

# SCB:s erfarenheter av digitalisering av Bidrag till Sveriges officiella statistik (BiSOS)

Rolf-Allan Norrmosse, SCB

## Dagordning

1. Bakgrund
2. Digitaliseringsprocessen
  - Bildfångst och bildfiler
  - OCR-tolkning
  - Framställning av publiceringsfärdig pdf
3. Utveckling och förnyelse
4. Resultat och förbättringar i digitala BiSOS jämfört med den tryckta förlagan
5. Presentation av BiSOS
6. Användning av BiSOS återstående medel

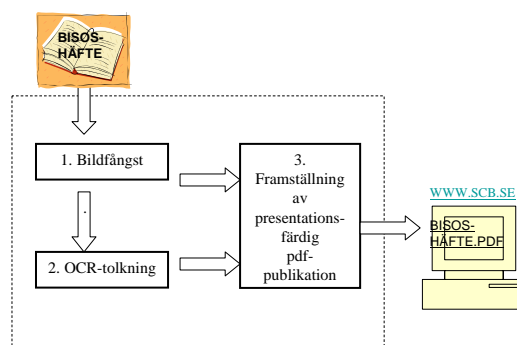
## BISOS (Bidrag till Sveriges officiella statistik)

- Svensk officiell statistik i 23 ämnesområden
- Publicerad 1859-1920 (1830-1917)
- Text, tabeller, utvik i svartvit, gråskala, färg
- Digitaliserad september 2007-april 2009

## 2006. Riksbankens Jubileumsfond

- Ansökan av digitalisering av BISOS i februari.
- 4 miljoner kr beviljades den 26 oktober för BISOS digitalisering 2007-2009.
- Tjänsten BiSOS digitalisering upphandlades våren 2007.
- Avtal med Logica, f.d. WM-data 2 juli 2007
- Utförare av tjänsten BISOS digitalisering är Infodisk Media, Riga – Logicas underleverantör

# Digitaliseringsprocessen



2009-10-14

Mötesplats inför framtiden 2009 - Rolf-Allan Norrmosse, SCB

5

## Bildfångst och bildfiler

- Tiff-format och 300 dpi
- Urn:nbn ska användas för häften och bilder.
- Varje sida en bildfil i svartvit / gråskala / färg med olika färgdjup
- Bilderna ska vara upprätade, beskurna, geometriskt korrigerade.
- Liggande tabeller/texter ska vridas till stående
- Tabeller över uppslag är två sidor och två bilder rader ska mötas horisontellt motsvarande för liggande tabeller ska kolumner mötas vertikalt
- Kontroll av bilder, kalibrering och dagligt underhåll

2009-10-14

Mötesplats inför framtiden 2009 - Rolf-Allan Norrmosse, SCB

6

## OCR-tolkning

- Slagning mot svensk ordlista (Dahlins ordbok 1855)
- OCR-tolkning: 99,9 % korrekthet på ord- och teckennivå på svensk text samt 99,4 % på fransk
- Annan text – OCR-tolkning: 85 % korrekthet på ord och teckennivå

## Utveckling och förnyelse (1)

- Granskat leverantörens arbetsmetoder och rutiner för att minimera SCB:s arbete med kvalitetssäkring
- Olof Dahlins ordbok 1855. Använd som ordlista vid teckenigenkänning (OCR-tolkning)
- Utnyttjat en den unik identifikatorn urn:nbn för häften och bilder
- Bevarat BiSOS paginering i pdf:ens sätt att paginera
- I den digitala versionen har införts

## Utveckling och förnyelse (2)

- Dokumentegenskaper med bibliografiska metadata, urn och filnamn
- Skapat beskrivningsblad för att kommunicera med leverantören av tjänsten
- Skapat innehållsförteckningar som saknas i 477 av 1495 häften
- Infört start settings för att presentera det digitala häftet på ett optimalt sätt för användaren

## Resultat

- Informationen i ett BiSOS-häfte ska vara fullständig. Saknas en sida, är skadad och inte finns hos SCB har sidan eller häftets lånats in från ett annat bibliotek för digitalisering. Den digitala versionen av häftet har det fullständiga exemplaret
- Bokverket BiSOS finns att läsa på SCB:s webbplats, [www.scb.se/bisos](http://www.scb.se/bisos)
- Alla häften, både tryckt och digitalt, håller på att registreras i LIBRIS
- Bildfiler i formatet tiff är av så hög kvalitet att vi får nytta av det i vårt kommande projekt

## Förbättringar i digitala BISOS jämfört med den tryckta förlagan

- Skapad innehållsförteckning om den saknas
- Innehållsförteckning alltid i början av ett häfte även om den finns på annan plats.
- Klickbara innehållsförteckningar
- Länk till annan innehållsförteckning eller tabellförteckning i innehållsförteckningen
- Bokmärken till lägsta nivå
- Liggande tabeller eller text vrides till stående
- Serie- och/eller volymuppgift visas alltid på titelsidan även om den saknas i den tryckta förlagan
- Inledningssida – vänstersida till titelsidan - med föregångare, efterföljare och översiktspublikation tillika med om och digitaliseringsinformation

2009-10-14

11

## Presentation av BISOS

- Pdf finns under Historisk statistik på SCB:s webbplats under Hitta statistik > Historisk statistik > BiSOS. Adress: [www.scb.se/bisos](http://www.scb.se/bisos).
- Kort presentation av serierna på BISOS seriesidor. T.ex. [Skogsväsendet](#)
- I SCB:s publiceringskalender på SCB:s webbplats (ännu inte klart)
- I LIBRIS katalogisering på häftesnivå av

2009-10-14

Mötesplats inför framtiden 2009 - Roll-Åkeri, Norrmösse, SCB

12

## Historisk statistik på [www.scb.se](http://www.scb.se)

Historisk statistik på SCB:s webbplats  
[www.scb.se](http://www.scb.se) > hitta statistik > historisk  
statistik

- BISOS. Adress: [www.scb.se/bisos](http://www.scb.se/bisos)
- [Serien Historisk statistik för Sverige](#)
- Annan historisk statistik. T.ex. [SCB:s stencilerade publikationer](#)

## BiSOS återstående medel

- Av 4 miljoner kr till BiSOS digitalisering har SCB använt 1,3 miljoner
- SCB har skrivit till RJ och lämnat förslag på användning av BiSOS återstående medel. RJ har beslutat att SCB får disponera återstående medel under 2009 och 2010 med slutrapportering 24/3 2011

## 1. Framtagande av verktyg för analys och teckentolkning av tabeller (VATT) - förstudie

- SCB vill i en förstudie undersöka, om det finns något tabellverktyg tillgängligt på marknaden, som kan anpassas till en statistikproducents behov, eller, om det inte finns, vill SCB beräkna kostnaden för att SCB utvecklar verktyget.
- I förstudien skall från det material som SCB har digitaliserat tas fram ett antal typfall på hur tabeller kan vara uppbyggda sedan 1811.
- Klar i okt. 2010. Ansökan om externa medel (fond) för anpassning/tillverkning av VATT samt tillämpning av VATT vid digitalisering av

2009-10-14

Mötesplats inför framtiden 2009 - Rolf-Allan Norrmosse, SCB

## 2. Fortsatt digitalisering av tryckta statistiska publikationer

- Större delen av BiSOS återstående medel kommer att användas till digitalisering av äldre statistikpublikationer både från tiden före och efter BiSOS
- Blå rad i efterföljande tablåerna betyder att serien är klar att publicera eller publicerad på SCB:s webbplats

2009-10-14

Mötesplats inför framtiden 2009 - Rolf-Allan Norrmosse, SCB

16



Föregångare till BISOS. 9 av BISOS 23 serier har föregångare. 27 000 s.

A. Kongl. Tabell-Commissionens underdåniga berättelse till Kongl. Maj:t angående nativitetens och mortalitetens förhållande ... samt Rikets folkmängd ... (1811/15-1857)

B. Justitie Statsministerns underdåniga berättelse om brottsmålen och civile rättegångsärenderna i riket (1830-1856)

C. Bergs-Collegii underdåniga berättelse om förhållandet med bergs-handteringen (1833-1857)

D. Commerce-Collegii underdåniga berättelse om fabrikerens och manufacturernes ställning (1830-1857)

E. Commerce-Collegii underdåniga berättelse om Sveriges inrikes sjöfart (1828-1857)

F. Commerce collegii underdåniga berättelse om Sveriges utrikes handel och sjöfart (1828-1857)

K. Sundhets-Collegii underdåniga berättelse om

2	Statistisk årsbok för Sverige. 41 000 sidor	1914 -
3	Folk- och bostadsräkningen. 28 000 sidor	<del>1998</del> -
4	Valstatistik. 16 700 sidor.	1911 -
5	Befolkning. I huvudsak. 45 000 sidor	<del>2001</del> -
6	Index till Statistiska meddelanden, 1963-2001. A4. 4 000 sidor	<del>2003</del> -
		2001

2009-10-14

Mötesplats inför framtiden 2009 - Rolf-Allan Norrmosse, SCB

18

7	[Statistiska centralbyråns stencilerade publikationer]. Stencilerat. 6 000	1952-1963
8	Statistiska meddelanden. Ser A. Tillfälliga statistiska undersökningar.	1912-1953
9	Statistiska meddelanden. Ser. C. 4 000 sidor	1912-1953
1	Statistiska meddelanden. Ser. D. 10 000 sidor	1913-1954
0	Järnvägs-statistiska meddelanden.	1954
1	Statistiska meddelanden. Ser. E. 8 000 sidor	1912-1953
1	Uppgifter om bankerna. 11 000 sidor	1953
1	Sociala meddelanden (Statistiska meddelanden. Ser. F). 15 000 sidor	1912-1967
2009-10-14		
Mötesplats inför framtiden 2009 - Rolf-Allan Norrmosse, SCB		
19		

1	Statistisk tidskrift, inkl Sveriges officiella statistik i sammandrag 1870-1913.	1860
3	Sammandraget är före-gångare till Statistisk årsbok för Sverige. 15 000 sidor	-
		1912
1	Arbetsstatistik. Serier: A-F, H, L. Utgivna af K. Kommerskollegii afdelning för arbetsstatistik. 4 000 sidor	1899
4		-
		1910
1	Meddelanden från K. Kommerskollegii afdelning för arbetsstatistik. 9 000 sidor	1904
5		-
1	Ekonomisk översikt / utarbetad inom Kommerskollegium. 14 000 sidor	1899
6		-
1	Kommersiella meddelanden / utgiven av Kunskommerskollegium. 33 000 sidor	1910
7		-
1	Återstående monografier i serien Historisk statistik för Sverige	
8		
2009-10-14		
Mötesplats inför framtiden 2009 - Rolf-Allan Norrmosse, SCB		
20		