

Bokstäver och andra tecken inom IT i statsförvaltningen

Internetdagarna 2003-10-08

Wolf Arfvidson

Statskontoret

2003-09-17

Programmet

- **Definitioner och historik**
- **Dagens teckenstandarder**
- **Behov i Sverige**
- **Att beakta**
- **Svenska tangentbordet**
- **Framtidens teckenuppsättning i svenska statsförvaltningen**
- **Frågor/diskussion**

Definitioner och historik

- **Tecken och kodning**
- **Teckenuppsättning**
- **Versal, gemen och glyph-begreppet**

- **7-bitskod >>> 95 tecken**
- **8-bitskod >>> 191-223 tecken**
- **16-bitskod >>> 50000+**

Dagens teckenstandarder 1

- **ISO/IEC 646 IRV = ANSI 7-bits ASCII**
- **SS 63 61 27 utgåva 2, version för sv egennamn**
- **ISO/IEC 8859/1 med LA 1**
- **ISO/IEC Reg 204 med €**
- **ISO/IEC 8859/15 med LA 9**
- **RSVs folkbokföringsteckenuppsättning =
Bokstäver enligt 8859/1 utom y**
- **CP 850 PC Multilingual**
- **CP 278 EBCDIC Finland/Sverige**
- **CP 1252 Windows Western**

Dagens teckenstandarder - 2

- **ISO/IEC 10646-1 och -2 UCS (Unicode)**
- **ISO/IEC 9995-alla delar Keyboard layouts**
- **SS 66 22 41:2000 Svenskt tangentbordsmönster**
- **ISO/IEC 14755:1997 Input methods**
- **CEN CWA 13873:2000 European Subsets**
 - MES 1 335 tecken**
 - MES 2 1052 tecken**
 - MES 3 ~ 2900 tecken (MES 3B 2819 tecken)**
- **CEN TR 14381:2003 European Fallback Rules**

ISO 646 IRV code table
 compared to Swedish 7-bit Names code table

	0	1	2	3	4	5	6	7
0			SP	0	Q	P	`	p
1			!	1	A	Q	a	q
2			"	2	B	R	b	r
3			#	3	C	S	c	s
4			\$	4	D	T	d	t
5			%	5	E	U	e	u
6			&	6	F	V	f	v
7			'	7	G	W	g	w
8			(	8	H	X	h	x
9			)	9	I	Y	i	y
A			*	:	J	Z	j	z
B			+	;	K	E	k	C
C			,	<	L	\	l	I
D			-	=	M	J	m	}
E			.	>	N	^	n	~
F			/	?	O	_	o	

Characters in unshaded positions exist in both sets with the same coding.
 Characters in shaded positions do not exist in Swedish 7-bit Names.

NOTE – In early editions of ISO 646 IRV, the character CURRENCY SIGN was assigned to code 24. This character was replaced by the DOLLAR SIGN in the third edition 1991, making the IRV identical to ASCII.

Swedish 7-bit Names code table
compared to ISO 646 IRV code table

	0	1	2	3	4	5	6	7
0			SP	0	É	P	é	p
1			!	1	A	Q	a	q
2			"	2	B	R	b	r
3			#	3	C	S	c	s
4			¤	4	D	T	d	t
5			%	5	E	U	e	u
6			&	6	F	V	f	v
7			'	7	G	W	g	w
8			<	8	H	X	h	x
9			)	9	I	Y	i	y
A			*	:	J	Z	j	z
B			+	;	K	Ä	k	ä
C			,	<	L	Ö	l	ö
D			-	=	M	Å	m	å
E			.	>	N	Ü	n	ü
F			/	?	O	_	o	

Characters in unshaded positions exist in both sets with the same coding.
Characters in shaded positions do not exist in ISO 646 IRV.

Latin-1 code table
compared to Latin-2 code table

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0			SP	0	@	P	`	p			NBSP	•	À	Ð	à	ø
1			!	1	A	Q	a	q			ı	±	Á	Ñ	á	ñ
2			"	2	B	R	b	r			€	²	Â	Ò	â	ò
3			#	3	C	S	c	s			£	³	Ã	Ó	ã	ó
4			\$	4	D	T	d	t			¤	´	Ä	Ô	ä	ô
5			%	5	E	U	e	u			¥	µ	Å	Õ	å	õ
6			&	6	F	V	f	v			¦	¶	Æ	Ö	æ	ö
7			'	7	G	W	g	w			§	-	Ç	×	ç	÷
8			<	8	H	X	h	x			¨	,	È	Ø	è	ø
9			)	9	I	Y	i	y			©	!	É	Ù	é	ù
A			*	:	J	Z	j	z			¨	*	Ê	Ú	ê	ú
B			+	;	K	Ç	k	ç			«	»	Ë	Û	ë	û
C			,	<	L	\	l				¬	¼	Ì	Ü	ì	ü
D			-	=	M	]	m	}			SHY	½	Í	Ý	í	ý
E			.	>	N	^	n	~			®	¾	Î	Þ	î	þ
F			/	?	O	_	o				¯	¿	Ï	ß	ï	ÿ

Characters in unshaded positions exist in both sets with the same coding.
Characters in shaded positions do not exist in Latin-2.

NOTE – The Latin-1 character LATIN CAPITAL LETTER ETH assigned to code D0 does not exist in Latin-2. However, the Latin-2 character LATIN CAPITAL LETTER D WITH STROKE assigned to the same code is generally represented by a similar glyph image. For display or printer output purposes these two characters could therefore be considered identical.

Behov i Sverige i praktiken 1

- **Svenska språket**
- **Samiska m fl indogena språk**
- **Invandrande språk: latinska, icke-latinska**
- **EUs officiella språk**
- **Östeuropeiska språk**
- **Egennamn, hela språk och specialtecken?**

Behov i Sverige i praktiken 2

- **Skriva namn så rätt som möjligt:
Folkbokföringen, transskribera till
vad? Fall-back, Schengen, icke-
latinska?**
- **Skriva namn enligt
Källan
Folkbokföringen
Engelska alfabetet**
- **Teckenuppsättningen är grundläggande**

Att beakta

- **Förståelsen** – mänskliga hanteringar
- **Inmatningen** – tagentbordets begränsningar *Design for all*
- **Arvet** – befintliga databaser och utrustningar
- **Translitterering/transskribering** – till vad
- **Krav i världen** – EU, IDA, pass, Internet
- **Namnlagen** *Kulturella särdrag*
- **Referensregistret** RSV – folkbokföringen

Svenska tangentbordet –

SIS 66 22 41:2 och framtiden

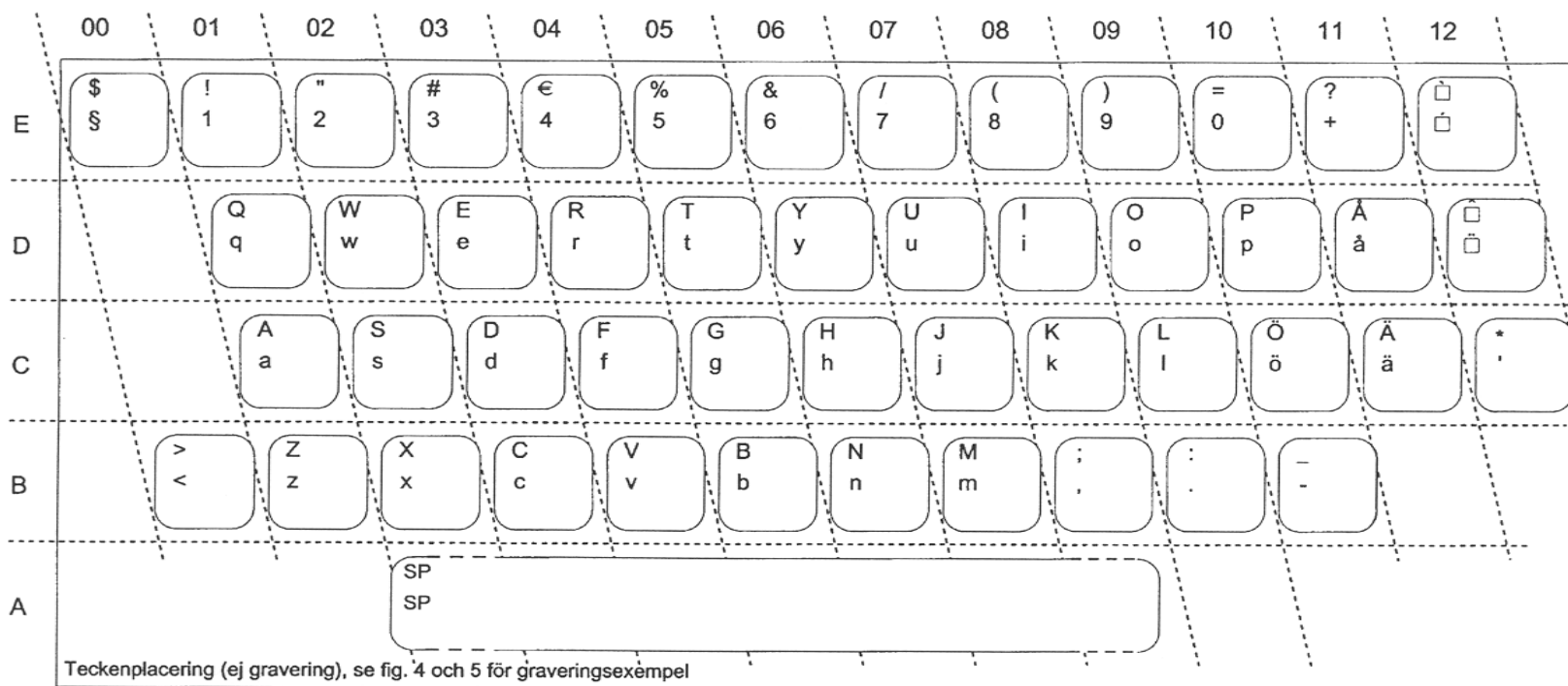
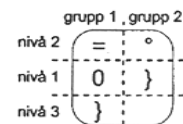
- **Modularisera tangentborden
inkl tre nivåer per grupp:**
 - Grupp 1 (enligt sv standard)**
 - Grupp 2 (enligt ISO/IEC och sv standard)**
 - Grupp 3 (en Sami-variant?)**
 - Grupp 4 (en annan variant?)**
 - Grupp 5**
- **Sv gruppregister? Sv växlingsregler?**

SS 66 22 41 Grupp 1 nivå 1 och 2

Tangenternas storlek och placering i figuren är bara ungefärliga

□ indikerar kombinativt tecken (dödtangent)

SP indikerar blanktecken



Figur 1 – Tangentområde för grafiska tangenter; placering av tecken i svensk grupp 1, nivå 1 och 2.

ANM. 1 – Dessa tangenter – 48 normalstora tangenter + tangenten för blanktecken – utgör det minsta tillåtna område på ett alfanumeriskt tangentbord som upptas av grafiska tangenter enligt denna standard.

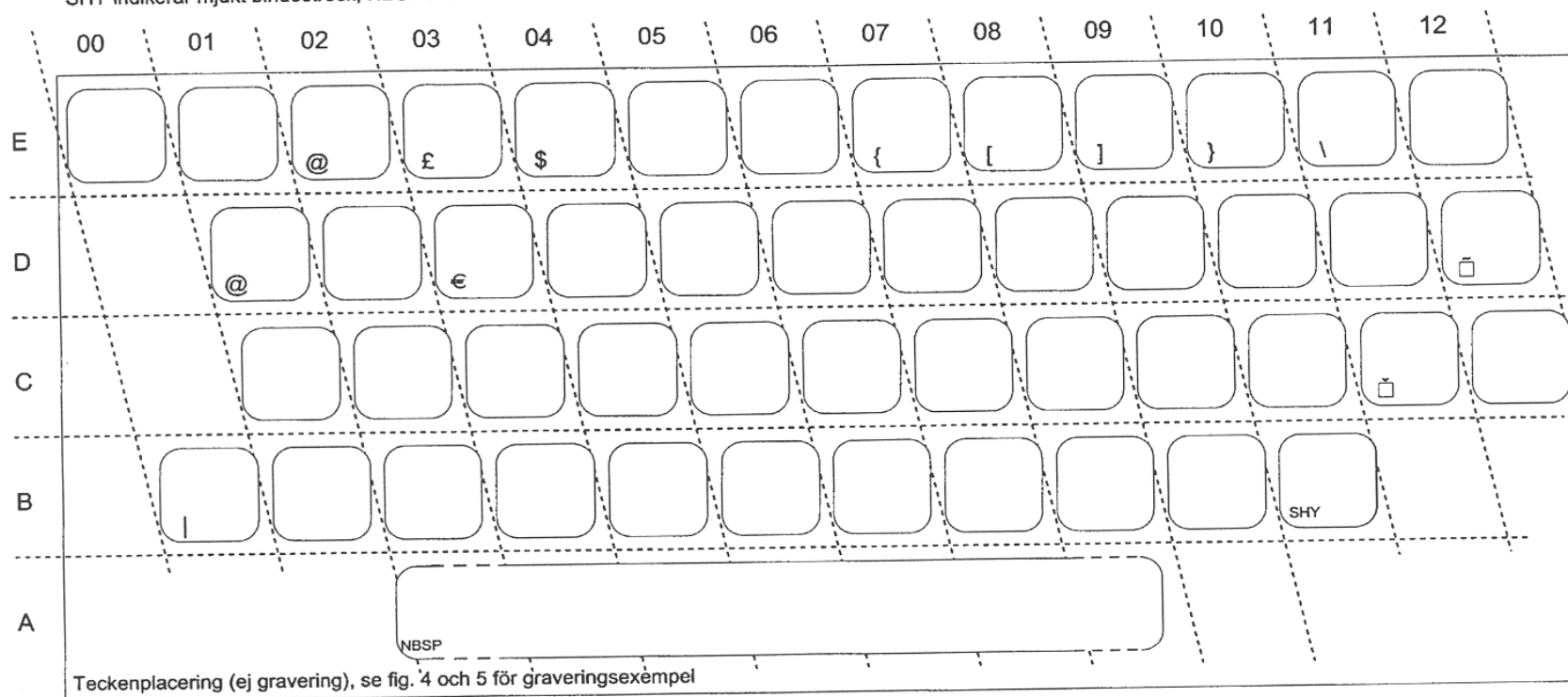
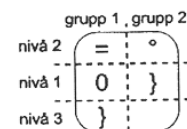
ANM. 2 – Tangentbord graveras normalt inte på detta sätt – se avsnitt 5.3.

SS 66 22 41 Grupp 1 nivå 3

Tangenternas storlek och placering i figuren är bara ungefärliga

□ indikerar kombinativt tecken (dödtangent)

SHY indikerar mjukt bindestreck, NBSP indikerar hårt blanktecken



Figur 2 – Placering av tecken i svensk grupp 1 nivå 3.

ANM. 1 – Tecknen i positionerna E02, E03, E04, D03, C11, B10 samt rad A är inte obligatoriska för nivå 3 – se avsnitt 5.2.2.

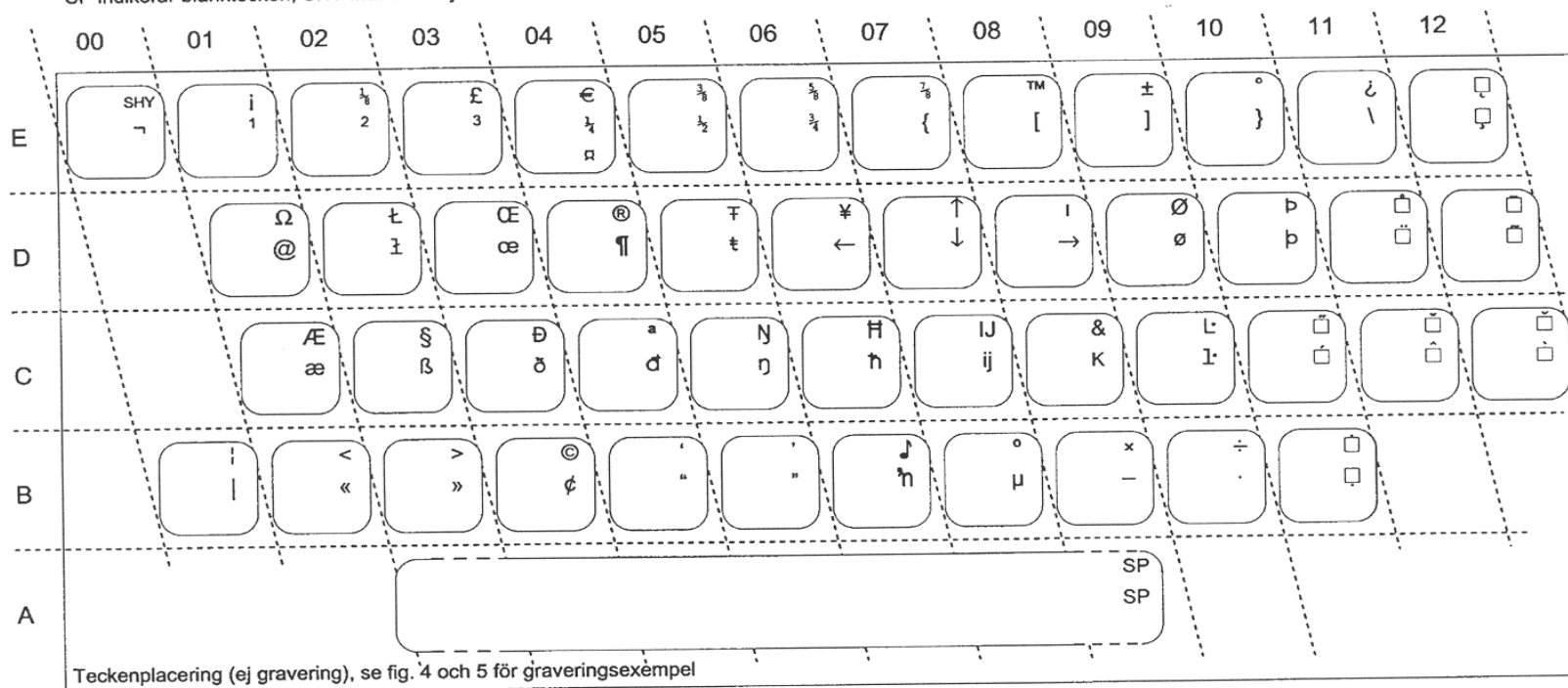
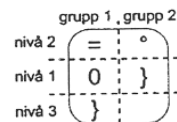
ANM. 2 – Skuggade teckenpositioner indikerar positioner som aldrig skall graveras; gäller E02, E04 och D03.

SS 66 22 41 Grupp 2

Tangenternas storlek och placering i figuren är bara ungefärliga

□ indikerar kombinativt tecken (dödtangent)

SP indikerar blanktecken, SHY indikerar mjukt bindestreck



Figur 3 – Placering av tecken i internationell grupp 2 i enlighet med ISO/IEC 9995-3, avsnitt 5, med preciseringar enligt denna svenska standard.
ANM. – Alla dessa tecken graveras normalt inte – se avsnitt 6.2.

Framtidens teckenuppsättning i svenska statsförvaltningen

- **Full teckenuppsättning enligt ISO/IEC 10646**
 - **Specialbruk**
- **Samtliga 191 bokstäver och andra tecken enligt 8859/1 - Normalbruk**
- **Internationell/engelsk ISO/IEC 646 IRV**
 - **Specialbruk**
- **Mellansteg MES 2 plus 6 kyrilliska tecken dvs sammanlagt 1058 tecken som kan fallbackas till IRV enligt CEN TR 14381:2003**
- **Kodning vanligen enligt ISO/IEC 10646**

Vad göra för detta?

- Vad göra i Sverige nu?
- Nämnden för elektronisk förvaltning inrättas f n - bl a för att fastställa bindande föreskrifter
- Föreslå **en föreskrift** för den statliga sektorns teckenuppsättning
- Komplettera med inmatningsregler
- Leverantörernas roll?
- Utarbета-remissbehandla-fastställa