An Ontological Representation of Sex and Gender Information

Paul FABRY[a,1], Adrien BARTON[a,b,1] and Jean-François ETHIER[a,1]

[a] *Groupe de Recherche Interdisciplinaire en Informatique de la Santé (GRIIS), Université de Sherbrooke, Quebec, Canada*
[b] *Institut de Recherche en Informatique de Toulouse (IRIT), CNRS, France*

Abstract. Sex and gender are important health determinants. It is therefore valuable to represent them in medical records. However, these entities are complex to define. In this paper we review some existing representations of sex and gender, with a special focus on the informational entities representing them. We detail our proposal for formalizing sex and gender informational entities according to the OBO Foundry methodology. In particular, we introduce classes enabling to represent information that may be either well-defined or ambiguous relatively to whether they represent sex or gender.

Keywords. Sex, Gender, Information content entity

1. Introduction

The terms "sex" and "gender", despite having clearly different meanings, are still sometimes used interchangeably in some contexts to characterize a human individual. While "sex" refers to characteristics that are biologically determined such as chromosomes distribution and reproductive/sexual anatomy and physiology, "gender" refers to socially constructed characteristics [1].

Sex is commonly categorized as female or male, especially in administrative documents, but there is a great variety of biological categories beyond those two. There is also a considerable diversity in gender differentiation, depending on how people perceive themselves (gender identity) and how they express their gender (gender expression).

Evidence of sex and gender differences have been reported in chronic disease, physiological processes and the impact of lifestyle on health [2]. Both sex and gender are important health determinants, therefore information about them is valuable to support care providers. This being the case, it must also be included as variables in clinical research in order to fully evaluate the impact of studied interventions.

Learning Health Systems (LHS) are conceptual frameworks that enable the tight coupling of care delivery, research and knowledge transfer. Starting from data generated

[1] Corresponding Authors. GRIIS, Université de Sherbrooke, 2500, boul. de l'Université, Sherbrooke (Québec), J1K 2R, Canada. IRIT, Université Toulouse III, q118, route de Narbonne, 31062 Toulouse Cedex 9, France. E-mails: paul.fabry@usherbrooke.ca, adrien.barton@irit.fr; jf.ethier@usherbrooke.ca.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

through care delivery, research results are more likely to be relevant and the findings are re-injected in care through knowledge transfer processes like audit/feedback tools and decision support tools. This then triggers a new cycle [3]. LHSs rely on a common, source-independent representation of clinical information in order to support interoperability between various data sources and across the different activities. The goal is to build a better understanding of a human individual through their data across multiple sources. Applied ontologies can provide such a model.

As part of the LHS PARS3 ("Plateforme apprenante en recherche en santé et en services sociaux" - https://griis.ca/en/solutions/pars3), we have developed several ontologies for various domains, such as the Prescription of Drugs Ontology PDRO [4,5] These ontologies can support the creation of a relational schema that is then mapped to databases from various healthcare institutions, in order to support a system of data mediation [6]. As a result, these ontologies are focused on the representation of informational entities (IAO:*information content entity* – ICE) pertaining to health information. The ontological representation of informational entities referring to sex and gender is therefore of primary importance to us.

Many electronic medical record systems integrate sex and gender information into patient demographic information. However, most medical records still only specify the sex of an individual, often labelled "sex assigned at birth", without providing the whole spectrum of possible biological sexes, or the possibility to specify a gender instead of a sex. Sometimes a field asking to choose "M" or "F" will not even specify explicitly whether it refers to sex or gender. Given the ambiguity surrounding both data capture (e.g. unclear question) or data storage (e.g. a database field labelled "M-F" without other information to clearly understand the nature of the information stored in the field), our ontology needs to be flexible enough to allow annotation of ambiguous data elements while providing the means to annotate with a much higher degree of precision. In this paper, we review existing terminological and ontological representations of sex and gender, detail our proposal for formalizing sex and gender informational entities in our demographic data ontology DEMO, and discuss some of its implications.

2. Representation of Sex and Gender

2.1. Representation of Sex and Gender in Health Data Standards

Distinction between sex and gender is acknowledged in the major international health standards such as HL7's Fast Healthcare Interoperability Resource (FHIR) or SNOMED CT [7], as well as in laboratory and radiology standards such as LOINC [8] or DICOM [9]. A comprehensive review is made available by Canada Health Infoway's Sex and Gender working group [10]. We will highlight here the most important aspects as they pertain to our goals.

While many health care standards provide distinct terms or placeholders for sex and gender, the definitions chosen for these elements make them more challenging to integrate into an ontology. For example, in the version 4.0.1 of FHIR, at least three elements are available for gender: *administrative gender*, *person gender* and *patient gender*. *administrative gender* is defined as: "The gender of a person used for administrative purposes." An extension adds *patient gender identity* defined as the gender the patient identifies with. The value set for *patient gender identity* contains more choices than for *patient gender* above. It is important to note that while these four elements contain the term “gender” and refer to it in their definitions, the term “gender” itself is never defined, leaving much uncertainty around the terms. Moreover, it is not clear whether the term refers to gender itself or to an informational entity that is about a gender. Lastly, the use of field value “Other” adds an epistemic dimension incompatible with a stable ontological representation (as the extension of the class “Other” depends on the extension of its sibling classes, which might change if new sibling classes are introduced [11]). So while multiple standards acknowledge the sex and gender terms, more work is required to integrate them in a realist ontology.

2.2. Representation of Sex and Gender in Ontologies

2.2.1. Sex and Gender Entities

The gender, sex and sexual orientation ontology (GSSO) identifies, categorizes and associates to other terminologies thousands of terms related to gender, sex, and sexual orientation [12]. GSSO is available through NCBO BioPortal. However, GSSO is only very partially aligned with the Basic Formal Ontology and articulated with the OBO Foundry principles [13] and it does not integrate the ICEs that are relevant to us.

Biological sex is formalized in the Phenotype And Trait Ontology (PATO) [14] as an organismal quality that determines the bearer’s ability to undergo sexual reproduction. *PATO:biological sex* has the subclasses *PATO:genotypic sex* and *PATO:phenotypic sex* depending on whether the biological sex quality inheres in the bearer’s composition of sex chromosomes or the physical expression of sexual characteristics.

Sex and gender classes in these ontologies are quality or social roles. However, related information content entities are necessary to support our LHS as explained above.

2.2.2. Informational Entities about Sex and Gender

Informational entities about sex are represented in the Vaccine Ontology (VO) [15]. *VO:biological sex datum** is categorized as a child of *IAO:measurement datum* and defined as: “A measurement datum that represents the biological sex of an animal.” Subclasses have been created for representing ICEs about multiple sex kinds. For example: *VO:female biological sex datum*$=_{\mathrm{def}}$. “A biological sex datum that represents the biological sex of an animal (including human) as being female.”

Terms referring to informational entities related to gender can be found in the Ontology of Medically Related Social Entities (OMRSE) [16] as subclasses of *OMRSE:social identity information content entity*. *OMRSE:gender identity information content entity** is defined as: “A social identity information content entity that is about whether some person identifies as some gender.” This class has also several subclasses representing ICEs about multiple gender kinds. For example: *OMRSE:female gender identity information content entity*$=_{\mathrm{def}}$. “A gender identity information content entity that is about some person’s identifying as female in gender.”

3. Sex and Gender in DEMO Ontology

As part of our ontology suite to represent clinical informational entities, we are developing an ontology, named “DEMO” (stands for DEMographics Ontology) that focuses on demographics data, including sex and gender. Our goal concerning those latter data is threefold:

- To formalize different informational entities about sex and gender;
- To not be limited to a male/female or man/woman binary representation, but be able to represent the other possibilities within sex or gender;
- To be able to annotate electronic health records containing representations that are ambiguous with respect to whether they refer to sex or gender (such as a field with two values 'M' and 'F', without more specifications about what those refer to).

In accordance to the OBO Foundry methodology, we imported the pre-existing classes mentioned above (*) in our ontology. However, those classes were limited to binary alternatives (man/woman for gender, male/female for sex) and we proposed to VO and OMRSE representatives to add classes about biological intersex and non-binary gender. As a result, the following classes were created and imported by DEMO:

- *VO:intersex biological sex datum*$=_{\text{def.}}$ "A biological sex datum that represents the biological sex of an animal (including human) as being intersex."
- *OMRSE:non-binary identity ICE*$=_{\text{def.}}$ "A gender identity ICE that is about some person's identifying as non-binary in gender."

To take into account ambiguous representations of sex and gender, we created the class *DEMO:biological sex or gender identity ICE*$=_{\text{def.}}$ "An ICE that is intended to denote a biological sex or a gender identity." This class is defined as a class equivalent to: *VO:biological sex datum* OR *OMRSE:gender identity ICE.* Three subclasses were added to account for multiple possibilities: DEMO:*female biological sex or woman gender entity ICE,* DEMO:*male biological sex or man gender entity ICE*, and DEMO:*intersex biological sex or non-binary identity ICE*. These subclasses are defined as "A biological sex or gender identity ICE that is intended to denote a female [resp. male, intersex] biological sex or a woman [resp. man, non-binary] gender identity".

The addition of informational entities about intersex and non-binary gender enables us to capture more diverse information about patients' sex and gender. Alongside "male", "female", "man" and "woman", they allow us to represent field values representing intersex, hermaphrodite or gender-fluid identity for example.

Moreover, when we are confronted with records where these fields are insufficiently defined, the *biological sex or gender identity ICE* class works as a catch-all solution. For example, if we have a data source with an undefined field labelled "Sex/Gender" and its possible value are "M" or "F", we can state that this field is about a biological sex or a gender. If this field has the value "F" for a given patient, we can classify it as an information content entity that denotes a female biological sex or a woman gender. While such data constructs are undesirable because of their ambiguity, they are certainly in existence; therefore, not allowing data access to such fields would not be acceptable to users.

4. Discussion and Conclusion

To represent ambiguous information, we could have instead introduced a class *ambiguous biological sex or gender identity information content entity*, whose instances are ICEs that denote ambiguously to either a sex or a gender. However, the ambiguity depends on the context and the receiver of the information: a piece of information can be ambiguous for a user (say, someone who is retrieving the information from an institution to which she does not belong), and non-ambiguous for another user (say, someone who knows how the database has been built). Therefore, the representation of an epistemically charged notion like an “ambiguous” representation is much more complex, and it is not clear that such complexity would bring commensurable gains. This is why we introduced the catch-all class *biological sex or gender identity ICE*.

Axioms using the IAO:**is about** relation should also be added in the future [17]. Note also that although they were created initially for medical records, these classes should be useful whenever sex and gender information are used. Finally, it is important to keep in mind that sex and gender data are highly sensitive data from an ethical point of view. For example, while analyzing gender information may allow us to discover specific health problems for more fragile populations, it entails a risk of categorizing people, possibly against their will or knowledge, in a way that might be detrimental to them.

References

[1] WHO | Gender and Genetics [Internet]. WHO. World Health Organization; [cited 2020 Aug 25]. Available from: https://www.who.int/genomics/gender/en/index1.html.

[2] Burgess C, Kauth MR, Klemt C, et al. Evolving Sex and Gender in Electronic Health Records. Fed Pract. 2019;36:271–277.

[3] Kaggal VC, Elayavilli RK, Mehrabi S, et al. Toward a Learning Health-care System – Knowledge Delivery at the Point of Care Empowered by Big Data and NLP. Biomed Inform Insights. 2016;8:13–22.

[4] Ethier J-F, Barton A, Taseen R. An ontological analysis of drug prescriptions. Appl Ontol. 2018;13:273–294.

[5] Barton A, Fabry P, Ethier J-F. A classification of instructions in drug prescriptions and pharmacist documents. Proceedings of the 10th International Conference on Biomedical Ontology (ICBO 2019). Buffalo, New York, USA; p. 1–7.

[6] Ethier J-F, Curcin V, Barton A, et al. Clinical data integration model. Core interoperability ontology for research using primary care data. Methods Inf Med. 2015;54:16–23.

[7] IHTSDO. SNOMED CT [Internet]. Leading healthcare terminology, worldwide. 2015 [cited 2015 Jan 12]. Available from: http://www.ihtsdo.org/snomed-ct/.

[8] Regenstrief. LOINC [Internet]. Logical Observation Identifiers Names and Codes (LOINC®). 2015 [cited 2015 Jan 12]. Available from: http://loinc.org/.

[9] Gibaud B. The DICOM Standard: A Brief Overview. In: Lemoigne Y, Caner A, editors. Molecular Imaging: Computer Reconstruction and Practice. Dordrecht: Springer Netherlands; 2008. p. 229–238.

[10] Sex and Gender [Internet]. [cited 2020 Jul 2]. Available from: https://infocentral.infoway-inforoute.ca/en/collaboration/wg/sex-gender.

[11] Bodenreider O, Smith B, Burgun A. The Ontology-Epistemology Divide: A Case Study in Medical Terminology. Form Ontol Inf Syst. 2004;2004:185–195.

[12] Kronk CA, Dexheimer JW. Development of the Gender, Sex, and Sexual Orientation ontology: Evaluation and workflow. J Am Med Inform Assoc. 2020;27:1110–1115.

[13] Smith B, Ashburner M, Rosse C, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nature Biotechnology. 2007;25:1251–1255.

[14] The Phenotype And Trait Ontology [Internet]. 2020 [cited 2020 Jul 2]. Available from: https://github.com/pato-ontology/pato.

[15] He Y, Cowell L, Diehl A, et al. VO: Vaccine Ontology. Nat Prec. 2009;1–1.

[16] Hicks A, Hanna J, Welch D, et al. The ontology of medically related social entities: recent developments. J Biomed Semantics. 2016;7:1–4.

[17] Schulz, Stefan, Martínez-Costa, Catalina, Karlsson, Daniel, et al. An Ontological Analysis of Reference in Health Record Statements. Frontiers in Artificial Intelligence and Applications [Internet]. IOS Press; 2014 [cited 2020 Sep 3].