

	SCG Guidance: How to convert Diacritic Characters within Patient Names			
	Programme	NPFIT	Document Record ID Key	
	Sub-Prog / Project	Standards Consultancy Group	NPFIT-FNT-TO-0067.07	
	Prog. Director	Ken Lunn	Status	Approved
	Owner	Keith Naylor	Version	1.0
	Author	Bharat Sharma	Version Date	21/10/2009

SCG Guidance: How to convert Diacritic Characters within Patient Names

Amendment History:

Version	Date	Amendment History
0.1	01/05/2009	Initial Creation
0.2	01/07/2009	Changes based on Comments
0.3	26/08/2009	Changes based on wider review Comments – Appendices and Clarification of Scope
0.4	03/09/2009	Added reference to European Alphabets for comparison to Appendices and ISO/IEC 10646
0.5	11/09/2009	Fifth draft including final correction to umlaut transliterations prior to submission to IQAP editorial board.
0.6	16/10/2009	Sixth draft prior to submission to external review and approval.
1.0	21/10/2009	Approved version

Forecast Changes:

Anticipated Change	When
Subject to Annual Review	01/04/2010

Reviewers:

This document must be reviewed by the following:

Name	Signature	Title / Responsibility	Date	Version
Davie Hay	Davie Hay	SCG and IG SMEs Lead	04/09/09	0.4
Rae Long	Rae Long	Standards Consultancy Lead (Acting)	04/09/09	0.4
Keith Naylor	Keith Naylor	Head of Data Standards and Products Implementation	04/09/09	0.4
Tim Chearman	Tim Chearman	Head of User Interface Design	04/09/09	0.4
Joanne Simpson	Joanne Simpson	PDS Data Manager	04/09/09	0.4
Philip Brennan	Philip Brennan	Technical Architect	04/09/09	0.4
John Martin	John Martin	Data Quality Manager	04/09/09	0.4
Jeremy Rogers	Jeremy Rogers	Principal Terminology Specialist	04/09/09	0.4
Edward Cheetham	Edward Cheetham	Principal Clinical Terminology Specialist	04/09/09	0.4
IQAP Editorial Board	IQAP Editorial Board	IQAP Editorial Board	12/10/2009	0.5

Approvals:

This document must be approved by the following:

Name	Signature	Title / Responsibility	Date	Version
Keith Naylor	Keith Naylor	Head of Data Standards and Products Implementation	20/10/2009	0.6

Distribution:

SHA Data Quality Leads

NHS Information Managers

NHS Data Quality Managers

NHS Medical Records Managers

NHS Trainers and Training Managers

Registration Staff in Primary and Secondary Care (PCT's, Acute, GP's)

LSPs and ESPs

Document Status:

This is a controlled document.

Whilst this document may be printed, the electronic version maintained in FileCM is the controlled copy. Any printed copies of the document are not controlled.

Related Documents:

These documents will provide additional information.

Ref no	Doc Reference Number	Title	Version
1	NPFIT-SHR-QMS-PRP-0015	Glossary of Terms Consolidated	13
2	NPFIT-FNT-TO-DQM-0024	IQAP Guidance on Ethnic Naming Conventions	5.0
3	http://ec.europa.eu/publications/bo- oklets/move/45/en.pdf	European Commission (2004) "Many tongues, one family. Languages in the European Union"	
4	http://standards.iso.org/ittf/Publicly AvailableStandards/c039921_ISO IEC_10646_2003(E).zip	ISO/IEC 10646:2003 - Information technology — Universal Multiple-Octet Coded Character Set (UCS). ISO_2003_charts_0000_33FF_BasicLatin_CJKCompat.pdf	

Glossary of Terms:

Term	Acronym	Definition
Information Quality Assurance Programme	IQAP	The Information Quality Assurance Programme (IQAP) was established by NHS National Programme for IT to ensure that guidance documents are issued to advise the NHS and Local Service Providers of the data quality related standards necessary for the NHS Care Records Service (NHS CRS).
International Organization for Standardization	ISO	ISO (International Organization for Standardization) is the world's largest developer and publisher of International Standards.
Transliteration		Transliteration is the practice of converting a text from one writing system into another in a systematic way.
Romanization		Rendering of written text from a foreign writing system into the Latin (Roman) alphabet, possibly supplemented by diacritical marks or additional characters.

Contents

1	Purpose	5
1.1	Out of Scope	5
1.2	Further Enquiries	5
2	Audience	5
3	Background	5
3.1	Existing Guidance.....	6
4	Transliteration of Diacritic Characters.....	7
	Appendix 1 – Additional Reference Materials	8
	Appendix 2 – Transliteration of Latin Diacritic Characters into English alphabet characters (A-Z) format	9
	Appendix 3 – Transliteration of Greek Diacritic Characters into English alphabet characters (A-Z) format	13
	Appendix 4 – Transliteration of Cyrillic Diacritic Characters into English alphabet characters (A-Z) format	15

1 Purpose

This document has been developed as guidance on how to convert diacritic characters that exist in a patient's name.

This issue was raised by the IQAP editorial board and was subsequently taken under the remit of the Standards Consulting Group (SCG). This was done with the understanding that there are no current NHS standards or guidance around the recording of a patient's name if it contains a diacritic character.

The scope of this document will focus on diacritic characters that exist in names from all official European Union (EU) writing scripts and at minimum all modern diacritic characters used in the alphabets of the 23 official EU languages.

1.1 Out of Scope

This document does not cover obsolete diacritic characters and does not intend to cover all characters used in the alphabets of European languages not officially recognised by the European Union (EU).

This document also does not cover non-EU writing scripts such as Mandarin or Hindi and consideration for these will be made in future versions of this guidance.

1.2 Further Enquiries

Enquiries about the contents of this document, or any of the requirements within it should be sent to nhsafh.scg@nhs.net.

The Standards Consulting Group (SCG) information is posted at <http://www.connectingforhealth.nhs.uk/systemsandservices/data/scg>.

2 Audience

The intended audience for this guidance is NHS Trainers and all NHS staff involved in registering, accessing and updating patient information including any other NHS employee who directly interface with patients.

3 Background

The NHS treats patient's from a diverse variety of backgrounds, cultures and origins from the UK and the rest of the world.

Diacritic characters are the result of marks utilised in alphabets that alter the existing character set to produce different pronunciations.

Example: é

It has been identified that there is a lack of consistent guidance around the recording of patient names that contain diacritic characters. This means that a patient may enter a healthcare organisation and have their name stored in a variety of different

ways, due to the misinterpretation or inability to enter the diacritic character into a keyboard or application using the English 'A-Z' alphabet.

Investigation has found not all NHS systems allow the storage and communication of diacritic characters.

This document aims to provide guidance on how to convert these characters until such a time that all NHS systems allow them.

3.1 Existing Guidance

- International standards currently exist in the form of ISOs that allow for and provide the standardised transliteration of writing scripts:
 - Latin to English (derived) - ISO 8859:1998 (Parts 1-10)¹
 - ISO/IEC 10646:2003 (Basic Latin)²
 - Cyrillic to Latin - ISO 9:1995³
 - Greek to Latin - ISO 843:1997⁴
- The United Nations (UN), in research into their own technical standard known as LOCODE, produced a Secretariat note⁵ containing transliteration of Latin diacritic characters into those used to write the English 'A-Z' alphabet. This documentation is based on ISO Standards.
- The International Civil Aviation Organisation (ICAO) has guidance that contains the transliteration of Latin and Cyrillic characters into those used to write the English 'A-Z' alphabet. Doc 9303⁶ "Machine Readable Official Travel Documents" contains recommended transliterations of national characters. This documentation is based on ISO standards.

¹ http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=28245

² http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=39921

³ http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=3589

⁴ http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=5215

⁵ <http://www.unece.org/cefact/locode>

⁶ <http://www2.icao.int/en/MRTD/Pages/Doc9393.aspx>

4 Transliteration of Diacritic Characters

In the event a patient name is provided with diacritical characters, transliteration can be used to convert diacritic characters into an English 'A-Z' alphabet form.

In the EU there are three official writing scripts that are currently recognised and can be used to write alphabets of the 23 recognised official languages⁷. Diacritical characters in these writing scripts have been mapped to the Latin characters used in writing the English 'A-Z' alphabet. These can be found in Appendices 2, 3 and 4.

ISO Standards (Section 3.1) were used as the standards to perform this transliteration and have been verified against the characters used in the alphabets⁸ below.

- Latin (Appendix 2) is used to write the alphabets of the following languages:
 - Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Hungarian, Irish, Italian, Latvian, Lithuanian, Maltese, Polish, Portuguese, Romanian, Slovak, Slovenian, Spanish and Swedish.
- Greek (Appendix 3) is used to write the alphabet of:
 - Greek
- Cyrillic (Appendix 4) is used to write the alphabet of:
 - Bulgarian

⁷ http://europa.eu/abc/european_countries/languages/index_en.htm

⁸ <http://www.evertype.com/alphabets/>

Appendix 1 – Additional Reference Materials

NHS CFH does not endorse or hold responsibility for the content of external internet sites.

United Nations Code for Trade and Transport Locations	http://www.unece.org/cefact/locode
Unicode Code charts – scripts (character encoding used in the representing of diacritic characters in this document)	http://www.unicode.org/charts/index.html
Wikipedia: Romanization of Greek – Table	http://en.wikipedia.org/wiki/Romanization_of_Greek
ISO 9 – Latin to Cyrillic ISO 9:1995	http://en.wikipedia.org/wiki/ISO_9
International Civil Aviation Organisation (ICAO): Doc 9303, Machine Readable Travel Documents	http://www2.icao.int/en/MRTD/Pages/Doc9393.aspx
International Standard Organization: ISO/IEC 8859-1:1998 Information technology -- 8-bit single-byte coded graphic character sets -- Part 1: Latin alphabet No. 1	http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=28245
International Standard Organization: ISO 9:1995 Information and documentation -- Transliteration of Cyrillic characters into Latin characters -- Slavic and non-Slavic languages	http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=3589
International Standard Organization: ISO 843:1997 Information and documentation -- Conversion of Greek characters into Latin characters	http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=5215
Patient Name Input and Display User Interface Design Guidance	http://www.cuiseecure.nhs.uk/CAPS/Information%20Entry%20and%20Display/Patient%20Name%20Input%20and%20Display.pdf
EUROPA - The EU at a glance - Languages in the EU	http://europa.eu/abc/european_countries/languages/index_en.htm
Evertype : The alphabets of Europe – Section 1.2.2	http://www.evertype.com/alphabets/

Appendix 2 – Transliteration of Latin Diacritic Characters into English alphabet characters (A-Z) format

NHS CFH does not endorse or hold responsibility for the content of external internet sites. Characters may differ slightly in appearance from how they are hand written due to the use of Unicode characters in this document.

This table was generated from transliteration of characters in the Latin writing system to characters used in the English alphabet, based upon the full descriptions of characters provided in ISO 8859:1998 (Parts 1-10) and ISO/IEC 10646:2003 (Basic Latin). Some exceptions to use of these descriptions do exist in order to respect historical and cultural conventions.

Upper Case Diacritic Character	Lower Case Diacritic Character	Description	Upper Case Transliterates to:	Lower Case Transliterates to:
Á	á	A acute	A	a
À	à	A grave	A	a
Â	â	A circumflex	A	a
Ä	ä	A diaeresis or umlaut	AE	ae
Ã	ã	A tilde	A	a
Ǻ	ǻ	A breve	A	a
Å	å	A ring	AA	aa
Ą	ą	A ring with acute	AA	aa
Ā	ā	A macron	A	a
Ȧ	ȧ	A ogonek	A	a
Ć	ć	C acute	C	c
Ĉ	ĉ	C circumflex	C	c
Č	č	C caron	C	c
Ċ	ċ	C dot accent	C	c
Ç	ç	C cedilla	C	c
Ð	ð	Eth	D	d
Ď	ď	D caron	D	d
É	é	E acute	E	e
È	è	E grave	E	e
Ê	ê	E circumflex	E	e
Ë	ë	E diaeresis or umlaut	E	e
Ě	ě	E caron	E	e

Upper Case Diacritic Character	Lower Case Diacritic Character	Description	Upper Case Transliterates to:	Lower Case Transliterates to:
É	é	E dot accent	E	e
Ē	ē	E macron	E	e
Ę	ę	E ogonek	E	e
Ě	ě	E breve	E	e
Ĝ	ĝ	G circumflex	G	g
Ğ	ğ	G breve	G	g
Ġ	ġ	G dot accent	G	g
Ģ	ģ	G cedilla	G	g
Ĥ	ĥ	H bar	H	h
Ĥ	ĥ	H circumflex	H	h
İ	ı	I without dot (Turkey)	I	i
Í	í	I acute	I	i
Ì	ì	I grave	I	i
Î	î	I circumflex	I	i
Ï	ï	I diaeresis or umlaut	I	i
Ĩ	ĩ	I tilde	I	i
Ĭ	ĭ	I dot accent	I	i
Ī	ī	I macron	I	i
Ĳ	ĳ	I ogonek	I	i
Ĳ	ĳ	I breve	I	i
Ĵ	ĵ	J circumflex	J	j
Ķ	ķ	K cedilla	K	k
Ł	ł	L slash	L	l
Ł	ł	L acute	L	l
Ľ	ľ	L caron	L	l
Ł	ł	L cedilla	L	l
Ń	ń	N acute	N	n
Ñ	ñ	N tilde	N	n
Ñ	ň	N caron	N	n
Ŋ	ŋ	N cedilla	N	n
Ŋ	ŋ	Eng	N	n

Upper Case Diacritic Character	Lower Case Diacritic Character	Description	Upper Case Transliterates to:	Lower Case Transliterates to:
Ø	ø	O slash	OE	oe
ø	ø	O slash with acute	OE	oe
Ó	ó	O acute	O	o
Ò	ò	O grave	O	o
Ô	ô	O circumflex	O	o
Ö	ö	O diaeresis or umlaut	OE	oe
Õ	õ	O tilde	O	o
Ő	ő	O double acute	O	o
Ō	ō	O macron	O	o
Ȫ	ȫ	O breve	O	o
Ŕ	ŕ	R acute	R	r
Ř	ř	R caron	R	r
Ŗ	ŗ	R cedilla	R	r
Ś	ś	S acute	S	s
Ŝ	ŝ	S circumflex	S	s
Š	š	S caron	S	s
Ş	ş	S cedilla	S	s
Ʀ	Ʀ	T bar	T	t
Ť	ť	T caron	T	t
Ṭ	ṭ	T cedilla	T	t
Ú	ú	U acute	U	u
Ù	ù	U grave	U	u
Û	û	U circumflex	U	u
Ü	ü	U diaeresis or umlaut	UE	ue
Ũ	ũ	U tilde	U	u
Ǔ	ǔ	U breve	U	u
Ǘ	ǘ	U double acute	U	u
Ů	ů	U ring	U	u
Ū	ū	U macron	U	u
Ț	ț	U ogonek	U	u
Ŵ	ŵ	W circumflex	W	w

Upper Case Diacritic Character	Lower Case Diacritic Character	Description	Upper Case Transliterates to:	Lower Case Transliterates to:
Ý	ý	Y acute	Y	y
ÿ	ÿ	Y circumflex	Y	y
ÿ	ÿ	Y diaeresis or umlaut	Y	y
Ž	ž	Z acute	Z	z
Ž	ž	Z caron	Z	z
Ž	ž	Z dot	Z	z
Þ	þ	Thorn (Iceland)	TH	th
Æ	æ	AE ligature	AE	ae
ij	ij	IJ ligature	IJ	ij
Œ	œ	OE ligature	OE	oe
ß	ß	S double (Germany)	SS	ss

Appendix 3 – Transliteration of Greek Diacritic Characters into English alphabet characters (A-Z) format

NHS CFH does not endorse or hold responsibility for the content of external internet sites. Characters may differ slightly in appearance from how they are hand written due to the use of Unicode characters in this document.

This table was generated from transliteration of Greek to Latin, based upon ISO 843:1997 as derived from Wikipedia: Romanization of Greek. Latin was then transliterated to English based upon ISO 8859:1998 (Parts 1-10) and ISO/IEC 10646:2003 (Basic Latin) as derived from International Civil Aviation Organisation (ICAO): Doc 9303 Machine Readable Travel Documents.

Upper Case Diacritic Character	Lower Case Diacritic Character	Description	Upper Case Transliterates to:	Lower Case Transliterates to:
A	α	Alpha	A	a
Ά	ά	Alpha tonos	A	a
B	β	Beta	V	v
Γ	γ	Gamma	G	g
Δ	δ	Delta	D	d
E	ε	Epsilon	E	e
Έ	έ	Epsilon tonos	E	e
Z	ζ	Zeta	Z	z
H	η	Eta (i macron)	I	i
Ή	ή	Eta tonos	I	i
Θ	θ	Theta	TH	th
I	ι	Iota	I	i
Ϊ	ί	Iota tonos	I	i
N/A	ϊ	Iota dialytika and tonos	N/A	i
K	κ	Kappa	K	k
Λ	λ	Lambda	L	l
M	μ	Mu	M	m
N	ν	Nu	N	n
Ξ	ξ	Xi	X	x
O	ο	Omicron	O	o
Ό	ό	Omicron tonos	O	o
Π	π	Pi	P	p

Upper Case Diacritic Character	Lower Case Diacritic Character	Description	Upper Case Transliterates to:	Lower Case Transliterates to:
P	ρ	Rho	R	r
Σ	σ	Sigma	S	s
T	τ	Tau	T	t
Υ	υ	Upsilon	Y	y
ΰ	ύ	Upsilon tonos	Y	y
ÿ	ÿ	Upsilon diaeresis or umlaut	Y	y
N/A	ϋ	Upsilon dialytika and tonos	N/A	y
Φ	φ	Phi	F	f
X	χ	Chi	CH	ch
Ψ	ψ	Psi	PS	ps
Ω	ω	Omega (o macron)	O	o
Ω	ώ	Omega tonos (o macron)	O	o

Appendix 4 – Transliteration of Cyrillic Diacritic Characters into English alphabet characters (A-Z) format

NHS CFH does not endorse or hold responsibility for the content of external internet sites. Characters may differ slightly in appearance from how they are hand written due to the use of Unicode characters in this document.

This table was generated from transliteration of Cyrillic to Latin, based upon ISO 9:1995 as found in International Civil Aviation Organisation (ICAO): Doc 9303 Machine Readable Travel Documents and Wikipedia: ISO 9. Latin was then transliterated to English based upon ISO 8859:1998 (Parts 1-10) and ISO/IEC 10646:2003 (Basic Latin) as derived from International Civil Aviation Organisation (ICAO): Doc 9303 Machine Readable Travel Documents.

Upper Case Diacritic Character	Lower Case Diacritic Character	Description	Upper Case Transliterates to:	Lower Case Transliterates to:
А	а	A	A	a
Б	б	Be	B	b
В	в	Ve	V	v
Г	г	Ge	G (except Belorussian and Serbian-Macedonian = H)	g (except Belorussian and Serbian-Macedonian = h)
Ґ	ѓ	Ge upturn (g grave)	G	g
Д	д	De	D	d
Ђ	ђ	Dje (d macron)	D	d
Ѓ	ѓ	Gje (g acute)	G	g
Е	е	Ye	E	e
Ё	ё	Yo (e diaeresis or umlaut)	E (except Belorussian = IO)	e (except Belorussian = io)
Є	є	Ye (e circumflex)	IE	ie
Ж	ж	Zhe (z caron)	ZH (except Serbian-Macedonian = Z)	zh (except Serbian-Macedonian = z)
Ж̣	ж̣	Zhe (z caron and cedilla)	Z	z
З	з	Ze	Z	z
С	с	Dze	DZ	dz

Upper Case Diacritic Character	Lower Case Diacritic Character	Description	Upper Case Transliterates to:	Lower Case Transliterates to:
И	и	I	I (except Ukrainian =Y)	i (except Ukrainian =y)
Ї	ї	I (i acute)	I	i
İ	ï	Yi (i diaeresis)	I	i
Й	й	Short I	I	i
Ј	ј	Je	J	j
К	к	Ka	K	k
Л	л	El	L	l
Л'	л'	El (l acute)	L	l
Љ	љ	Lje (l circumflex)	LJ	lj
М	м	Em	M	m
Н	н	En	N	n
Њ	њ	Nje (n circumflex)	NJ	nj
О	о	O	O	o
Œ	œ	OE ligature	OE	oe
П	п	Pe	P	p
Р	р	Er	R	r
С	с	Es	S	s
Т	т	Te	T	t
Ћ	ћ	Tshe (c acute)	C	c
Ќ	ќ	Kje (k acute)	K	k
У	у	U	U	u
Ў	ў	Short U (u breve)	U	u
Ф	ф	Ef	F	f
Х	х	Kha	KH (except Serbian-Macedonian = H)	kh (except Serbian-Macedonian = h)
Ц	ц	Tse	TS (except Serbian-Macedonian = C)	ts (except Serbian-Macedonian = c)
Ч	ч	Che (c caron)	CH (except Serbian-Macedonian = C)	ch (except Serbian-Macedonian = c)

Upper Case Diacritic Character	Lower Case Diacritic Character	Description	Upper Case Transliterates to:	Lower Case Transliterates to:
Č	č	Che (c circumflex)	C	c
Ĉ	ĉ	Che (c grave)	C	c
Ċ	ċ	Dzhe (d circumflex)	DZ	dz
Š	š	Sha (s caron)	SH (except Serbian-Macedonian = S)	sh (except Serbian-Macedonian = s)
Щ	щ	Shcha (s circumflex)	SHCH (except Bulgarian = SHT)	shch (except Bulgarian = sht)
Ъ	ъ	Hard sign (Yer)	Y	y
Ы	ы	Yery	Y	y
Ь	ь	Soft sign (Yeri)	Y	y
Э	э	E (e grave)	E	e
Ю	ю	Yu (u circumflex)	IU	iu
Я	я	Ya (a circumflex)	IA	ia
Ѹ	ѹ	izhitsa	Y	y
Ң	ң	Ng (n cedilla)	N	n
Ө	ө	Barred O (o circumflex)	O	o
Ү	ү	Straight U with stroke (u dot)	U	u