N (Section TBD)

N.n Special Purpose

Non-host Organism Identifiers (OI)

The Non-host Organism Identifiers (OI) domain is used for representing the various levels of organism taxonomy that have been identified for a non-host organism that is the direct object of a test or other observation. This includes viruses and organisms such as pathogens or parasites, but could also be used for non-pathogenic organisms such as normal intestinal flora. The biological classification of a non-host organism typically stops at the taxonomic rank of 'species'. Scientific taxonomic nomenclature below the rank of species is not clearly defined, lacks a globally-accepted standard terminology, and is frequently organism-dependent. The OI domain addresses this issue in the taxonomic nomenclature of non-host organisms. The SDTM variables –NSPCES (non-host species) and –NSTRN (non-host strain) are often inadequate for this purpose for two reasons: 1) the --NSTRN variable allows for identification of only one sub-species taxon (strain) whereas there are often multiple sub-species taxa for a given group of organisms, and 2) the term "strain" is not taxonomically and scientifically defined, hence does not always align with the first, or any, sub-species taxon name for a given group of organisms. The table below shows the variety in both the numbers of sub-species taxa and the nomenclature conventions for those taxa using five species of viruses as examples.

Species	HIV	Influenza A	Hepatitis C	Hepatitis B	HPV
Subspecies Level 1	Type	Subtype	Genotype	Genotype	Type
Subspecies Level 2	Group	Strain	Subtype	Sub-genotype	
Subspecies Level 3	Subtype (or Clade)			Recombination Type	
Subspecies Level 4	Subclade				

Table 1: Sub-species taxa for five species of viruses. Each value appearing below a species name represents the name of a taxon for that species. The taxa progress from less specific to more specific as you move down a column. Each of these taxa have multiple possible values (not shown), and could therefore be viewed as similar to variable names.

Note that if SDTM were to accommodate specific variables corresponding to the values of the various sub-species levels described in the table above, a total of 8 unique variable names would be required. Additionally, at least one of those variables (Subtype) would represent different levels of subspecies nomenclature in

different viruses. Finally, even where the sub-species nomenclature does align, the meaning of the taxon name is not identical (consider "Type" in the context of HIV and HPV—both represent the first level below species, but an HIV "Type" is not the same thing as an HPV "Type" despite the fact that both would share the same variable name). The ultimate identification of more nomenclature conventions for different viruses and other pathogens or parasites will likely compound this problem.

The OI domain solution to this problem is to contain this hierarchical nomenclature in a separate dataset thereby removing the burden from any parent domain where observations about the organism are contained. All unique organisms are assigned a sponsor-defined non-host organism identifier (NHOID) in the OI domain. Uniqueness is defined by the specific values of the organism's entire taxonomy (described by pairs of taxon name and taxon value) to whatever level is known. Any domain that contains observations about this known organism can make use of NHOID, which can in turn be used to link to the OI domain for the full taxonomy of that organism. Sometimes a test seeks to determine the identity of an organism. In these cases the organism's taxonomy is unknown. NHOID should not be used in these cases as the identification would be the result as opposed to something already established.

Note that this approach will also entail deprecation of the –NSTRN variable. –NSPCES will still be available to allow for findings based on organisms where the species name alone is sufficient, or for sorting records based on species in a co-infection study.

OI – Description/Overview for Non-host Organism Identifiers Domain Model

The Non-host Organism Identifiers domain is for storing the levels of taxonomic nomenclature of microbes or parasites that have been either experimentally determined in the course of a study, or are previously known as in the case of lab strains used as reference in the study. It is not to be used for host species identification such as for animals used in pre-clinical studies, nor should it be used to represent other, non-taxonomy characteristics of non-host species such as drug susceptibility, growth rates, etc.
Note: This draft domain contains a new proposed variable, *NHOID*.

OI – Specification for Non-host Organism Identifiers Domain Model

oi.xpt, Non-host Organism Identifiers —Special Purpose, Version 1.0. One record per non-host organism identifier per non-host organism, Tabulation.
Note: This domain table follows the convention set by the Device Identifiers (DI) domain, including both "Definition" and "Implementation Notes" columns.

Variable Name	Variable Label	Type	Controlled Terms, Codelist or Format	Role	Definition	Implementation Notes	Core
STUDYID	Study Identifier	Char		Identifier	Unique identifier for a study.	Unique identifier for a study.	Req
DOMAIN	Domain Abbreviation	Char	OI	Identifier	Two-character abbreviation for the domain.	Two-character abbreviation for the domain.	Req

NHOID	Sponsor Non-host Organism Identifier	Char		Identifier	Sponsor-defined identifier for a non-host organism	NHOID should be populated with an intuitive name based on the identity of the organism as reported by the lab. It must be unique for each unique organism as defined by the specific values of the organism's entire known taxonomy described by pairs of OIPARMCD and OIVAL .	Req
OISEQ	Sequence Number	Num		Identifier	Sequence number given to ensure uniqueness within a parameter within an organism (NHOID) within dataset		Req
OIPARMCD	Non-host Organism Identifier Element Short Name	Char	*(will request CDISC-controlled extensible code list)	Topic	Short name of the taxon being described	Examples: GROUP, GENTYP, SUBTYP	Req
OIPARM	Non-host Organism Identifier Element Name	Char	*(will request CDISC-controlled extensible code list)	Synonym Qualifier	Name of the taxon being described	Examples: Group, Genotype, Subtype	Req
OIVAL	Non-host Organism Identifier Element Value	Char	*	Result Qualifier	Value for the taxon in OIPARMCD/OIPARM for the organism identified by NHOID .	Examples: "M" when OIPARM="Group" for an HIV-1 virus, "2" When OIPARM="Genotype" for HCV virus, "H3N2" when OIPARM="Subtype" for Influenza A virus	Req

Indicates variable may be subject to controlled terminology, (Parenthesis indicates CDISC/NCI codelist code value)

OI – Assumptions for Non-host Organism Identifiers Domain Model

1. Non-Host Organism Identifiers (OI) is a special-purpose domain that provides a mechanism for combining multiple taxonomic identifiers in a single identifier variable (NHOID) to describe the taxonomic nomenclature for each unique non-host organism identified.
2. NHOID is sponsor defined, with the following constraints:
   a. A unique NHOID must represent a unique identity as represented in its combination of OIPARMCD/OIVAL pairs. If two organisms share the same first two levels of taxonomy with regard to OIPARMCD/OIVAL, but one is identified to a third level and the other is not, they should be assigned two unique NHOIDs.
   b. Study sponsors should populate NHOID with intuitive name values based on either:

---

i.   the name of the organism as reported by a lab or specified by the investigator, or
ii. published references/databases where applicable and appropriate (e.g., when reference strain H77 is used in a HCV study, NHOID for this strain should be populated with "H77" or "HCV1a-H77").
3. NHOID can be used in any domain where observations about these organisms are being represented, allowing end-users to determine what is known about the organism's identity by merging on NHOID, or by otherwise referring to the OI domain.
4. OIPARMCD and OIPARM must represent parameters for the identification of non-host organisms with regard to nomenclature only
   a. Mostly, this will represent taxonomic ranks (i.e. Species) as well as commonly used grouping terms (taxa that aren't officially ranked) such as "subtype", "group", "strain", etc.
   b. They may also include other nomenclature terms that are less widely known but are used frequently for organism identification in a specific field of study (e.g., "spoligotype" in tuberculosis).
   c. They should be listed in the OI dataset in hierarchical order of least to most specific with increasing OISEQ values.
5. Variables not listed in the OI domain table above should not be used in OI data sets.

OI – Examples for Non-host Organism Identifiers Domain Model

Example 1
This example shows drug sensitivity testing for HIV and Hep C at one timepoint in one subject participating in a co-infection study. An OI dataset example follows the MS dataset example below.

Rows 1 and 4:   Shows the response of the virus extracted or derived from the subject based on drug concentrations required to produce 50% inhibition of the standard virus growth.
Rows 2 and 5:   Show a reference viral sample response based on drug concentrations required to produce 50% inhibition of the standard virus growth.
Rows 3 and 6:   Show the fold change of the response of the virus extracted from the subject compared to the reference viral sample response based on drug concentrations required to produce
50% inhibition of the standard virus growth. This is the subject sample result divided by the reference result. Because these records are derived, VRDRVFL=Y.
All Rows: Note the use of NHOID as a sponsor-defined non-host organism identifier. This ID serves as a link from findings on the virus to that virus' taxonomic identifiers in the Organism Identifiers (OI) dataset.

ms.xpt

Row	STUDYID	DOMAIN	USUBJID	MSSEQ	MSGRPID	NHOID	MSTESTCD	MSTEST	MSDRUG	MSORRES	MSORRESU	MSSTRESC
1	COINF1	MS	COINF1-01	1	1	HIV1MC	IC50S	IC50 Subject Result	Experimenavir	0.2	nM	0.2
2	COINF1	MS	COINF1-01	2	1	HIV1MB	IC50R	IC50 Reference Control Result	Experimenavir	0.21	nM	0.21
3	COINF1	MS	COINF1-01	3	1		IC50FCR	IC50 Fold Change from Reference	Experimenavir	0.95		0.95
4	COINF1	MS	COINF1-01	4	2	HCV2C	IC50S	IC50 Subject Result	Heprevir	1.35	nM	1.35
5	COINF1	MS	COINF1-01	5	2	H77	IC50R	IC50 Reference Control Result	Heprevir	1.21	nM	1.21

This example shows the Non-host Organism Identifiers (OI) dataset which is used to represent the taxonomic identifiers for the pathogens identified by NHOID in the MS dataset above.

All Rows: NHOID is a unique non-host organism ID used to link findings on that organism in other datasets with details about its identification in OI. OIPARM shows the name of the individual taxa identified and OIVAL shows the experimentally determined values of those taxa.
Rows 1 to 4: Show the taxonomy for the HIV organism given the NHOID of HIV1MC. This virus has been identified as HIV-1, Group M, Subtype C.
Rows 5 to 8: Show the taxonomy for the HIV organism given the NHOID of HIV1MB, which was used as a reference. This virus has been identified as HIV-1, Group M, Type B.
Rows 9 to 11: Show the taxonomy for the HCV organism given the NHOID of HCV2C. This virus has been identified as HCV 2c.
Rows 12 to 14: Show the taxonomy for the HCV organism given the NHOID of H77. This virus is a known reference strain of HCV 1a.

oi.xpt

Row	STUDYID	DOMAIN	NHOID	OISEQ	OIPARMCD	OIPARM	OIVAL
1	STUDY123	OI	HIV1MC	1	SPCIES	Species	HIV
2	STUDY123	OI	HIV1MC	2	TYPE	Type	1
3	STUDY123	OI	HIV1MC	3	GROUP	Group	M
4	STUDY123	OI	HIV1MC	4	SUBTYP	Subtype	C
5	STUDY123	OI	HIV1MB	1	SPCIES	Species	HIV
6	STUDY123	OI	HIV1MB	2	TYPE	Type	1
7	STUDY123	OI	HIV1MB	3	GROUP	Group	M
8	STUDY123	OI	HIV1MB	4	SUBTYP	Subtype	B
9	STUDY123	OI	HCV2C	1	SPCIES	Species	HCV
10	STUDY123	OI	HCV2C	2	GENTYP	Genotype	2
11	STUDY123	OI	HCV2C	3	SUBTYP	Subtype	C
12	STUDY123	OI	H77	1	SPCIES	Species	HCV
13	STUDY123	OI	H77	2	GENTYP	Genotype	1
14	STUDY123	OI	H77	3	SUBTYP	Subtype	A