ.

För att testa olika metoder för att skanna geofysikdata så blev det i ett inledande skede av projektet valt ut ett begränsat material. En beskrivning av testerna finns i bilaga 1.

Testresultat

Skanningen blev utförd i flera pass där olika upplösning, gråskala och svart-vitt blev utvärderat. Det blev på ett tidigt stadium klart att kartor som innehåller konturerade mätvärden (vilket är en stor del av materialet) skapar problem för OCR programvaran. Talvärden blir maskerade av konturlinjerna, något som medför att de inte är läsbara på ett meningsfullt sätt. Detta kräver en hög grad av manuell korrigerings – något som är tidskrävande. Kartor med endast mätvärden och rutnät är mindre problematiska, men under OCR-konverteringen kan det uppstå problem när mätvärden och rutnät överlappar. Detta är fallet för de flesta av de undersökta kartorna.

Även om denna test är baserad på endast ett fåtal skanningar, så är slutsatsen att det finns följande realistiska alternativ:

- 1) Manuell inmatning av data.
- 2) Semiautomatiserad inmatning av data.
- 3) Skanning av specialist (extern konsultfirma).

Utvärdering av alternativ

- *Manuell inmatning av data.*
Detta alternativ innebär att det från varje karta manuellt skall avläsas och inskrivas mätvärden och koordinatinformation. Detta kan vara bästa sättet på de kartor som har få mätvärden, men mycket tidskrävande när det är många.
- *Semiautomatiserad inmatning av data.*
Detta innebär att man OCR-skannar till ett koordinatsystem (t.ex. ett Excel ark) och korrigerar manuellt. Detta kan vara ett realistiskt alternativ när det inte finns konturer eller annat som maskerar mätvärden. Ett problem är dock att många kartor innehåller linjer och konturer.
- *Skanning av specialist (extern konsultfirma).*
En prisförfrågan har ägt rum där vi inhämtat information från tre externa skanningsspecialister/konsulter angående digitalisering av markgeofysik. De tillfrågade firmorna är:

PICAB	- http://www.picab.se
ASTACUS AB	- http://www.astacus.se
PROCAPTURA AS	- http://www.procaptura.com

Prisförfrågan berörde skanning och digitalisering av:

- 10 000 kartor i storlek A1, gråskala
- 4000 kartor i storlek A0, gråskala
- 4000 kartor i storlek A0, färg

PICAB är en firma i Jokkmokk som bland annat är specialiserad på digitalisering av elnät. Firman kan emellertid utföra det mesta inom digitalisering. En representant från PICAB besökte Mineralinformationskontoret i Malå 2005-11-02. Möjliga digitaliseringsalternativ blev diskuterade och PICAB fick med sig testmaterial från arkivet som användes som underlag för prisinformation och leveranstid.

En offert sändes oss 2005-11-16 och en tidsstudie för ev. inmatning av markgeofysiska värden några dagar senare. Det framkommer att manuell inläsning är det enda alternativet och vid en sammanhängande totalentreprenad på detta jobb kan de lämna ett pris på 275 SEK/tim. Kostnaden för manuell inläsning av uppskattningsvis 5 miljoner värden skulle kosta ca 1,87 miljoner kr. Skanning av ovan nämnda kartor skulle totalt kosta 526 000 kr.

Astacus AB är en firma i Vikingstad utanför Linköping, som är specialist på digitalisering. Firman använder en kombination av skanning i Sverige och manuell digitalisering i "högteknologiska låglöneländer" bl.a. i Indien. 2005-10-17 blev det skickat en förfrågan till Astacus AB med begäran om

prisuppgifter etc. för digitalisering av geofysikmaterialet. Testmaterial skickades 2005-11-22 och några veckor senare erhöll vi Astacus prisförslag. Kostnaden för manuell inläsning uppskattas hos Astacus till 1,34 miljoner kr. Skanning av ovan nämnda kartor skulle totalt kosta 306 000 kr.

Procaptura AS är en norsk firma som tidigare hette Kongsberg Scanners. SGU anlidade tidigare denna firma till digitalisering av berggrundskartor och har goda erfarenheter av detta samarbete. 2005-11-23 skickades en förfrågan till Procaptura varefter testkartor skickades. Det visade sig att de haft tidigare erfarenheter av bl.a. hydrografiska kartor av olika slag och utvecklat en programvara för att kunna få fram önskad information på ett lätt och 100 % säkert sätt. Efter utvärdering av de tillsända kartorna kunde de ge oss en prisbild. I nuläget skulle det kosta 10 000 kr per karta. Det betyder att det skulle kosta 150 miljoner kronor att digitalisera 15 000 kartor. Skanning av kartor ingår i priset. Priset skulle kunna gå ner till 2000-5000 kr per karta om vi först sätter upp ett pilotprojekt till en kostnad av ca 1 miljon kr.

Utlokalisering av digitaliseringsjobb betyder att originalmaterial från arkivet i Malå måste transporteras från Malå till produktionsorten. Detta medför en risk för att materialet kan skadas eller tappas bort. En lösning på detta problem kan vara att SGU själva organiserar transport t.ex. med tjänstebil från Malå till produktionsstället.

Förslag till metadatastruktur

Till varje skannad geofysikkarta bör följa lämpliga metadata och vi anser det möjligt att använda samma struktur som för geologikartorna, (se sid. 36), men med följande tillägg: lok_NW, lok_SE samt NW och SE.

För de digitaliserade kartorna, (där varje mätvärde på kartan har en koordinat i det lokala staxsystemet) kan befintliga databastabellstrukturer (olika för olika metoder) användas, se bilaga 2.

Markgeofysiska data i pärmar

Utöver de kartor med markgeofysiska data som här utretts finns även ett antal pärmar med olika typer av markdata. Många av dessa värden kan återfinnas på de markgeofysiska kartorna men i en del fall är värdena i pärmarna de enda som finns av de mätningar som är gjorda. Det finns 150 st. ringpärmar, (mestadels i A4-format) med i genomsnitt 122 sid. i varje. Markdata i dagboksformat finns 1210 st. av med ca 143 sid. i varje. Dessa är skrivna med blyerts och bör skannas i gråskala.

Datalagringsbehov

Varje geofysikkarta antas ta ca 100 MB i TIFF-format. Vid skanning av 18 000 kartor blir det totala lagringsbehovet ca 1,8 TB.

Rekommendation

Geofysikmaterialets heterogena och svåra karaktär gör att det sannolikt vill vara mest effektivt att överlåta digitaliseringsjobbet av detta material till en extern digitaliseringsfirma.

En vanlig skanning av kartorna till digitala bildfiler bör göras av säkerhetsskäl och för att underlätta sökning och distribution av data. Detta *kan* göras av SGU, men om materialet skickas till konsult för digitalisering av mätvärden så kan det vara kostnadseffektivt att låta konsulten göra detta samtidigt.

Metodik vid eventuellt utskick till extern digitaliseringsfirma:

- I den omfattning det är möjligt skall det användas kopior och inte originalmaterial.
- Det krävs en noggrann genomgång av alla 320 lådor innan materialet skickas iväg. Detta för att sortera bort eventuella dubletter och för att dokumentera innehållet i detalj.
- Alla kartor numreras fortlöpande med ett ID-nr. innan de skickas iväg.
- Det görs en Excel-fil med metadata som använder ovannämnda ID-nr som unik identifikation.

Metadata är typspecifika; dvs. att Mag, Slingram, IP och VLF har olika metadata.

Borrhålsprotokoll och kemiska analyser av borrhärnor

Material: A4 pärmar med loggar, mätprotokoll, analyser m.m. från borrhål och borrhärnor som finns i SGUs borrhärnearkiv.

Antal pärmar: 700 st.

Antal sidor: ca 300.000-350.000.

Målsättning: Att digitalisera all borrhålsinformation. En målsättning är att den kemiska informationen som beskriver borrhärnorna skall extraheras och omformas till ett format som möjliggör inmatning i sökbar databas.

Beskrivning och skanning

I dagsläget finns ca 25 000 borrhärnor i SGUs borrhärnearkiv. I SGUs pappersarkiv i Malå finns loggar, analyser och mätprotokoll vidrörande dessa borrhärnor arkiverade (fig. 16). Det finns dessutom redan en digital databas med koordinaterna till borrhålen.



Figur 16. Borrhåls pärmar. Totalt 700 stycken med 350-500 sidor vardera.

En fördel med dessa borrhåls pärmar är att alla dokument finns i ringpärmar, vilket underlättar hantering och ADF-skanning.

Ett stort problem med materialet är att pappersmaterialet är mycket heterogent. I en och samma pärm kan det finnas borrhålsprotokoll, kemiska analyser, loggar, kartor etc. Dessa kan dessutom vara av varierande dokumentstorlek. Stora dokumentformat kan vara hopvikta och ligga omlott emellan vanliga A4 sidor. Kvaliteten på kopiorna varierar från mycket bra till nästan oläsbara. Det kan vara stora kontrastskillnader från sida till sida, något som gör det svårt att hitta optimala skannerinställningar. Innan digitalisering måste varje pärm därför ses igenom. Eventuella ”svåra” dokument kan då plockas bort för individuell digitalisering. Detta kommer sannolikt att ta mer tid än själva skanningen.

Det finns även 13 separata pärmar plus en lös bunt om ca 25 cm med kemiska analyser på både borrhärlor, block och håll (blandat). Borrhärloranalyser måste här separeras ut och jämföras med de eventuella kemiska analyser som finns i borrhärlorpärmarna (en del av analysmaterialet finns redan i borrhärlorpärmarna i form av kopior). Det är i nuläget okänt hur många analyser detta omfattar, men detta kommer sannolikt att bli en tidskrävande uppgift.

Det är sammanlagt 21 borrhärlorpärmarna som skannats inom projektet.

Resultat av OCR-konvertering

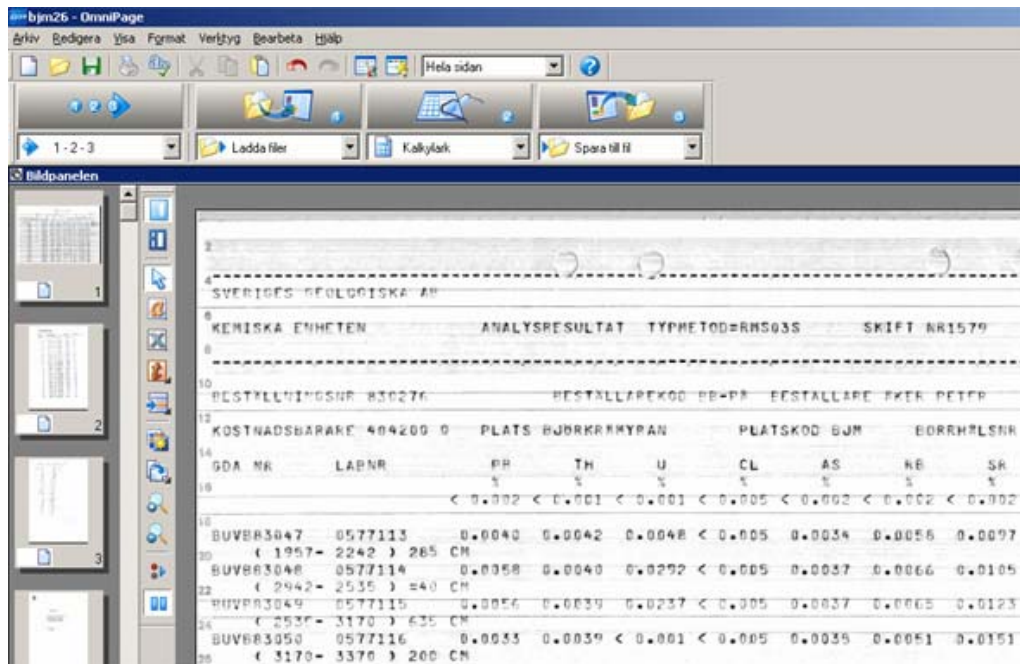
En prioriterad uppgift är att få alla geokemiska analyser inmatade i en sökbar databas. Detta ställer krav på OCR programvaran och källmaterialet. Som påpekat ovan så består många pärmar av kopior med mycket variabel kvalitet.

Konvertering av borrhärlorpärm i Adobe Acrobat 7.0 visar på stora problem med att konvertera textmaterialet, de beskrivande delarna av loggarna. Texten som konverterats i testen är av medelgod kvalitet, i jämförelse med många andra pärmar, men ändå blir det många fel i den OCR-lästa texten. Vissa sidor som har ramar kring sig känner programvaran av endast som bilder.

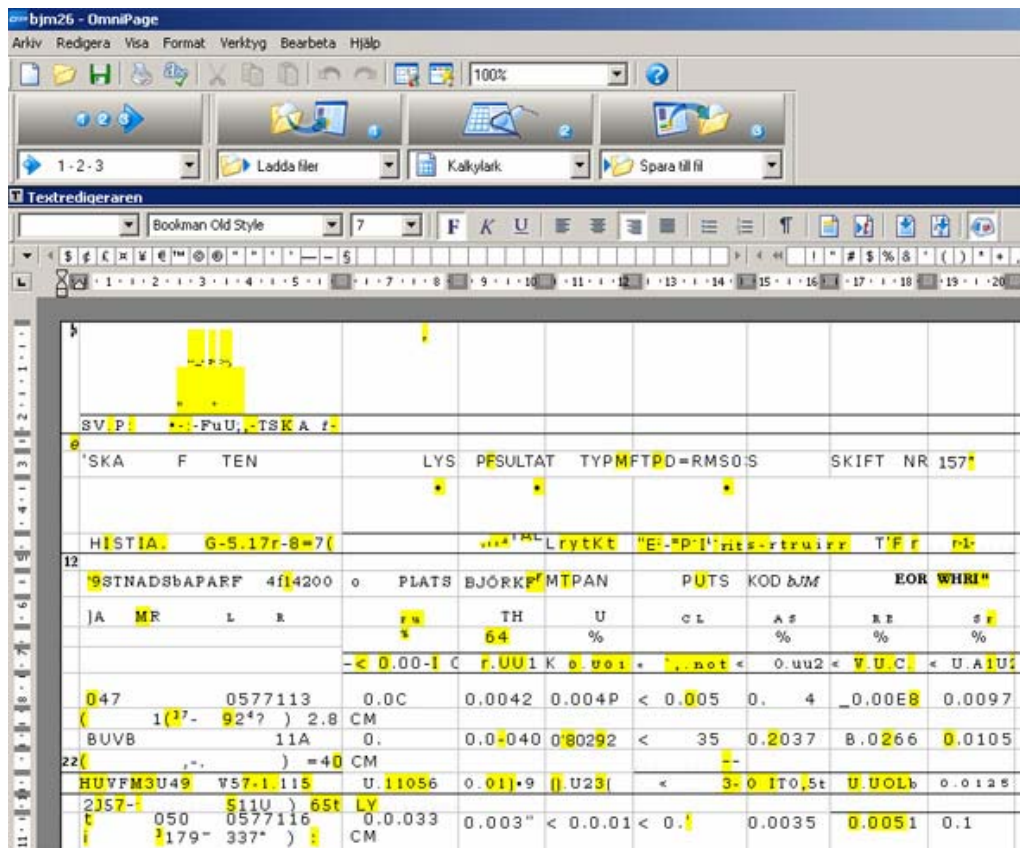
Konvertering av borrhärlorpärm i Omnipage Professional uppvisar flera slags problem. Dels går oftast kolumntext ut i delen där analysvärdena står (fig. 17) vilket skapar problem för programmet när värdena ska sparas som Excel. Det krävs då mycket manuellt arbete för att få värdena på rätt plats i tabellen.

Ett annat problem är att en del av borrhärloranalystabellerna i sig är sneda kopior eller kan vara snett inskannade vilket gör att Omnipage inte kan konvertera dessa rakt och därmed inte lyckas spara i en läsbar Excel-fil. Förövrigt är det också blandad tryckkvalitet på analyserna som också skapar problem vid konvertering.

Konklusionen är att borrhärlor materialet utgör ett mycket svårt material för OCR (fig. 18). I nuläget så bedöms det som ett högriskprojekt. I många fall är det enda realistiska alternativet manuell inmatning av text och analysvärden. Dock uppnås relativt bra resultat när tryckkvaliteten på kopiorna är god och tabellerna är översiktliga och raka (fig. 19, 20).



Figur 17. Skannad kemisk analys i Omnipage innan OCR konvertering. Kolumntexten går ut i delen där analysvärdena står och skapar problem för OCR tolkningen. Kopians tryckkvalitet är dessutom något svag.



Figur 18. Analysen i Figur 17 efter OCR konvertering i Omnipage. Oläsbart resultat.

294 - OmniPage

Arkiv Redigera Visa Format Verktyg Bearbeta Hjälp

Hela sidan

1-2-3 Ladda filer Kalkylark Spara till fil

Bildpanelen

105

106

107

SCU DATASEKTICN
PRCGRM DJULIS6
SKIFT 3408 TYP 34

DATUM: 1977- 2-22

IDNR	PROVET MÄRKT	CU PPM	PB PPM	ZN PPM	MO PPM	NI PPM	CR PPM
0530065	PR1- BH59207 STEKENJCKK	1250	1057>	4859	55	12	28
0530086	PR2 BH59207 STEKENJCKK	300	136	2020	14	67	247
0530087	C393 BH59209 STEKENJCKK	>13000>	1450>	5000	70	18	54
0530088	C394 BH59209 STEKENJCKK	880	148	919	37	< 10	6
0530089	C397 BH59209 STEKENJCKK	620	62	898	25	< 10	12
0530090	C398 BH59209 STEKENJCKK	>13000>	1638>	5000	43	11	48
0530091	C493 BH62201 STEKENJCKK	2000	218	2030	34	< 10	22
0530092	C494 BH62201 STEKENJCKK	4500>	1391>	5000	48	23	40
0530093	C311 BH62001 ANKARVATTNET	8580	921>	3706	< 10	13	34
0530094	C312 BH62001 ANKARVATTNET	7660	1056>	4038	< 10	14	46
0530095	C313 BH62001 ANKARVATTNET	1620>	1074>	3892	< 10	15	51
0530096	C314 BH62001 ANKARVATTNET	5770	> 783>	2922	< 10	< 10	12
0530097	C315 BH62002 ANKARVATTNET	>13000	300>	5000	< 10	21	79
0530098	C316 BH62002 ANKARVATTNET	220	65	1300	10	44	187
0530099	C317 BH62002 ANKARVATTNET	980>	1056>	3737	< 10	18	86
0530100	C318 BH62002 ANKARVATTNET	3190	789>	3608	< 10	11	39
0530101	C319 BH62002 ANKARVATTNET	210	374	1035	16	45	181
0530103	C321 BH62002 ANKARVATTNET	4900	> 763>	2639	14	< 10	14
0530104	C337 BH62006 ANKARVATTNET	9310	708>	4098	13	22	34
0530105	C338 BH62006 ANKARVATTNET	7860	> 995>	3849	11	17	26

Figur 19. Skannad kemisk analys i OmniPage innan OCR konvertering.

294 - OmniPage

Arkiv Redigera Visa Format Verktyg Bearbeta Hjälp

100%

1-2-3 Ladda filer Kalkylark Spara till fil

Textredigeraren

Style 1 Lucida Console 8

SCU 04TASEKTICN
PR2CGB1AM 0JULIS6
SKIFT 3408 TYP 34

DATUM: 1977- 2-22

IDNR	PROVET MÄRKT	CU PPM	Pt1 PPM	ZN PPM	MO PPM	NI PPM	CR PPM
0533385	PR1- 0m59207 ST2KENJCKK	1250	1057>	4359	55	12	28
0530086	PR2 8859207 STEKENJCKK	300	136	2020	14	67	247
0530087	0393 dH59209 STEKENJCKK	>13900>	1450>	5000	70	18	54
0530088	C3/4 8859209 STEKENJCKK	880	143	919	37	< 10	6
C53...C09	C397 BH59209 STEKENJCKK	620	62	898	25	< 10	12
0530300	0393 8854209 STEKENJCKK	>13003>	1638>	5000	43	11	48
C53...091	0453 8862201 STEKENJCKK	2000	218	2030	34	< 10	22
0530092	0494 8862201 STEKENJCKK	4530>	1391>	5300	48	23	43
3530093	0311 8862001 ANKARVATTNET	8580	921>	3706	< 10	13	34
0530094	C312 8862001 ANKARVATTNET	7660	1056>	4038	< 13	14	46
0530395	C313 8862001 ANKARVATTNET	1620>	1074>	3892	< 10	15	51
0530096	C314 8862001 ANKARVATTNET	5770	> 783>	2922	< 10	< 10	12
C:33097	C315 3662002 ANKARVATTNET	>13000	300>	5003	< 13	21	79
C530308	0316 8862002 ANKARVATTNET	220	65	1300	10	44	187
0533299	C317 8852002 ANKARVATTNET	980>	1056>	3737	< 10	18	86
0530100	0318 8862002 ANKARVATTNET	3190	789>	3608	< 10	11	39
0530101	C319 9862002 ANKARVATTNET	210	374	1035	16	45	181
C530103	0321 8862002 ANKARVATTNET	4900	> 763>	2639	14	< 10	14
0530104	0337 81-162006 ANKARVATTNET	9310	708>	4098	13	22	34
0530135	0338 8862006 ANKARVATTNET	7860	> 995>	3849	11	17	26

Figur 20. Samma analys som i Figur 19 efter OCR konvertering i OmniPage. Resultatet av den här konverteringen är acceptabel, men kräver ändå en viss grad av manuell korrigerig.

Rekommendation

Det måste tagas bort dubletter i pärmarna och ersätta kopior med original av analyser, före skanning. Skanning bör företas i svart-vitt (bara gråskala där det är nödvändigt) i största möjliga mån för att minimera storleken på PDF-filerna som annars lätt kan bli ohanterliga. Detta gäller speciellt vid bearbetning, t.ex. om kontrast, ljusstyrka etc. skall korrigeras i OCR-programmet.

Geologiska kartor

Material: Kartor i olika format som är tryckta/ritade på papper och på transparenta plastfolier. Materialet inkluderar geologiska detaljkartor, observationskartor, blockkartor, borrhålsloggar mm.

Häll och geologikartor:	ca 12 000 st.
Blockkartor:	ca 4000 st.
Div. kartor	ca 3000 st.
Regionalkartor	ca 1000 st.
Äldre inbundna gruvkartor	ca 3000 st.
Övriga gruvkartor	ca 1300 st.
Generalstabskartor	ca 735 st.
Totalt:	ca 25 000 st.

Målsättning: Att skanna alla kartor till digitala bilder. Dessa kan därefter kopplas till metadata för att underlätta arkivering och sökning. Detta ökar också tillgängligheten av materialet.

De geologiska kartorna omfattar ett antal olika slags geologiska kartor. Det finns bl.a. geologiska översiktskartor, detaljkartor, material associerat med borrhningar, häll- och blockkartor mm.

Huvudsakligen är de geologiska kartorna sorterade efter kartblad i 340 st. kartlådor (fig. 21) i ca 1200 mappar. Utöver dessa finns 735 st. kartor sorterade enligt det gamla generalstabskartsystemet i en separat avdelning. Materialet är mycket heterogent, arbetsmaterial och tryckta kartor ligger blandat. Storlek och kvalitet på kartorna är variabel. De största kartorna är i A0 format men merparten av kartorna är i A1 format.



Figur 21. Lådor med geologikartor i pappersarkivet.

Resultat av skanning

Skanning av A1-och A0-kartor företogs vid huvudkontoret i Uppsala vid skannern Colortrac Smart LF 4080. Under sessionen skannades 170 kartor, varav 48 st. geologikartor (övriga 117 st. kartor var A1-och A0-bilagor till prospekteringsrapporter och borrhålskartor). Fig. 22 och fig. 23 visar exempel på en skannad geologisk karta. Kartorna blev skannade till TIFF format i färg och i gråskala. Det tog ca tre gånger så lång tid att skanna i färg som i gråskala; A1 gråskala 300 dpi ca 0,5 min. A1 färg 300 dpi ca 1,5 min. En del av kartmaterialet är så fragilt att det först måste fästas på annat underlagspapper innan det skannas vid A0-skanner. Andelen fragila kartor uppskattas dock till < 5%.

Kvaliteten på de skannade kartorna är variabel beroende på att ljusstyrkan på den skanner som användes inte är homogen över hela bilden.

Förslag till metadatastruktur

För att kunna söka efter kartor bör varje karta vara beskriven. De äldre generalstabskartorna (735 st.) som förvaras i en separat del av arkivet, är de enda som i nuläget är sökbara. Projektet har utarbetat ett förslag till metadata utifrån dessa befintliga metadata som finns som register. Kartorna får ett löpnummer vid skanning som är detsamma som id-numret för kartan. De geologikartor som skannats i projektet är också metadata inlagda för.

Följande metadatafält föreslås:

ID_NO
TITEL
AMNE
PROJEKT
FORF
AR
KARTBLAD
LAN
KOMMUN
LOKAL_OBJ
NYCKELORD
KARTTYP
SKALA
KOORDINATSYSTEM

Datalagringsbehov

De 48 geologikartor som är testskannade tar 6,9 GB diskutrymme. I genomsnitt 143 MB per karta (TIFF format) ger för det totala antalet geologikartor (ca 25 000 st.) ett datalagringsbehov på 3,57 TB.

Utvärdering av alternativ

Det finns två alternativ för digitalisering av de geologiska kartorna.

- Digitalisering av extern digitaliseringsbyrå.
- Skanning på plats i Malå.

Rekommendation

Det rekommenderas att skanningen utförs på plats i Malå. Om skanningen görs på plats så kan arkivet vara operativt och kunna yta service till kunder också under skanningsperioden. Dessutom så undviker man att skicka i väg materialet, något som medför risk för att materialet kan skadas eller tappas bort.

Ett viktigt moment är dessutom att materialet måste gås igenom av en geolog innan skanning. Vissa av kartorna kan innehålla mycket lite information, t.ex. endast en punkt med ett nummer. Denna punkt bör då överföras till annan karta eller dubblett över samma område som innehåller mer information.

Metadata skall skrivas till kartorna i samband med skanning. Metadataregistrering kan vara en tidskrävande uppgift när kartuppgifter i många fall saknas. Uppskattningar indikerar att sortering och metadata-sättning kommer att kräva flera arbetstimmar än själva skanningen (se sid. 48).

Borrhålskartor

Material: Kartor med borrhålslägen samt profiler, huvudsakligen i A1/A0 format tryckta på papper och transparenta plastfolier.

Antal: ca 4300 st.

Målsättning: Att skanna alla kartor till digitala bilder. Dessa kan därefter kopplas till metadata för att underlätta arkivering och sökning. Detta ökar tillgängligheten av materialet.

Borrhålskartorna visar var borrhål är placerade och är ett viktigt komplement till den information som finns i borrhålsparmarna. Kartorna är i dag sorterade kartbladsvis.

Resultat av skanning

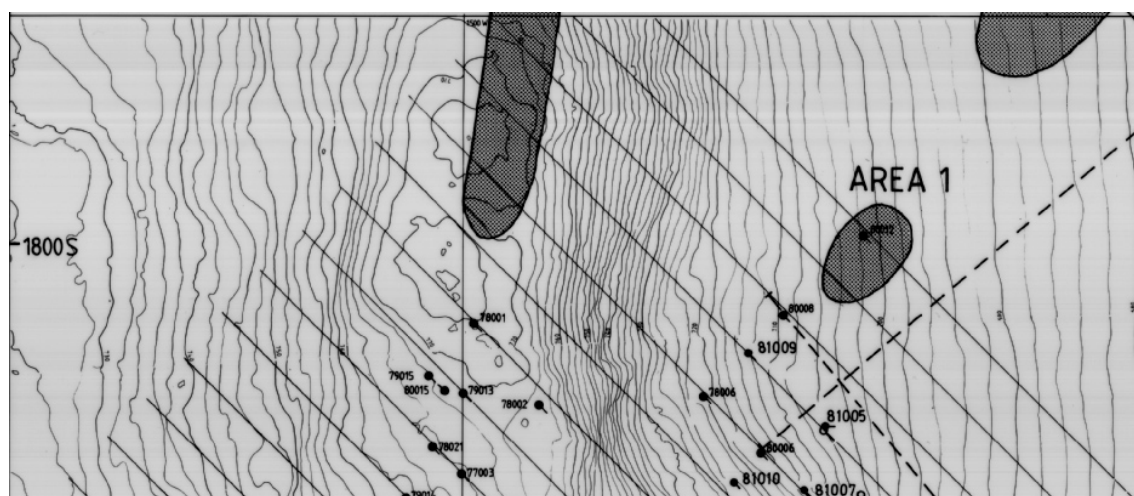
Skanning av åtta stycken borrhålskartor ägde rum vid A0-skannern i Uppsala. Det tar ca 0,5 – 1 minut att skanna en karta i gråskala, 300 dpi, till TIFF, vilket är att rekommendera. Resultatet av kvaliteten på de skannade kartorna är variabel beroende på att ljusstyrkan på den skanner som användes inte är homogen över hela bilden, något som drabbar framförallt kartor tryckta på transparenta plastfolier, vilket många av borrhålskartorna är (se fig. 24).

Metadata

För att kunna söka efter en borrhålskarta är det viktigt att varje karta beskrivs efter bestämda kriterier. Samma typ av metadatastruktur kan användas för borrhålskartorna som för geologikartor, (se sid. 36).

Datalagringsbehov

Borrhålskartor saknar oftast färg, därför kräver de mindre diskutrymme än färgkartor. Genomsnittstorleken på en skannad borrhålskarta som TIFF-fil är ca 84 MB, vilket gör det totala lagringsbehovet till 360 GB.



Figur 24. Utsnitt av skannad borrhålskarta. Originalen är en plastfolie (translar).

Geokemiska data av olika typer samt kartor

Material: 43 stycken A4 ringpärmar med analyslistor, ca 10 000 A4-sidor sammanlagt.
Kartor mestadels i A1 format: ca 25 000 st.

Målsättning: Att skanna in de analyser som ej finns digitalt samt OCR konvertera dessa så att analysvärden kan importeras in i befintliga databaser vid SGU.
Att skanna kartorna samt föra in metadata för varje karta för att öka tillgängligheten och därmed användningen av materialet.

De geokemiska analyserna omfattar prov av jord, myrkant, bäcktorv, bäcksediment och tungmineral tagna under perioden 1964 – 1983. En del av detta material finns redan idag digitalt. Tid har använts till inventering av pärmar för att se vilket analogt material som behöver skannas. Inventeringen indikerar att 43 pärmar återstår. Av dessa är nu 13 stycken skannade inom projektet.

De geokemiska kartorna visar bl.a. var proverna är tagna. Det är inte alltid ovan nämnda analyser har någon koordinat varför kartmaterialet är värdefullt. Det totala antalet geokemiska kartor i arkivet är uppskattade till ca 25 000 stycken baserat på räknande av 12 slumpvis utvalda lådor av sammanlagt 240 st. lådor som är indelade kartbladsvís. Regional geokemi tillkommer där 32 st. lådor innehåller tillsammans ca 1000 kartor.

Beskrivning och resultat av skanning

Analyslistorna är av mycket skiftande kvalitet. Trots att de sitter i ringpärmar krävs en del förarbete.

Många analyser finns på blad med häftklamrar. Dessa häftklamrar måste tas bort innan skanning. Många analysblad sitter dessutom ihop (listpapper) och dessa måste först separeras. Denna procedur tar ca 20 minuter per pärm. Några blad har hålen på olika platser så man måste göra nya hål efter skanningen vilket tar ca 5 min.

Analyslistorna har testats både för skanning i gråskala och svart-vitt. Det visar sig att det bästa är svart-vit skanning när kopiorna i pärmarna är av tillräckligt bra tryckkvalitet. Oftast krävs dock att gråskale-skanning företas. Av de 13 pärmar om ca 250 sidor vardera som skannats ser man att värdering måste göras från fall till fall och kontrast och ljusstyrka kan förbättras via inställningar på skannern.

Det tar sammanlagt ca tre timmar att skanna en pärm till PDF-fil, vilket förstås varierar beroende på tryckkvalitet.

Resultat av OCR-konvertering

Vid OCR-konvertering sker import av PDF-filer direkt in i Omnipage Professional. Det tar ca 7 minuter att ladda 100 sidor, (i gråskala 300 dpi). Själva konverteringen tar sedan ca 9 minuter för 100 sidor. Efter konvertering kan man spara resultatet som en Excel-fil.

Resultatet är mycket varierande beroende på ursprungsmaterialets uppställning av kolumner och givetvis tryckkvalitet, (jfr. borrhålsanalyser). Kontrast och ljusstyrka går dock att elaborera med i Omnipage Professional för att nå bättre resultat. Det går också att välja var kolumnavskiljare skall placeras före konvertering. Resultaten av testkonverteringen visar att delar av geokemianalyserna (kanske upp till ca 50 %) går att OCR-konvertera med mindre handpåläggning.

Excellfiler med analyser visar på 11-20 % fel för analystabeller av bättre kvalitet och kolumnuppställning. Manuell korrigering av excellfilerna är tidskrävande men uppskattas ändå till att ta kortare tid än manuell inmatning av alla värden från grunden. Dock kan manuell inmatning vara nödvändig för en del av materialet. Total tidsåtgång uppgår till 1250 timmar. Manuell inmatning och korrigering är då inkluderad.

Problemen med att OCR-läsa geokemianalyserna gör att detta moment i nuläget måste betraktas som ett högriskprojekt. Det kan påpekas att det är mycket viktigt att analysvärdena som läggs in i databasen är korrekta och den grad av noggrannhet som vi kräver kan inte uppnås med de OCR-program som vi har testat. Om man väljer att OCR-läsa först och därefter manuellt korrigera så är detta en tidskrävande uppgift.

Datalagringsbehov

De skannade analyserna i PDF beräknas ta 7,7 GB i utrymme. Storleken på Excel-filerna blir försumbara. Lagringsbehovet för geokemikartorna, som skannas till TIFF-format, kan beräknas ur storleken på de skannade geologikartorna som är av liknande slag och beräknas då ta ca 3,5 TB.

Metadata

För att kunna söka bland geokemikartor bör metadata skrivas till varje karta. Metadatastrukturen för geologikartorna kan även användas för geokemikartorna. Metadatastrukturer är en tidskrävande uppgift och det beräknas ta 3125 timmar baserat på att man metadatasätter 8 kartor i timmen.

Rekommendation

Det rekommenderas att alla analyspärmarna skannas till PDF-filer. Inmatning av data i databaser kan göras med en kombination av OCR och manuell korrigering eller inmatning. Kartorna skannas till bildfiler (TIFF format) på samma sätt som geologikartorna.

Block- och hälldata samt fältdagböcker

Material: Block och hällbeskrivningar samt analyser i A4-format huvudsakligen, totalt ca 124 000 sid. Block- och hällbeskrivningar i dagboksformat (A6), totalt ca 14 500 sid.

Målsättning: Att digitalisera materialet för att öka tillgängligheten och användningen av det.

Beskrivning och skanning

Block- och hällbeskrivningar är en mycket inhomogen samling data som insamlats under perioden 1940 – 1991. Inventering av befintligt material visar att det finns:

- 90 ringpärmarna med huvudsakligen blockbeskrivningar samt blockanalyser
- 36 ringpärmarna blandat häll- och blockobservationer samt analyser
- 25 A5-pärmarna block- och hällobservationer
- 395 ringpärmarna hällobservationer och kopierade dagböcker, 15 % A5, 85 % A4.
- 169 inbundna dagböcker (A6 format), register finns
- 109 inbundna dagböcker (A6 format), register finns ej
- 13 ringpärmarna med analysintyg (+ 25 cm tjock bunt) block, häll och borrhål blandat

Ringpärmarna är huvudsakligen insorterade i kartbladsordning.

I nuläget finns ca 40 000 block och hällar inregistrerade i en databas, dock finns inte all tillhörande information till blocken och hällarna inlagda i databasen. Kemiska analyser finns för ett stort antal (kan ej uppskattas) block och hällar. En del av detta finns digitalt.

Sex stycken pärmarna är skannade i projektet. Stora variationer i tryckkontraster och att en del material är handskrivet med blyerts (fig. 25 och fig. 26) gör att man får vara noggrann med att ändra mellan gråskala och svart-vitt vid skanning.

Rekommendation

Det rekommenderas att alla block- och hällpärmarna samt dagböcker skannas till PDF-format trots att delar av informationen redan finns i databasformat idag, därför att det är en allt för tidskrävande uppgift att utröna vilka uppgifter som inte finns digitalt.

Uranarkivsmaterial

Material: Uranarkivet består av olika typer av material, bl.a. protokoll och dokumentation över radioaktiva objekt, autoradiogram, geokemiskt analysmaterial samt olika typer av kartor. Materialet är bestått huvudsakligen av A4-ark i ringpärmar samt kartor i A1 format.

Totalt antal A4 sidor som skall skannas: ca 170 000

Totalt antal A1-A0 kartor som skall skannas: ca 10 000

Totalt antal A4 sidor på mikrofilm som skall skannas: 1456

Målsättning: Att digitalisera materialet för att öka tillgängligheten och användningen av det.

Beskrivning

Uranprospektering bedrevs av SGU (Malmbyrå) mellan åren 1969-1986. Radioaktiva mineraliseringar registrerades då kontinuerligt i form av protokoll, kartverk, dokumentation och interna rapporter. Dokumentationen av radioaktiva objekt består av observations- och bearbetningsmaterial i form av dagboksblad, lägeskisser, geologiska och radiometriska kartor (i A4- eller A3- format), foton, autoradiogram, kemiska analyser, kortfattade arbetsrapporter mm.

Mellan åren 1982-92 var uranarkivet underhållet av SGAB i Luleå, därefter av MRM konsultbolag (Mark- radon- miljö) också i Luleå, fram till år 2000 då det flyttades till Mineralinformationskontoret i Malå. Materialet har här splittrats; en del, specifik för uranprospektering, har behållits intakt under samma namn medan mer generella geologiska, geofysiska och geokemiska data och kartor har insorterats efter disciplin i det övriga pappersarkivet.

Materialet består huvudsakligen av A4-ark i ringpärmar, ca 612 st. med i medeltal 279 sidor per pärm (baserat på 15 st. slumpmässigt utvalda pärmar). Det finns även ca 10000 geologiska och radiometriska kartor i A1-A0 format.

Resultat av skanning

Pärmarna kan ha ett blandat innehåll där handskriven text, tryckt text, analysprotokoll och gammalloggar ligger om vartannat. Vid skanning måste man växla mellan gråskala och svart-vitt vilket gör att det tar mellan en till tre timmar att skanna en pärm. Det är sju pärmar skannade i projektet.

Bland uranarkivsmaterialet finns unika geokemidata m.m. på microfich (1456 st. A4 sidor på 7 st. microficher). Test har utförts att försöka skanna dessa med en vanlig skanner. Detta försök misslyckades då siffrorna på microficherna är alldeles för små. SGUs bibliotek har en microficheläsare med utskriftsmöjlighet, men tyvärr är den inbyggda skrivaren ur funktion. En microfichskanner kostar ca 30 000 kr.

Vi har emellertid också undersökt möjligheten att låna utrustning, och eventuellt framtida projekt kan använda microfichskanningsutrustningen vid Forskningsarkivet vid Umeå universitetsbibliotek för en liten summa (mindre än 1000 kr).

Datalagringsbehov

En uranpärm tar i genomsnitt 296 MB i PDF och en urankarta ca 110 MB i TIFF vilket gör att det totala datalagringsbehovet för uranmaterialet är ca 1,3 TB.

Rekommendation

Det första som bör göras med uranmaterialet är att skanna microficherna, eftersom det är material som annars lätt kan förkomma.

Gruvkartor samt arkivakter med handlingar vid Bergsstaten

Material: Bergsstaten har i sina arkiv i Luleå och Falun ca 20 000 gruvkartor samt ca 170 000 sidor tillhörande arkivakter med handlingar.

Målsättning: Att digitalisera materialet för att öka tillgängligheten och användningen av det. Kartorna skall även metadatasättas för att möjliggöra sökning.

Bergsstaten är inte bara förvaltningsmyndighet för frågor enligt minerallagen. De besitter även gruvkartor från början av förra seklet och framåt. Under slutfasen av projektet har Bergsstaten visat intresse för att få dessa gruvkartor samt tillhörande arkivhandlingar digitaliserade. Gruvkartor är ofta av stort intresse för de prospektörer som vänder sig till Mineralinformationskontoret, varför det vore av stor vikt att ta med även detta material vid ett fullskaligt digitaliseringsprojekt.

Mängden kartor samt handlingar uppskattas till att vara totalt 20 000 gruvkartor (A1-A0-format) samt 170 000 sidor (A4-format) handlingar. Detta material finns vid Bergsstatens huvudkontor i Luleå och filialkontoret i Falun.

En närmare undersökning av kvaliteten på materialet har ej hunnit företas men det går att anta att dessa äldre gruvkartor är av fragilare karaktär. Därför uppskattar vi att det tar en något längre tid att skanna dessa (i jämförelse med modernare kartor) eftersom man antagligen måste fästa upp delar av kartmaterialet på underlagspapper innan skanning för att undvika att de på något sätt slits sönder.

Övrig prospekteringsintressant information

Det finns en del material som ej ännu är insorterat under ovanstående kategorier men som ändå kan vara prospekteringsintressant. Detta kan t.ex. vara kartor eller kartsnitt i A4-format, hällobeskrivningar och analyser som är gjorda inom ett visst projekt under en viss tid. Dessa är nu samlade i s.k. projektpärmar och är sorterade i kartbladsordning. Det finns ca 183 st. projektpärmar med blandad geologi. Det kan uppskattas att vara ca 180 sidor i varje pärm baserat på genomsnittsberäkningar. Dessa skannas på liknande sätt som beskrivet under kapitlet block- och hälldata.

Sammanfattning av arbetstid och kostnader

Utifrån inventeringar, beräkningar, testskanningar och test av OCR-konverteringar uppskattas digitalisering av hela arkivet ta ca 41 500 timmar, inklusive projektledartid men exklusive utbildningstid.

Tidsuppskattning av olika arbetsmoment

	antal (st/sid)	antal/timme	timmar (tot)
Prospekteringsrapporter			
skanning inkl. ur- och insättning, spara & märkn.- exkl.stora (>A3) bilagor	235 000	148	1 588
skanning stora bilagor (exkl. infogning)	28 500	21	1 357
Geofysik			
urplockning digitaliserade kartor + dubletter	27 000	33	818
skanning dagböcker med markdata	173 000	267	648
skanning ringpärmar med markdata	18 300	166	110
metadatasättning	15 000	13	1 154
Borrhålspärmar			
skanning (inkl. plocka i och ur pärm samt rengöring)	300 000	125	2 400
sortering	300 000	89	3 371
OCR-konvertering - kemiska analyser	25 000	8	3 125
Borrhålskartor			
skanning inkl. i och urplockning ur lådor	4 300	17	253
Geologikartor			
sortering, överflyttning, märkning	20 000	21	952
metadatasättning	24 300	8	3 038
skanning	20 000	21	952
skanning, äldre kartor i pärmar	4 300	17	253
Geokemianalyser			
skanning	7 000	83	84
OCR-konvertering	10 000	8	1 250
Geokemikartor			
sortering, överflyttning, märkning	25 000	21	1 190
metadatasättning	25 000	8	3 125
skanning	25 000	21	1 190
Block- och hållpärmar (ringpärmar)			
skanning inkl. ur- och insättning	23 800	83	287
Hällobservationer & dagböcker (ringpärmar)			
skanning inkl. ur- och insättning	100 000	83	1 205
Dagböcker med håll- & blockobservationer (inbundna)			
skanning	14 500	83	175

Uranarkivet				
	skanning pärmar	170 000	167	1 018
	metadatasättning kartor	10 000	8	1 250
	skanning A1/A0 kartor (inkl. märkning)	10 000	17	588
Bergstatens gruvkartor				
	skanning	20 000	15	1 333
	metadatasättning	20 000	8	2 500
Bergstatens arkivakter med handlingar, Luleå				
	skanning	16 500	125	132
Bergstatens arkivakter med handlingar, Falun				
	skanning	150 000	125	1 200
Projektpärmar				
	skanning	18 000	150	120
Antal timmar digitalisering				36 667
	varav skanningsarbete			18 265
	varav sorterings- metadatasättnings- konverteringsarbete			18 402
Projektledartid* (75%)		4 827		4 827
Totalt antal timmar (digitalisering & projektledartid)				41 494

*Projektledartid inkluderar förutom att leda personal i sitt dagliga arbete samt budget- och rapporteringsarbete, även kontroll och organisering utav producerade filer.

Kostnader

	Tid (tim)	Kostnad (kr)
Digitalisering	36 667	7 579 476
varav skanningsarbete	18 265	2 739 750
varav sort- metadatasättn- konverteringsarbete	18 402	4 839 726
Utbildning	960	198 240
skanningspersonal, 1 mån x 3		72 000
sortering-metadata-konvert-personal, 1 mån x 3		126 240
Utlokalisering av geofysikmtrl (inkl.frakt)		1 700 000
Projektledning	4 827	1 573 602
Påslag (179 kr/timme)	42 454	7 599 266
Hårdvara		434 500
Datorer och skärmar x 6		90 000
A3 Skannrar x 3		133 500
A0-skanner x 1		211 000
Programvaror		75 512
Omnipage Professional x 1		4 500
Adobe Acrobat x 5		18 150
Document Express with DjVu x 1		49 362
Photoshop Elements x 5		3 500
Resor		50 000
Totalt		19 210 596

Skanningsarbete är beräknat på SGUs standardpris för tim- och arvodesanställd personal, grupp Z8, 150 kr/tim. Sorterings- metadatasättnings- och konverteringsarbete är beräknat på standardprisgrupp B5, 263 kr/tim. Projektledningsarbete är beräknat på ett timpris av 326 kr, standardprisgrupp B4.

Utbildningstid för projektets medarbetare är uppskattad till fyra veckor per person.

Utlokalisering av digitalisering av geofysikmaterial är beräknat på den billigaste offerten, den från Astacus AB (se sid. 27). Digitalisering av mätvärden och skanning ger en kostnad på 1 646 000 kr. Fraktkostnad är adderad med 54 000 kr, något som är en osäker kostnad beroende på hur materialet ska fraktas.

Påslaget är SGUs overheadkostnader och är beräknat på det lägsta påslaget för 2006, 179 kr per arbetad timme.

Datorer och skärmar för personal uppgår till 90 000 kr och omfattar sex datorer med 2 GB minne och 250 GB hårddisk och 20"-skärmar.

A3-skannrar kostar 44 500 kr per styck och sammanlagt tre skannrar förväntas behövas under digitaliseringsprojektets levnadstid.

A0-skanner till projektet förväntas kosta mellan 200 000 till 250 000 kr. Priset på den skanner vi anser optimal för projektet kostar 211 000 kr i nuläget.

Sammanfattning av datalagringsbehov

Behovet för diskutrymme beror på upplösning och format av de producerade filerna.

Det optimala är att skanna all text i svart-vitt med 200-300 dpi upplösning. Filstorleken kommer då att minimeras. En annan fördel med detta alternativ är att utskriftskvaliteten oftast blir bättre därför att kontrasten i svart-vit dokument är högre än i gråskala. OCR-konvertering fungerar dessutom oftast bättre på sidor med god kontrast.

Figurer, kartor, ritningar osv. bör däremot skannas i gråskala eller färg till TIFF format som ger större filstorlekar.

En uppskattning baserat på en genomgång av delar av materialet och testskanning av olika dokumenttyper indikerar att det totala lagringsbehovet är ca 14 TB. En detaljerad översikt över datalagringsbehovet till de olika materialtyperna kan ses nedan. Detta är en enormt stor datamängd och kräver investering i ett datalagringsystem som kan hantera så stora datamängder på ett effektivt sätt. Vi föreslår att IT-avdelningen i Uppsala får ansvaret för att utvärdera och ev. upphandla detta.

Av det totala lagringsbehovet på ca 14 TB så kommer geologikartor, geokemikartor, geofysikkartor, urankartor och gruvkartor utgöra > 90%. Det beror på att dessa kartor blir stora bildfiler i TIFF format. Om man väljer ett bildformat som kräver mindre utrymme så kan lagringsbehovet minskas.

Det kan vara värt att notera att om man sparar hela arkivet i DjVu format så indikeras ett totalt lagringsbehov på < 1 TB.

Dokumenttyp	Filformat	Antal	Storlek per st. (MB)	Totalt utrymmesbehov (GB)
Prosp. rapporter	pdf	7 000	22	153
Geofysik	tiff	18 000	100	1 800
Borrhålsparmar	pdf	700	411	288
Geologikartor	tiff	25 000	143	3 575
Borrhålskartor	tiff	4300	84	360
Geokemipärmar	pdf	44	175	8
Geokemikartor	tiff	25 000	143	3 575
Block- & hållpärmar	pdf	164	150	25
Hällobes & dagböcker	pdf	395	200	79
Inbundna dagböcker	pdf	288	30	9
Uranpärmar	pdf	612	296	181
Urankartor	tiff	10 000	110	1 100
Bergst. gruvkartor	tiff	20 000	143	2 860
Bergst. handlingar s.	pdf	170 000	0,63	107
Projektpärmar	pdf	183	180	33
Geofysik inbundna	pdf	267	30	79
Geofysik pärmar	pdf	166	84	14
Totalt				14 246

Sammanfattande rekommendationer för digitalisering av hela arkivet

Personal

Ett storskaligt digitaliseringsprojekt kräver olika slags kompetenser. Utifrån tidsbehov för olika delmoment under digitaliseringen (se sid. 48) föreslås följande typ av personalstyrka:

- 18 402 timmar utförs av personal med någon form av geovetenskaplig bakgrund. Denna kompetens behövs i arbetsmoment som t.ex. metadatasättning, sortering och uppdatering av kartor.
- 18 265 timmar utförs av personal utan geovetenskaplig kompetens. Olika arbetsmoment som kan utföras av denna personal inkluderar bl.a. skanning och produktion av filer. Till detta krävs det inga formella geovetenskapliga kunskaper, men egenskaper som noggrannhet, ansvarskänsla, datorvana och kommunikationsförmåga är nödvändig.
- En koordinator, projektledaren, som skall sköta den dagliga arbetsledningen, uppföljning, kvalitetskontroll och problemlösning, bör ha geovetenskaplig bakgrund.
- Befintlig kompetens vid Mineralinformationskontoret bör vara tillgänglig för projektet ett antal timmar för frågor rörande materialet.
- IT avdelningen i Uppsala måste involveras i frågor som berör datalagring, back-up av filer och databaser.

Prioriteringsordning och arbetsplan

Allt material i arkivet kan vara intressant ur prospekteringssynpunkt, ändå efterfrågas vissa typer av material mer än andra. I innehållsförteckningen av denna rapport listas de olika materialtyperna upp. Det är den ordning som föreslås gälla rörande ordningsföljd av det som skall digitaliseras. Vad gällande Bergstatens gruvkartor bör de eventuellt få högre prioritet än vad som föreslagits. Prioriteringsordningen kan komma att ändras om efterfrågan från kund väsentligen förändras. Det går också att tänka sig att olika materialtyper digitaliseras parallellt.

Uppstartsfasen till projektet (en till två månader) bör användas till att bygga upp en ”produktionslinje” i arkivet. Detta inkluderar inköp och installation av utrustning, testkörning, lösning av praktiska problem etc. Detta bör vara avklarat innan anställning av personal för skanning för att undvika att förseningar i inkörningsfasen drabbar full personalstyrka.

Det totala antalet timmar för digitaliseringsarbete, exklusive projektledare uppskattas till ca 36 700 timmar. Ett manår beräknas på SGU vara 1609 timmar. Om det anställs tre personer för skanningsarbete så kan skanningen slutföras inom fyra år. Tre personer arbetar parallellt med sortering, metadatasättning och OCR-konvertering.

Detaljerad arbetsplan bör upprättas för effektiv synkronisering av dessa två personalgrupper.

Beroenden

Skanningen skall företas i arkivet, platsen där dokument och kartor är lokaliserade. Filialen i Malå har för 2006 fått finansiering för en datalinje mellan kontorsbyggnaden och pappersarkivet. Det krävs att datalinjen är färdiginstallerad före ett digitaliseringsprojekt kan starta.

Ett system för lagring av producerade datafiler måste vara implementerad innan skanningen påbörjas.

Det finns även krav att ställa på den fysiska arbetsmiljön. Det måste avsättas plats i arkivet för utrustning, färdigställande av arbetsplatser med nätverksanslutning etc. Kontorsutrustning (stolar, arbetsbord, hyllor m.m.) måste vara på plats före ett digitaliseringsprojekt kan starta upp.

Risker

Projektet antas att ha en levnadstid på flera år, något som kan medföra att personal slutar eller byts ut. Det påför därmed projektet extra belastning t.ex. i form av vakanser och upplärning av nya medarbetare.

Om projektledaren eller någon annan i projektet, är medarbetare vid Mineralinformationskontoret finns det risk att kontorets programrelaterade arbete (t.ex. service till kontorets kunder) får prioritet på bekostnad av digitaliseringsprojektet. Det medför att tids- och kostnadsberäkningarna för projektet kommer att öka.

OCR

Vi har i projektet utvärderat OCR funktionen i Adobe Acrobat 7.0, Omnipage Professional 15 och Document Express with DjVu. De tester som vi har utfört med OCR programvarorna indikerar att delar av materialet, t.ex. borrhålsparmar och geokemiska analyser, är svårt att OCR läsa på ett meningsfullt sätt (se t.ex. fig. 53). Textdokument av god kvalitet (t.ex. prospekteringsrapporter) kan däremot OCR läsas med ovanförmända program och konverteras utan större problem.

Det rekommenderas att svårt material skannas till vanliga PDF-filer först. Därefter kan data med hög prioritet extraheras med en kombination av OCR läsning och manuell inmatning.

Det skall noteras att finns flera andra OCR-produkter på marknaden som vi ej har testat t.ex. programmet Readiris Pro 10.

Användning av extern digitaliseringsbyrå

Huvuddelen av digitaliseringen utförs på plats i Malå för att behålla arkivets funktionalitet. Det finns dock geofysiskt material som bedöms att vara alltför svårdigitaliserat och tidskrävande. Tre olika digitaliseringsbyråer har bedömt detta material och lämnat offerter.

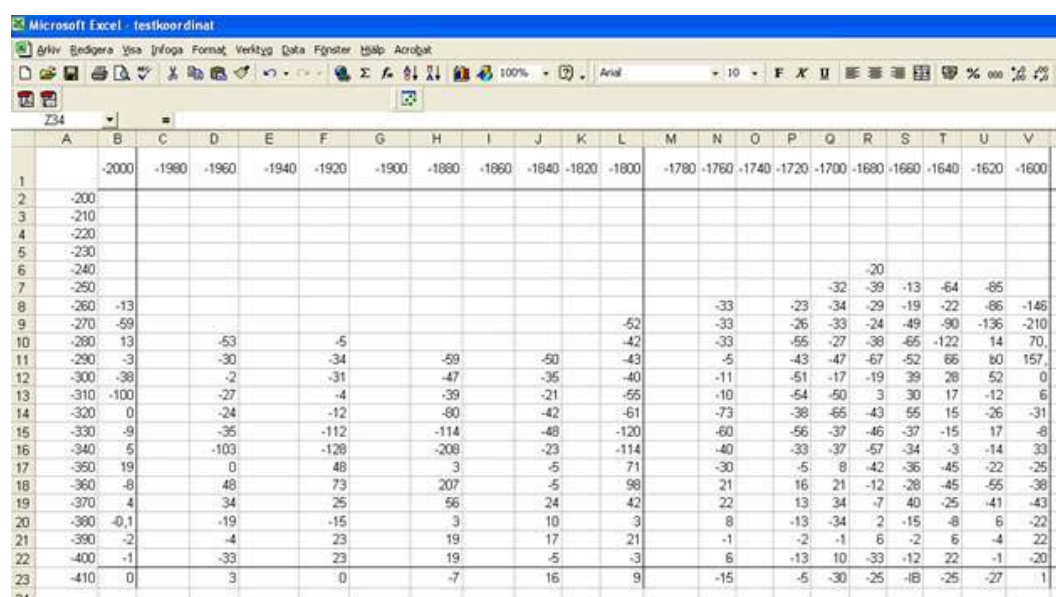
Bilaga 1: Testskanning av markgeofysikmaterial

För att testa metoder för att skanna markgeofysiska data så blev det i ett inledande skede av projektet valt ut ett begränsat material. Testmaterialet består av geofysikkartor som innehåller:

- 1) A1 karta (transparent folie) med rutnät innehållande koordinatinformation (lokalt staksystem) och siffror med mätvärden utsatta på kartan bredvid mätpunkten.
- 2) A1 karta (transparent folie) med rutnät innehållande koordinatinformation (lokalt staksystem) och siffror med mätvärden utsatta på kartan bredvid mätpunkten. Det finns i tillägg konturlinjer som binder ihop områden med lika mätvärden.

Detta experiment blev utfört i en tidig fas av projektet med en A4 skanner, något som medför att varje A1 karta måste skannas i flera delar. Detta har emellertid ingen betydelse för utvärderingen av huruvida det är möjligt att OCR skanna mätvärdena. Efter skanning blev mätvärden inlästa med OCR programvaran Omnipage och omformat till text. Därefter blev talvärdena exporterade till MS Excel. Placeringen till mätpunkten på kartan ger koordinatinformationen till mätvärdena. Detta betyder att punkternas (mätvärdens) placering måste behållas efter OCR konverteringen för att inte förlora denna koordinat informationen. Detta kan åstadkommas under konverteringen till kalkylbladsformat med att de individuella cellerna i kalkylbladet utgör ett koordinatsystem (fig. 1.1).

Man kan t.ex. sätta in lokala x- och y- koordinater i MS Excel och använda en formel för att extrahera värdena (fig. 1.2). För att undvika att få värdet "0" där värde saknas kan en fast x-koordinat och en varierande y-koordinat användas.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1		-2000	-1980	-1960	-1940	-1920	-1900	-1880	-1860	-1840	-1820	-1800	-1780	-1760	-1740	-1720	-1700	-1680	-1660	-1640	-1620	-1600
2		-200																				
3		-210																				
4		-220																				
5		-230																				
6		-240																				
7		-250																				
8		-260	-13																			
9		-270	-59																			
10		-280	13																			
11		-290	-3																			
12		-300	-38																			
13		-310	-100																			
14		-320	0																			
15		-330	-9																			
16		-340	6																			
17		-350	19																			
18		-360	-8																			
19		-370	4																			
20		-380	-0,1																			
21		-390	-2																			
22		-400	-1																			
23		-410	0																			

Figur 1.1. MS Excel blad som visar exempel på OCR skannad markgeofysikmätning. Placeringen till mätpunkten anger koordinaten till mätningen i det i lokala staksystem (rad 1 och kolumn 1).

Bilaga 2: Metadatastruktur för markgeofysikmaterial

Följande metadatastruktur för olika typer av markgeofysikdata kan användas:

VLF	SLINGRAM	MAGNET	IP	GRAVIMETRI
LOK_N_S	LOK_N_S	LOK_N_S	LOK_N_S	IDNR
LOK_E_W	LOK_E_W	LOK_E_W	LOK_E_W	OBJEKT
RE	RE	OBS.T,	IP	METOD
IM	IM	MMPT, 2 , 6	RES	PARAMETER
MVHR	RE	DATUM	SP	SCALEFACTOR
MVHI	IM	PROFILNR	DATUM	YEAR
DATUM	DATUM1	XRAK	PROFILNR	LOK_N_S
PROFILNR	PROFILNR	YRAK	XRAK	LOK_E_W
XRAK	XRAK		YRAK	XRAK
YRAK	YRAK			YRAK
				VALUE

Bilaga 3: Programvara

Document Express with DjVu



<http://www.lizardtech.com>

Programmet återförsäljs i Sverige av bl.a. Headway Technology och en prisförfrågan till denna leverantör den 10 oktober gav följande:

1 LizardTech Document Express Professional Edition 5.0

NOTE! Special introduction price during 2005 Version 5.0 GUI-based application for batch conversion of unlimited scanned documents. First year support included. OCR for full text searching. Also includes virtual printer. Includes: Multi-page scanning capability via TWAIN, Full Annotation capability, Batch Virtual Printer. Free Plugin viewer available over the web. System Requirements: 600MHz PC, 128MB of RAM, CD ROM, Win 95/98/2000 SP2, NT4 SP6, IE5 or above or Netscape 4.0 - 4.7

Unit Retail Price 3 002 SEK Part No: 201-048-1

2 LizardTech Document Express Enterprise Edition 5.1 Win ECM Ver

Version 5.1 ECM Bundle includes 25,000 page cartridge first year of support and one Professional Editor License. Version 5.1 GUI-based application for batch conversion of 25,000 scanned documents. OCR for full text searching. Includes: Multi-page scanning capability via TWAIN, Full Annotation capability, Batch Virtual Printer. Free Plugin viewer available over the web. System Requirements: 600MHz PC, 128MB of RAM, CD ROM, Win 95/98/2000 SP2, NT4 SP6, IE5 or above or Netscape 4.0 - 4.7

Unit Retail Price 11 391 SEK Part No: 201-053-1

3 LizardTech Document Express Version 5.1 25,000 Page Cartridge Refill

Allows Document Express Enterprise to convert 25,000 images to DjVu.
This is a licence key code.

Unit Retail Price 3 791 SEK Part No: 201-053-6

4 LizardTech Document Express Enterprise Edition 5.0 UE

NOTE! Special introduction price during 2005. Version 5.1. Enables high-volume conversion of colour or bitonal scanned documents into DjVu format. Support for 1st year included. Document Express Enterprise Edition contains a GUI-based application for Windows platforms. OCR for full text searching. Includes: Unlimited processing capacity, first year support, three copies of Desktop. Works using "Watched Folders." Can Convert from DjVu back to Tiff if required. System Requirements: 400 Mhz PC, 128MB RAM, CD-ROM drive, Win 98/2000 SP2 or NT4 SP6. Sufficient hard disk space for original images, output images, and temporary files created during the encoding process.

Unit Retail Price 49 362 SEK Part No: 201-050-1

Beskrivning av Document Express with DjVu Enterprise Edition

DjVu använder en segmenteringsteknik som separerar text från grafik och bilder. Generellt är text det som ställer högsta krav på kompressionsteknik. Detta är på grund av att text har skarpa kanter och linjer som har en tendens till att bli något "taggiga" vid kompression. Vid en separation av texten till ett separat lager kan man använda aggressiva komprimeringsalgoritmer på grafiken som oftast är mindre känslig för kompression, medan texten komprimeras separat med en mindre aggressiv teknik. Detta medför att kompressionsförhållande för sidan som helhet blir större än vid traditionell kompression.

Typiskt kompressionsförhållande för DjVu är mellan 300:1 och 1000:1. Ett A4 dokument i DjVu format med 300-400 dpi upplösning vill typiskt ha en storlek på

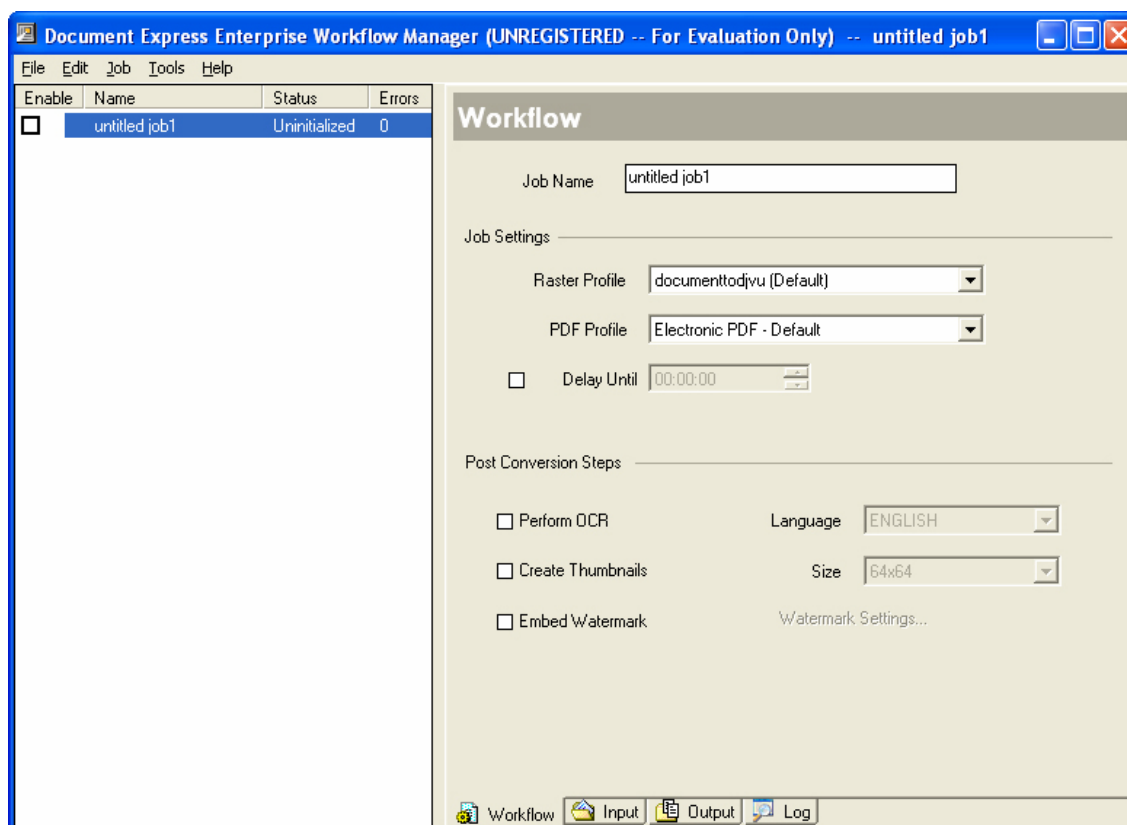
endast 50 KB. DjVu dokumentbilder är uppemot 1000 gånger mindre än TIFF filer, och mellan 10-100 gånger mindre än JPEG och PDF filer.

DjVu segmenterar en sida på följande sätt:

- ”Foreground layer” – här finns bl.a. text och linjer. Detta lager segmenteras vidare i ett svart-vitt lager och ett färglager.
- ”Background layer” – detta lager innehåller foto, grafik etc.

När en skannad sida är segmenterad så uppnås dessutom bättre resultat för OCR när man skannar komplexa sidor som innehåller mycket färg etc. därför att texten finns i ett separat lager.

Programmet har ett mycket enkelt användargränssnitt som gör det smidigt att använda. Stommen i programmet är en så kallad ”Work-Flow Manager” (fig. 3.1) som gör det lätt att organisera konverteringsjobb med olika inställningar. Man kan spara olika ”jobs” där det är möjligt att bl.a. definiera upplösning, OCR (med ordlista för många språk) med mera.



Figur 3.1. Workflow manager i Document Express.

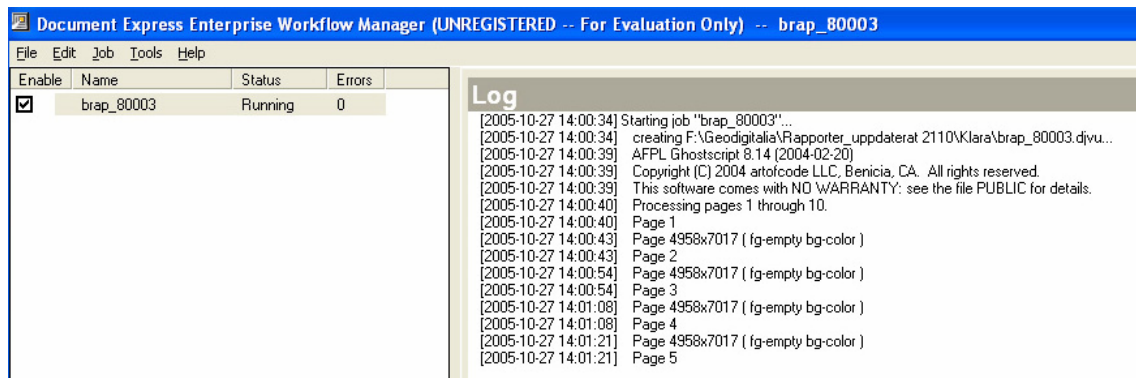
En mycket bra finess är s.k. ”Watched Folder” som möjliggör automatisk övervakning av en specifik mapp där man kan lägga filer som skall konverteras till DjVu format. Detta är praktiskt om man skannar dokument till PDF eller TIFF format först, och därefter konverterar till DjVu. Om man till exempel använder en nätverksskanner så kan man t.ex. skanna PDF eller TIFF filerna till en mapp på en server i nätverket som övervakas av en ”Watched Folder” Job. Dessa filer konverteras

därefter till DjVu format automatiskt. Applikationen flyttar därefter originalet till en förhandsdefinierad mapp.

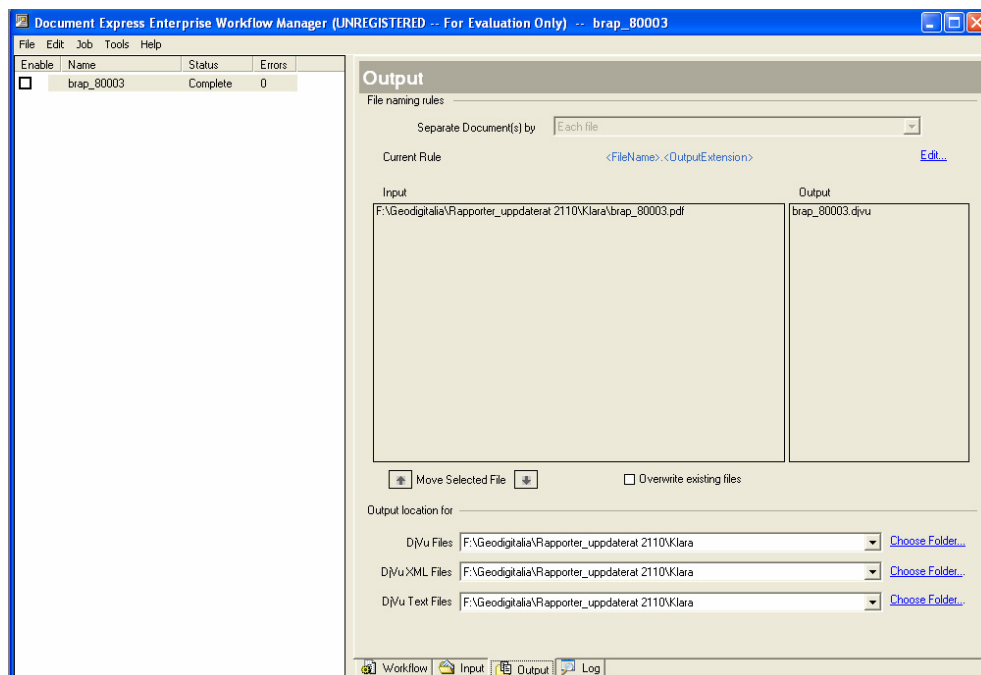
I ”Log”- fönstret kan man se status till konverteringsprocessen (Fig 3.2). Här vill eventuella fel som uppkommer under konverteringen visas.

Det finns möjlighet för att exportera OCR text från DjVu filer till XML (fig. 3.3).

”Digital To DjVu”- är en komponent i programmet som möjliggör konvertering direkt från digitala dokument (ex. MS Word, MS Excel, MS Powerpoint, HTML, PDF etc.) till DjVu. Komponenten har inte provats i pilotprojektet, men på Lizarstechs hemsida så kan man läsa att denna funktion inte konverterar dokumentet till en bitmap först, utan programmet analyserar sidelementen (ord, bilder, linjer) och avgör om elementet skall placeras i ”foreground” eller ”background” lagret. Detta motsvarar segmentationsprocessen som används på skannade dokument. Därefter komprimeras dokumentet som vanligt.



Figur 3.2. Log fönstret visar progression och status under konverteringen.



Figur 3.3. Output fönstret har bl.a. möjlighet för att exportera det dolda OCR-text lagret som XML.

Bilaga 4: Skannrar och annan hårdvara

SGU har genom VHS avtalen åtta avtalsleverantörer av skannrar att välja från. En förteckning över dessa finns på:

http://www.avropa.nu/templates/ramavtalsomrade____86.aspx

A3 skannrar till ett fullskaligt digitaliseringsprojekt kan t.ex. vara av typen:

Ricoh Aficio IS760D



Allmänt

Skannertyp: A3 flatbedsskanner för dubbelsidigt och färg med automatisk dokumentmatare
Upplösning: Optisk: 400 x 400 dpi
Utskrift: 100/200/300/400/600 dpi (75 - 800 dpi/i steg om 1 dpi)
Bitdjup: 24 bits/pixel
Skanningshastighet: Svartvitt: 70 ppm (116 spm)₁
Fullfärg: 38 ppm (60 spm)₁
Originalformat: ADF: Min: 69 x 90 mm
Max: 305 x 2 000 mm
Täckplatta: Min: 5 x 5 mm
Max: 305 x 432 mm
ADF kapacitet: 200 ark (A4, 64 g/m₂)
ADF dokumentvikt: 40 - 157 g/m₂
Exponeringskontroll: 7 nivåer₃
Förinställda grader: 7 förminskningar & 5 förstoringar₃
Zoom: 25 - 400% (i steg om 1%)₃
Uppvärmning: Mindre än 20 sekunder (4 sekunder från energisparläge)
Mått
(B x D x H): 490 x 695 x 332 mm
Vikt: Mindre än 34 kg
Strömkälla: 220 - 240 V, 50/60 Hz
Strömförbrukning: Max: Mindre än 180 W
Energisparläge: Mindre än 11 W

(Nätverks-) anslutning

Gränssnitt: Standard: SCSI3
USB 2.0
Tillval: FireWire (IEEE 1394)₂
Ethernet 10 base-T/100 base-TX₃
Drivrutiner: TWAIN, ISIS
Protokoll: TCP/IP
Nätverkshantering: SNMP

Uppgifterna är hämtade ur Ricohs reklambroschyr.

Bilaga 5: Digitaliserat material i Geodigitalia

Prospekteringsrapporter (421 st. kompletta med alla bilagor)

B_8513	K_8625	Prap_83017	Prap_83088	Prap_83517	Prap_84551	S_8703
B_8804	K_8627	Prap_83018	Prap_83089	Prap_83520	Prap_84552	
B_9004	K_8629	Prap_83019	Prap_83090	Prap_83521	Prap_84554	
B_9008	Ki_8219	Prap_83020	Prap_83091	Prap_83523	Prap_84555	
Brap_00779	Ki_8220	Prap_83021	Prap_83092	Prap_83524	Prap_84556	
Brap_00888	Ki_8221	Prap_83022	Prap_83093	Prap_83525	Prap_85001	
Brap_00916	Ki_8328	Prap_83023	Prap_83094	Prap_83526	Prap_85004	
Brap_00917	Ki_8329	Prap_83024	Prap_83095	Prap_83527	Prap_85030	
Brap_75001	Ki_8330	Prap_83025	Prap_83096	Prap_83528	Prap_85031	
Brap_77004	Ki_8331	Prap_83026	Prap_83097	Prap_83529	Prap_85500	
Brap_77006	Ki_8332	Prap_83027	Prap_83098	Prap_83530 del1	Prap_85502	
Brap_77007	Ki_8333	Prap_83028	Prap_83099	Prap_83531	Prap_85504	
Brap_78005	Ki_8334	Prap_83029	Prap_83100 del1	Prap_83532	Prap_85505	
Brap_78503	Ki_8405	Prap_83030	Prap_83100 del2	Prap_83533	Prap_85511	
Brap_79008	Ki_8406	Prap_83031	Prap_83101	Prap_83534	Prap_85513	
Brap_79022	Ki_8407	Prap_83032	Prap_83102	Prap_83535	Prap_85516	
Brap_79023	Ki_8408	Prap_83033	Prap_83103 del1	Prap_83537	Prap_85520	
Brap_79028	Ki_8411	Prap_83034	Prap_83103 del2	Prap_83538	Prap_85525	
Brap_79047	Ki_8412	Prap_83035	Prap_83104	Prap_83541	Prap_85527	
Brap_79056	Ki_8413	Prap_83036	Prap_83105	Prap_83547	Prap_85532	
Brap_79058	Ki_8414	Prap_83037	Prap_83106	Prap_83548	Prap_85533	
Brap_79403	Ki_8415	Prap_83038	Prap_83107	Prap_83555	Prap_85536 del1	
Brap_80002	Ki_8416	Prap_83039	Prap_83108	Prap_83558 del1	Prap_85536 del2	
Brap_80003	Ki_8417	Prap_83040	Prap_83109	Prap_83558 del2	Prap_85538	
Brap_80004	Ki_8418	Prap_83041	Prap_83110	Prap_83559	Prap_85542	
Brap_80005	Ki_8420	Prap_83042	Prap_83111	Prap_83560 del1	Prap_85547	
Brap_80006	Ki_8421	Prap_83043	Prap_83112	Prap_83560 del2	Prap_85551	
Brap_80007	Ki_8422	Prap_83044	Prap_83113	Prap_84002	Prap_86047	
Brap_80008	Ki_8423	Prap_83045	Prap_83114	Prap_84004	Prap_86088	
Brap_80009	Ki_8424	Prap_83046	Prap_83115	Prap_84005	Prap_86513	
Brap_80015	Ki_8425	Prap_83047	Prap_83116	Prap_84006	Prap_86520	
Brap_80024	Ki_8426	Prap_83048	Prap_83117	Prap_84008	Prap_86539	
Brap_80026	Ki_8427	Prap_83049	Prap_83118	Prap_84009	Prap_86549	
Brap_80028	Ki_8428	Prap_83050	Prap_83119	Prap_84010	Prap_86559	
Brap_80030	Ki_8435	Prap_83051	Prap_83120	Prap_84011	Prap_87006	
Brap_80033	Ki_8436	Prap_83052	Prap_83121	Prap_84012	Prap_87026	
Brap_80037	Ki_8439	Prap_83053	Prap_83122	Prap_84072	Prap_87526	
Brap_80039	Ki_8441	Prap_83054	Prap_83123 del1	Prap_84060	Prap_87533	
Brap_80040	Ki_8442	Prap_83055	Prap_83123 del2	Prap_84089	Prap_88030	
Brap_80042	Ki_8507	Prap_83056	Prap_83124	Prap_84105	Prap_88041	
Brap_80043	Ki_8531	Prap_83057	Prap_83125	Prap_84501	Prap_88507	
Brap_80044	Ki_8532	Prap_83058	Prap_83126	Prap_84502	Prap_88520	
Brap_80046	Ki_8540	Prap_83059	Prap_83127	Prap_84504	Prap_88541	
Brap_80048	Mink_0330	Prap_83060	Prap_83128	Prap_84506	S_8401	
Brap_80414	Mink_0331	Prap_83061	Prap_83129	Prap_84507	S_8402	
Brap_80531	Mink_0332	Prap_83062	Prap_83130	Prap_84509 del1	S_8403	
Brap_81024	Mink_0333	Prap_83063	Prap_83131	Prap_84509 del2	S_8404	
Brap_81032	Mink_0334	Prap_83064	Prap_83132	Prap_84510	S_8405	
Brap_81050	Mink_0335	Prap_83065	Prap_83133	Prap_84511	S_8407	
Brap_81058	Mink_0336	Prap_83067	Prap_83134	Prap_84512	S_8408	
Brap_81066	NSG_89124	Prap_83068	Prap_83135	Prap_84514	S_8409	
Brap_81077	NSG_90044	Prap_83069	Prap_83136	Prap_84515	S_8410	
Brap_81078	NSG_90062	Prap_83070	Prap_83137	Prap_84516 del1	S_8411	
Brap_81084	NSG_90112	Prap_83071	Prap_83138	Prap_84518	S_8412	
Brap_82062	Prap_83001	Prap_83072	Prap_83139	Prap_84520	S_8415	
Brap_82090	Prap_83002	Prap_83073	Prap_83501	Prap_84521	S_8416	
Brap_82109	Prap_83003	Prap_83074	Prap_83502	Prap_84523	S_8417	
Brap_83515,	Prap_83004	Prap_83075	Prap_83503 del 1	Prap_84524	S_8419	
brap_83700-704	Prap_83005	Prap_83076	Prap_83503 del2	Prap_84525	S_8420	
Brap_83686	Prap_83006	Prap_83077	Prap_83504	Prap_84527	S_8421	
Brap_83687	Prap_83007	Prap_83078	Prap_83505 del1	Prap_84528	S_8427	
Brap_83688	Prap_83008	Prap_83079	Prap_83505 del2	Prap_84530	S_8506	
Brap_83689	Prap_83009	Prap_83080	Prap_83506	Prap_84531	S_8524	
Brap_83690	Prap_83010	Prap_83081	Prap_83507	Prap_84535	S_8527	
Brap_83691	Prap_83011	Prap_83082	Prap_83509	Prap_84542 del2	S_8528	
Brap_83692	Prap_83012	Prap_83083	Prap_83510	Prap_84544	S_8529	
Brap_83801	Prap_83013	Prap_83084	Prap_83512	Prap_84546	S_8602	
Brap_87003	Prap_83014	Prap_83085	Prap_83513	Prap_84548	S_8609	
K_8612	Prap_83015	Prap_83086	Prap_83515	Prap_84549	S_8616	
K_8615	Prap_83016	Prap_83087	Prap_83516	Prap_84550	S_8618	

Prospekteringsrapporter: (DjVu format. 35 st. kompletta)

Brap_75001
Prap_83004
Brap_77007
Prap_83005
Brap_79008
Prap_83006
Brap_79056
Prap_83009
Brap_80003
Prap_83012
Brap_80004
Prap_87026
Brap_80006
Ki_8220
Brap_80008
Ki_8411
Brap_80009
Brap_80015
Brap_80024
Brap_80026
Brap_80042
Brap_80044
Brap_80046
Brap_81050
Prap_83001
Brap_83003
Ki_8412
Ki_8423
Ki_8615
Ki_8627
S_8402
S_8403
S_8404
S_8410
S_8618

Borrhålsprotokoll (21 st. kompletta skannade till pdf).

Pärm nr	Filnamn/Borrhålsnr
26	Bjm_26
27	Bjm_27
28	Bjm_28
29	Bjm_29
246	ROUTEVARE/ROVASUANDO
286	STEGRUVAN
287	STE1
288	STE61
289	STE101
290	STE59001
291	STE61001
292	STE63001
293	STE63301
294	STE200
295	STE64101
296	STE68001
297	STE70001
298	STE79100
299	STE82101
323	TJARROJÄKKA
325	TJARROJÄKKA

Geokemianalyser (13 st. kompletta pärmar skannade till pdf).

Block- och hållpärmar (13 st. kompletta pärmar skannade till pdf).

Uranpärmar (7 st. kompletta pärmar skannade till pdf).

Projektpärmar (2 st. kompletta pärmar skannade till pdf).

Geologi och borrhålskartor (56 kartor i A0 är skannade till TIFF och metadatasatta).