Exploring Complexity in Health: An Interdisciplinary Systems Approach
A. Hoerbst et al. (Eds.)

How to Work with Vocabulary Correctly, Exemplified with Gender Coding?

Frank OEMIG[a,b,c,1], Bernd BLOBEL[c,d] and Peter GEIBEL[b]
[a] *Deutsche Telekom Healthcare and Security Solutions GmbH, Mülheim, Germany*
[b] *German Association of Health IT Vendors – bvitg e.V., Berlin ,Germany*
[c] *eHealth Competence Center Bavaria, Deggendorf Institute of Technology, Germany*
[d] *Medical Faculty, University of Regensburg, Germany*

Abstract. Popular data exchange standards facilitate the use of vocabulary in different ways. Quite often, the usage is based on, and in favor of, simple implementations not considering reality. In most cases the vocabulary is assembled in form of simple code lists. Well-known bad practice is neither to define proper concept descriptions for the individual codes leaving their interpretation to the developer nor to identify the underlying vocabulary domain resulting in a mixture of something. This paper takes the gender use case and analyses the use of codes within some data exchange standards and makes recommendations for improvements in handling and managing code systems correctly.

Keywords. Interoperability; Vocabulary; Code system

1. Introduction

The new e-health law in Germany intends to improve semantic interoperability among healthcare information systems [1]. But reality demonstrates that each individual organization claims their right to specify data exchange standards on their own, thus not considering any interoperability requirements. This ignorance starts at the syntactic level, continues at the semantic level, and finally has no chance to ever reach service level. Data exchange specifications do not only touch those different levels, but also include vocabulary fundamentally. Quite common is the specification of "tables" without further meta-data or versioning information and no management procedure in place [2].

This paper analyses common practice and suggests an alternative to this behavior exemplified with coding of gender aspects.

2. Methods

International data exchange standards (CDC, NETSS, NAACCR, USA Census, HL7 v2, HL7 V3, X12, ASTM, DICOM, OBIS, UBIF) are examined for their use of codes in a gender context. Table 1 provides an overview about standards and their

[1] Corresponding Author: Frank Oemig, PhD; Deutsche Telekom Healthcare and Security Solutions GmbH, Germany; Email: frank.oemig@t-systems.com

incorporation of certain concepts (covering human beings and animals). The only commonality for all of them is the support of "male" and "female", all other concepts are diverging and orthogonal in their semantics. However, all standards intend to use a single code from its list.

Table 1. Gender Coding with International Data Exchange Standards

Standard / Description	ISO/IEC 5218 [3]	ISO/IEC 5218 extended	CDC	NETSS, NCVHS [6]	ECML [8]	ICAO	Visa-Schengen	NAACCR	US Census	HL7 V3	HL7 v2	X12	ASTM	DICOM [13]	OBIS [16]	UBIF [17]
not known	0	0		9	U	X	<	9		U	U	U		U	U	?
Male	1	1	M	1	M	M	M	1	1	M	M	M	M	M	M	m
Female	2	2	F	2	F	F	F	2	2	F	F	F	F	F	F	f
both (male + female)															B	
transgender male		3											FC	FC		fc
transgender female		4											MC	MC		mc
Transitional															T	
not applicable	9										N					
Other			O					3			O			121102		
Transsexual								4								
non-sexed												N				
Unsexable												X		121103	I	i
unknown sex													U			
male pseudo-hermaphrodite													MP	MP		
female pseudo-hermaphrodite													FP	FP		
Hermaprodite													H	H	H	h
simultaneous hermaphrodite																sh
hermaphrodite male phase																hm
hermaphrodite female phase																hf
Ambiguous											A		A	121102		a
mixed sex																x

3. Results

This short summary clearly indicates that some standards mix different kind of concepts into a single value set. Other standards also introduce the notion of time as can be seen with codes for a "transition" process.

Common practice is to use only a single code. Most – if not all – applications only allow for a single value to be stored with a patient. Normally, i.e. in 99.9% of the solutions, an individual is associated with a single gender that persists over the whole life. Hence, there is no possibility to capture and maintain specifics like transsexual or transgender aspects. This is a rare exception. In some jurisdictional environments – like in Germany - keeping track of this kind of information is even illegal. Nevertheless, an appropriate representation is necessary and a correct handling is an essential precondition to manage future and upcoming challenges in vocabulary usage.

Table 2.a and 2.b demonstrate that the administrative gender, which is, e.g., used in a hospital for assigning beds, is more or less identical to what a person or natal register will maintain. The only distinction is that a register allows for not stating an explicit gender. In Table 1 this is listed as "non-sexed". Furthermore, it is worth mentioning, that administrative gender is used for simple administrative reasons within hospitals and has nothing to do with sexual orientation, so that a reduction should not introduce any problems beyond possible discrimination perception of the person in question.

Note that the term "gender" pertains to gender roles and gender identity, whereas "sex" refers to the biological sex of a person, which comprises chromosomal, gonodal, genital and hormonal sex. It is common practice that in clinical studies no clear distinction is made between these concepts, potentially introducing a statistical bias since transsexual and intersexual persons are mixed with "biological" men in women.

Table 2. a) Administrative Gender, b) Person Register

a)

Code	Description
M	Male
F	Female
U	unknown

=

b)

Code	Description
M	Male
F	Female
X	Undetermined

Tables 3.a, 3.b and 3.c address different clinical aspects. Transgender issues normally only require the distinction between transgender (no assignment/definition of a gender or sexual orientation) and non-transgender. In principle, a Boolean value would be sufficient then. But for non-transgender persons, a gender assignment may become necessary again. Therefore, a level is introduced indicating, that male and female are specializations of non-transgender persons.

Table 3. a) intersexual, b) transsexual, c) transgender

a)

Code	Description
M	Male
W	Female

b)

Code	Description
M	Identifies as a man
F	Identifies as a woman

c)

Code	Level	Description
T	1	transgender
NT	1	Non-transgender

a)

Code	Description
IS	intersexual

b)

Code	Description
U	Unknown

c)

Code	Level	Description
M	2	Male
F	2	female

Discussing necessary and appropriate codes usually ends up with introducing a new code for "other" not specifying the real underlying meaning because it is unknown. With the introduction of specialized code systems, this behavior is misleading and not necessary any more.

For some clinical, medical and biological questions, the genetic gender becomes necessary. For such a purpose, Table 4 suggests values for an appropriate code system.

Table 4. Genetics

Concept	Code	Description
46,XX	46_XX	Caryotype (female)
46,XY	46_XY	caryotype (male)
45,X	45_X	Turner-Syndrome with female phenotype
47,XXY	47_XXY	Klinefelter-Syndrome with male phenotype
48,XXXY	48_XXXY	Klinefelter-Syndrome with male phenotype (rare)
49,XXXXY	49_XXXXY	Klinefelter-Syndrome with male phenotype (rare)
47,XXX	47_XXX	Triplo-X-Syndrome
mos45,X/46,XX	mos45_X46_XX	Mosaic
mos45,X/46,XY	mos45_X46_XY	Mosaic
chi46,XX/46,XY	chi46_XX46_XY	Chimersm
47,21	47_21	Down-Syndrome (Trisomy 21)
47,18	47_18	Edwards-Syndrome (Trisomy 18)
47,13	47_13	Pätau-Syndrome (Triosomy 13)
47,8	47_8	Trisomy 8
48,XXXX	48_XXXX	
49,XXXXX	49_XXXXX	
47,XYY	47_XYY	XYY-Syndrome
48,XXYY	48_XXYY	Y-Polysomy
49,XYYYY	49_XYYYY	Y-Polysomy

Table 4 clearly indicates the necessity for clean code systems with a solid description of the concepts. Each table (2.a to 4) represents a code system with clearly defined concepts. These code systems can be used in conjunction with each other.

4. Discussion and Conclusions

Instead of mixing unclear or not well-defined concepts into a single value set, multiple codes from clean code systems should be conveyed. Paired with timing aspects such a transmission allows for a complete and correct patient history, if jurisdictional, legal and data privacy aspects are neglected.

References

[1] Geibel P, Oemig F, Kassner A: „Fallstudie: Das Geschlecht in den Datensätzen der Health-IT“, in eHealthCom, Magazin für Health IT, vernetzte Medizintechnik und Telemedizin, 06/15, p.48 www.e-health-com.eu

[2] Oemig, Blobel: Practical Implications of Value Set Definitions: an Example. Proceedings of the 15th International HL7 Interoperability Conference, p 25-27. Prague: emina; 2015.

[3] ISO/IEC 5218:2004 (http://standards.iso.org/ittf/PubliclyAvailableStandards/c036266_ISO_IEC_5218_2004(E_F).zip)

[4] Canada National Body: ISO/IEC JTC1 SC36 N0834 - Business Case and for Rationale

[5] Incorporating “Multilingual Capability and Equivalency as one of the Basic Principles, at the Architectural and Structural Levels, for the Development of the Multipart MLR standard”. (http://old.jtc1sc36.org/doc/36N0834.pdf) old.jtc1sc36.org, 19. August 2004, S. 6, Fußnote 2

[6] UK Government Data Standards:- Person Gender at Registration (http://www.datadictionary.nhs.uk/version2/data_dictionary/data_field_notes/p/person_gender_at_registration_de.asp?shownav=0), Feb.15th 2016

[7] Standards and Liaison Committee Health Information and Surveillance Systems Board: Common Data Elements Implementation Guide Version 2.4 (http://www.cdc.gov/DATA/impv24.pdf) (PDF; 338 kB), U.S. Department of Health and Human Services, 3. Januar 2000, S. 35 ff.; Proposed CDC Health Data Standards (http://www.cdc.gov/DATA/INDEX.HTM), Website des Centers for Disease Control and Prevention

[8] RFC 4112: Electronic Commerce Modeling Language (ECML) - Version 2 Specification. Juni 2005, Motorola Laboratories

[9] Multilateral Interoperability Program, NATO (Hrsg.): The Joint C3 Information Exchange Data Model (JC3IEDM Main) (http://www.mip-site.org/publicsite/04-Baseline_3.0/JC3IEDMJoint_C3_Information_Exchange_Data_Model/JC3IEDM-Main-GBRDMWG-Edition_3.1b_2007-12-13.pdf) (PDF) Edition 3.1b, 13. Dezember 2007, mip-site.org

[10] Julie Butler: X Marks the Spot for Intersex Alex. (http://www.sageaustralia.org/docs/West%20Australian%20Newspaper.pdf) In: The Western Australian, Perth, 11. January 2003; HTML (http://www.tgcrossroads.org/news/?aid=597)

[11] Human Rights Commission (Hrsg.): To Be Who I Am - Report of the Inquiry into Discrimination Experienced by Transgender People (http://www.hrc.co.nz/hrc_new/hrc/cms/files/documents/06-May-2008_10-35-46_Trans_Gender_Inquiry_Report_Final_Word_May_08.doc) (MS Word; 1,6 MB), Januar 2008, New Zealand, ISBN 978-0-478-28643-4

[12] Gemeinsame konsularische Instruktion an die diplomatischen Missionen und die konsularischen Vertretungen, die von Berufskonsularbeamten geleitet werden (http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52005XG1222(01):DE:NOT), 22. Dezember 2005, eur-lex.europa.eu

[13] National Electrical Manufacturers Association (Hrsg.): PS 3.3-2007 - DICOM Part 3: Information Object Definitions (http://medical.nema.org/dicom/2007/07_03pu.pdf) (PDF; 4,5 MB) medical.nema.org, 2007, Version: 28. Dezember 2006

[14] ISIS & CGI Information Systems and Management Consultants Inc. (Deepali Thussu): *ZIMS - PM224*

[15] Animal Management - Version 2.00 (http://zims.isis.org/System%20Use%20Cases%20%20PDF/PM224%20-%20Page%20Dictionary%20-%20Animal%20Management%20v2.0.pdf) (Login required), 5. Oktober 2006

[16] The OBIS SchemaVersion 1.1, http://www.iobis.org/tech/provider/schemadef1/, 14. 8. Juli 2005

[17] TDWG-SDD: TWiki - SDD Web - Version101 - Version1dot1, http://wiki.tdwg.org/twiki/bin/view/SDD/Version1dot1, SDD1.1.zip accessed March 19. 2007, there SDD1.1/documentation/UBIF_EnumData.html