# Proposal for Specification Change

|  |
| --- |
| Request for Comments (As the reviewer reads through this proposal, please consider and comment on the following:)   * How valuable would this change be to your current or foreseeable program of research (and please elaborate if you care to)? * If implemented in the next 6 months, would the change disrupt any current study? * How would the proposed change enhance HCSRN's ability to respond to current or foreseeable funder priorities (e.g., what's currently or about to be "hot")? * Is this implementable at your site? * Does the revised spec maximally preserve the information commonly used in your local data? If not, how could it be changed so that it does? * Would you guess this is a low, medium, or high-effort change at your site (and please elaborate if you care to)? |

|  |  |
| --- | --- |
| Name of Work Group: | Enroll/Demog |
| Request submitted by: | Roy Pardee |
| Date of submission: | 19-sep-2019 |
| Addition or Change? | Both |
| Target Vote Date | ? |
| Target Implementation Date | ? |
| Anticipated Impact | medium effort |
| Motivation for Addition/Change This proposal is intended to cure a defect in the current Demographics specification–namely that it combines the distinct concepts of biological sex and gender identity in a single variable. Existing Spec The current spec sets out a single field called ‘gender’ which may hold either gender (subjective gender identity) or sex (biological sex), or a mixture of the two, depending on which data is available at the site. The spec currently reads as follows:   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Variable Name | Definition | Type(Len) | Values | Implementation Guidelines | | gender | The person’s gender and/or sex; if both gender and sex are known, this variable should hold gender | char(1) | M = Male F = Female O = Other including transgendered U = Unknown |  |   While we expect that the vast majority of values are in fact sex assigned at birth, users currently have no way of knowing whether any given value is sex or gender. | |
| Proposed Addition/Change We propose to:   1. Deprecate/remove the existing **gender** variable. 2. Add a new variable to hold *only* gender identity information (as of last ascertainment) 3. add new fields to hold:    1. sex at birth    2. administrative sex    3. gender identity | |
| Anticipated Impacts (on users and implementers)On Users It would be hard to overstate the degree to which the existing **gender** variable is used. There are probably far more applications that access **gender** than those that do not, and so we don’t take removing it lightly. That said, we consider the proposal a distinct improvement in that under the current spec, users don’t actually know whether they are getting the preferred concept gender, or rather biological sex. If the change is approved, users will have much more control over their applications.  We expect that there will be significant amounts of older code that will need to be revised (including 4 standard macros) for cases where either sex is really what was desired, or where the existing combination-concept scheme is preferable to a pure gender\_identity field. But the revisions will by and large be easy to make, and we judge that the short-term pain is worth the longer-term gain of conceptual clarity and additional information. One advantage to removing the old **gender** field is that old code that is not rewritten will cause errors, which will call attention to the issue. On Implementers It will require work to identify sources for the new variables and incorporate them into existing ETL code for the Demographics file, as well as possibly disentangling gender from sex. We anticipate that at most sites this will be a low-to-medium effort process (do please comment to let us know if this is the case). On the Workgroup We will need to revise the QA program to warn (and eventually fail) implementations that still feature the old **gender** field, and add checks and descriptives for the new fields. | |
| Proposed Specification  |  |  |  |  |  | | --- | --- | --- | --- | --- | | Variable Name | Definition | Type(Len) | Values | Implementation Guidelines | | sex\_admin | The person’s “administrative sex”–the value stored in official/legacy sources at last ascertainment. | char(1) | F = Female M = Male X = Neither Male Nor Female O = Other U = Unknown /uncertain / missing | Legacy sources may designate the information we mean here as either ‘sex’ or ‘gender’ without regard for the distinction between these concepts. In the absence of specific knowledge that a given legacy source codes gender identity, implementers should default to placing legacy data in this variable.  In general, any data collected before your organization began collecting detailed SOGI data should go in this field.  The workgroup expects that sex\_admin will be most similar to the existing gender field in terms of having informative values on >99% of records and will be the best single-field substitute for it.  This is **not** compatible with [PHVS\_AdministrativeSex\_HL7\_2x](https://phinvads.cdc.gov/vads/ViewValueSet.action?id=06D34BBC-617F-DD11-B38D-00188B398520) but should better accomodate the data we expect to be available at the sites. | | sex\_at\_birth | The person’s sex as assigned at birth. | char(1) | F = Female M = Male I = Intersex O = Other U = Uncertain , Unknown or Not recorded on birth certificate C = Choose not to disclose | The existence of the [‘Intersex’](https://en.wikipedia.org/wiki/Intersex) and ‘Other’ options renders this **in**compatible with [LOINC LL3324-2](https://loinc.org/LL3324-2/). Those values appear in our data and are foreseeably useful enough to research that we are comfortable departing from LOINC compatibility.  It is expected that sex\_admin may change over time as a person changes their legal and/or physical sex–this variable should not change. Missing values should be coded as Unknown. | | gender\_identity | The person’s gender identity as subjectively experienced, on last ascertainment. | char(2) | FF = Female MM = Male FM = Female to Male transgender MF = Male to Female transgender GQ = Genderqueer or non-conforming or non-binary or genderfluid OT = Other ND = Choose not to disclose UN = Unknown | Compatible with [PHVS\_GenderIdentity\_CDC](https://phinvads.cdc.gov/vads/ViewValueSet.action?id=660779DA-64E9-E611-A856-0017A477041A) and [LOINC LL3322-6](https://loinc.org/LL3322-6/). Values of ‘unsure/questioning’ should be coded as Other. Missing values should be coded as Unknown. | | |
| Additional Information (Optional)Notes The concept of gender identity is not static, but may change over time. This proposal does not seek to accommodate full information regarding what changes occurred when, but rather just represent the current, best-known information at the time the table is created or updated. This should be sufficient to serve researcher’s needs to identify populations of interest, without unduly burdening users who rely on Demographics’ one-record-per-person structure.  It will be possible to back-translate values from the proposed variables to the old (existing) **gender** scheme. The workgroup will provide SAS code for doing so (possibly in a macro) once the spec for the new vars is finalized. Sexual Orientation We originally attempted to incorporate sexual orientation fields into this change request, but after numerous calls and e-mail threads found that we could not come to a good specification that would integrate the information we expect to be available and be useful for known research. With time and experience using SO data outside of the VDW we may realize that it can be done well (or that it’s worth putting in a separate table). | |