DPCC Data Standard Reference for Virus Reagent v1.2

	Project_Identifier	Contributing_Institution	Reagent_Name
Input Type	Text Field	Text Field	Text Field
Definition	A unique Project Identifier generated by the DPCC by combining the Center-generated Project Code and a random 4-digit number	The institution code for the group that CREATED THE REAGENT. Must be one of the CEIRS institution codes assigned by NIAID.	The WHO or ICTV strain name of the virus submitted as a reagent
Format	Project_Code_XXXX Maximum length: 21 characters	Center three-letter code followed by three digits. Maximum length: 6 characters	Influenza A virus: Antigenic Type/Host of Origin/Geographical Origin/Strain Number/Year of Isolation (Subtype), or rg-Antigenic Type/Host of Origin/Geographical Origin/Strain Number/Year of Isolation [rg details] (Subtype) Influenza B, C, or D virus: Antigenic Type/Host of Origin/Geographical Origin/Strain Number/Year of Isolation, or rg-Antigenic Type/Host of Origin/Geographical Origin/Strain Number/Year of Isolation [rg details] SARS-CoV-2 and other viruses: Virus Name/Host of Origin/Geographical Origin/Strain Number/Year of Isolation, or rg-Virus Name/Host of Origin/Geographical Origin/Strain Number/Year of Isolation [rg details] Maximum length: 150 characters
Value List	None	None	Text U
Curation	The entry must be a Project Identifier value registered with the DPCC.	The entry must be an Institution Code value registered with the DPCC. Center 3-letter codes are case-sensitive and must be entered in all-caps.	For influenza viruses, strain names most follow WHO naming convention: Fields must be ordered as follows and separated with the '/' character: 1. The antigenic type (e.g., A, B, C, D) 2. The host of origin (e.g., swine, equine, chicken. For human-origin viruses, no host of origin designation is given.) 3. Geographical origin (e.g., Denver, Taiwan) 4. Strain number (e.g., 15, 7) 5. Year of isolation (e.g., 2009, 1934) 6. For influenza A viruses, the hemagglutinin and neuraminidase antigen description in parentheses (e.g., (H1N1), (H3N2)) For SARS-CoV-2 viruses, strain names most follow ICTV naming convention: Fields must be ordered as follows and separated with the '/' character: 1. Virus name (e.g., SARS-CoV-2) 2. The host of origin (e.g., human. Human-origin viruses must include the origin designation.) 3. Country of geographical origin as a three-letter code from the DPCC's Country Codes list (e.g., USA, MEX, CAN) 4. Strain number (e.g., 15, 7) 5. Year of isolation (e.g., 2019, 2020) For all other viruses, please use the following convention: Fields must be ordered as follows and separated with the '/' character: 1. Virus name (e.g., MERS-CoV, Bat-CoV, etc.) 2. The host of origin (e.g., human, bat, camel. Human-origin viruses must include the origin designation.) 3. Geographical origin, either regional locality or country (e.g., Denver, Taiwan) 4. Strain number (e.g., 15, 7) 5. Year of isolation (e.g., 2019, 2020) If the strain is a reverse genetic virus, prefix Strain_Name with rg-. Provide reverse genetic details in brackets after the year of isolation and before the subtype.
Examples	SJCProj02_4001	SJC101	For influenza: A/Hong Kong/1/1968 (H3N2), A/chicken/Fujian/4/2002 (H3N6), A/chicken/Fujian/4/2002 (HxNx), A/chicken/Fujian/4/2002 (mixed), A/swine/Iowa/233-56/2011 (H3N2), A/duck/Alberta/35/1976 (H1N1), B/Hong Kong/432/2014, C/Texas/19876/2011, D/swine/Oklahoma/1334/2011, or rg-A/Puerto Rico/8/1934 (H1N1), or rg-A/Egypt/N03072/2010 [PR8 internal R 6+2] (H5N1) For SARS-CoV-2: SARS-CoV-2/human/USA/NY-PV08486/2020 For other viruses: PHEV-CoV/swine/USA/15TOSU25049/2015
Notes			WHO Reference for influenza viruses: http://www.cdc.gov/flu/about/viruses/types.htm ICTV Reference for SARS-CoV-2 viruses: https://pubmed.ncbi.nlm.nih.gov/32123347 (HxNx) can be used in cases where a partial subtype has been determined (e.g., H5Nx, HxN2). If there are mixed subtypes contained within a sample use A/chicken/Fujian/4/2002 (mixed) for Strain_Name and enter additional subtype information in the Comments field.
Dependent Fields			
Validation	Project_Identifier should be a valid project identifier.	Must be an Institution Code value registered with the DPCC	Validate field length
Message Code	Error_9_PROJECT_NOT_FOUND	Error_1_INVALID_VALUE	Error_70_INVALID_FIELD_LENGTH
Validation	Project_Identifier exists but user does not have permission to access or edit the project.		Geographical origin element of strain name must be alphanumeric, underscore, dash, period, or single quote: a-z, A-Z, 0-9, _, -, ., '
Message Code	Error_4_DENIED_USER_ACCESS		Error_138_INVALID_LOCATION_STRAIN_TEXT
Validation			Strain number element of strain name must be alphanumeric, underscore, dash, or period: a-z, A-Z, 0-9, _, -, .
Message Code			Error_139_INVALID_STRAIN_NUMBER_STRAIN_TEXT
Validation			Year of isolation element of strain name must be 4-digit year.
Message Code			Error_110_INVALID_STRAIN_YEAR
Validation			If strain is influenza A, subtype must be present as text between parentheses.
Message Code			Error_112_INVALID_STRAIN_SUBTYPE
Validation			Brackets cannot be present if strain does not have rg- prefix.
Message Code			Error_119_INVALID_STRAIN_BRACKETS
Validation			If included, the strain must have a matched pair of opening and closing brackets or parentheses.
Message Code			Error_158_INVALID_STRAIN_OPEN_BRACKETS
Validation			
Message Code			
Validation			
Message Code			
Validation			
Message Code			
Validation			
Message Code			
Validation			
Message Code			

	Sample_Identifier	Passage_History	Parent_Sample_Identifier	Parent_Strain_Name
Input Type	Text Field	Text Field	Text Field	Text Field
Definition	Identifier assigned to the virus used to create the reagent	Description of the passage history of the virus strain	A series of identifiers referencing the parent virus or viruses used to create the reagent	The name of the parent strains used to create the reverse genetic or reassortant virus strain
Format	Center-specific Allowed characters include alphanumeric, hyphen, and underscore: a-z, A-Z, 0-9, -, _ Maximum length: 50 characters	Text Maximum length: 50 characters	Maximum of eight comma-separated Sample_Identifiers. Allowed characters include alphanumeric, hyphen, and underscore: a-z, A-Z, 0-9, -, _ Maximum length: 800 characters	Eight comma-separated Strain_Names Antigenic Type/Host of Origin/Geographical Origin/Strain Number/Year of Isolation (Subtype) Maximum length: 800 characters
Value List	Text NA	Text X	Text U NA	Text U NA
Curation	It should match with the Sample_Identifier provided in related data submissions.	To indicate a change in passage history at one institution, combine passage host and number: E5C2 To indicate a change in passage history for samples between users or institutions, separate passage host and number with a slash '/': E5/C2	Each Sample_Identifier provided must match with a Sample_Identifier provided in related data submissions. For virus reagents derived from a surveillance sample, provide the Sample_Identifier of the surveillance record. For reverse genetic or reassortant virus reagents, provide the Sample_Identifier associated with each gene segment in the following comma-separated order: PB2, PB1, PA, HA, NP, NA, MP, NS	The entry must list a value for each of the eight gene segments. Strain_Names for each gene segment must be listed in the following comma-separated order: PB2, PB1, PA, HA, NP, NA, MP, NS Strain_Names must follow WHO strain naming convention.
Examples	21145-001a	E3, E2C2, EX, or E5/C2	13250,13250,U,21145,13250,21145,13250,13250, or NA	A/Puerto Rico/8/1934 (H1N1),A/Puerto Rico/8/1934 (H1N1),U,A/Puerto Rico/8/1934 (H1N1),A/Puerto Rico/8/1934 (H1N1),A/Vietnam/1203/2004 (H5N1),A/Puerto Rico/8/1934 (H1N1),A/Puerto Rico/8/1934 (H1N1), or NA
Notes	Enter NA if the reagent is not related to a surveillance submission.	Indicate lab host and number of passages: E = Embryonated chicken eggs C = MDCK cells S = MDCK-SIAT cells M = Monkey kidney cells For other hosts, write out common name and include passage number e.g., Mouse2, E(duck)1, 293T1. Use X if the host or passage number is unknown, e.g., EX, CXE2. Enter X if passage history information is unknown. Enter NA if virus was not passaged. Additional information relating to viral passaging can be entered in the Comments field.	If the reagent is a reverse genetic or reassortant virus, eight comma-separated entries must be present. If an individual Sample_Identifier is unknown, enter U for the specific gene segment. Enter NA once if the strain is not a reverse genetic or reassortant virus.	The Strain_Name for each segment of the reverse genetic or reassortant virus must be named. If an individual Strain_Name is unknown, enter U for the specific gene segment. Enter NA once if the strain is not a reverse genetic or reassortant virus.
Dependent Fields			Parent_Strain_Name	Parent_Sample_Identifier
Validation	Validate field length	Validate field length	Validate field length	Validate field length
Message Code	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH
Validation			Fields Parent_Sample_Identifie and Parent_Strain_Name must have same number of values	Fields Parent_Sample_Identifier and Parent_Strain_Name must have same number of values
Message Code			Error_63_INVALID_NUMBER_ENTRIES	Error_63_INVALID_NUMBER_ENTRIES
Validation				Count number of segments. It should be 8
Message Code				Error_1_INVALID_VALUE
Validation				If Parent_Sample_Identifier is NA then Parent_Sample_Strain should be NA
Message Code				Error_1_INVALID_VALUE
Validation				Host of origin element of strain name must be alphanumeric, underscore, dash, period, or single quote: a-z, A-Z, 0-9, _, -, ., '
Message Code				Error_137_INVALID_HOST_SPECIES_STRAIN_TEXT
Validation				Geographical origin element of strain name must be alphanumeric, underscore, dash, period, or single quote: a-z, A-Z, 0-9, _, -, ., '
Message Code				Error_138_INVALID_LOCATION_STRAIN_TEXT
Validation				Strain number element of strain name must be alphanumeric, underscore, dash, or period: a-z, A-Z, 0-9, _, -, .
Message Code				Error_139_INVALID_STRAIN_NUMBER_STRAIN_TEXT
Validation				Year of isolation element of strain name must be 4-digit year.
Message Code				Error_110_INVALID_STRAIN_YEAR
Validation				If strain is influenza A, subtype must be present as text between parentheses.
Message Code				Error_112_INVALID_STRAIN_SUBTYPE
Validation				Year of isolation element of strain name must be 4-digit year.
Message Code				Error_110_INVALID_STRAIN_YEAR
Validation				If strain is influenza A, subtype must be present as text between parentheses.
Message Code				Error_112_INVALID_STRAIN_SUBTYPE
Validation				Brackets cannot be present if strain does not have rg- prefix.
Message Code				Error_119_INVALID_STRAIN_BRACKETS
Validation				If included, the strain must have a matched pair of opening and closing brackets or parentheses.
Message Code				Error_158_INVALID_STRAIN_OPEN_BRACKETS
Validation				The strain must have four or five elements separated by a slash '/' character.
Message Code				Error_159_INVALID_NUMBER_OF_STRAIN_ELEMENTS

	Mutations	BEI_Number	Contact_Name	Contact_Email
Input Type	Text Field	Text Field	Text Field	Text Field
Definition	A description of the mutation(s) or non-coding change(s) made in each segment of the reverse genetic or reassortant virus	The identification number assigned to the reagent by BEI	The name of the point-of-contact for reagent requests or the name of the BEI depositor	Email address for the point-of-contact
Format	An HGVS-formatted list describing sequence changes Maximum length: 400 characters	BEI Number Maximum length: 20 characters	FirstName LastName FirstName MI. LastName Maximum length: 50 characters	Text@domain.com Maximum length: 50 characters
Value List	Text NA	Text NA	None	None
Curation	The entry must use the HGVS reference for annotation guidelines to describe mutations, insertions, or deletions in any gene segment order. http://www.hgvs.org/mutnomen/quickref.html	The entry must be a valid BEI number.	None	The entry must be a valid email address.
Examples	HA:p.[N220K; Q222L],NS:p.[H155del],PB2:p.[E627K]	NR-12277	Jane Smith, or Davey M. Jones	jsmith@email.edu
Notes	<ins>Format for amino acid mutation: protein_name:p.[mutation_information]</ins> For a single mutation, list the original amino acid, its position, and the new amino acid in brackets: PB2:p.[E627K] For a multiple mutations within a single protein, separate individual mutations by a semicolon: HA:p.[N220K;Q222L] For a deletion within a single protein, list the amino acid, its position, and 'del': NS:p.[H155del] For a stop codon, use an asterisk: NA:p.[K471_*472insHHHHHH] <ins>Format for non-coding mutation: gene_segment_name:c.[mutation_information]</ins> For a single mutation: enter the position, the original nucleotide, >, and the new nucleotide: PB2:c.[273C>T] For multiple mutations within a single gene, separate individual mutations by a semicolon: PB1:c.[273C>T;274A>G;275G>T] For a deletion: enter the position, del, and the nucleotide deleted: NA:c.[73delT] To list both amino acid and non-coding changes, comma-separate the individual mutation information in any order: HA:p.[N220K;Q222L],HA:c.[273C>T;274A>G;275G>T],PB1:c.[273C>T;274A>G;275G>T] Enter NA for non-reverse genetic or reassortant viruses, or viruses with no known mutations.	Enter NA if the reagent was not deposited with BEI.		
Dependent Fields				
Validation	Validate field length	Validate field length	Validate field length	Validate field length
Message Code	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH
Validation				Must be a valid email address.
Message Code				Error_114_INVALID_EMAIL
Validation				
Message Code				
Validation				
Message Code				
Validation				
Message Code				
Validation				
Message Code				
Validation				
Message Code				
Validation				
Message Code				
Validation				
Message Code				
Validation				
Message Code				
Validation				
Message Code				
Validation				
Message Code				

	Availability	Publication_PMID	Comments
Input Type	Text Field	Text Field	Text Field
Definition	Whether the reagent is available to share with other investigators on the reagents page of the CEIRS Public Website	The PubMed Unique Identifier (PMID) for the publication in which the reagent was published.	Text describing anything else of interest related to the submission
Format	Text Maximum length: 1 character	PMID Maximum length: 50 characters	Text Maximum length: 2000 characters
Value List	Y N	Text NA	Text NA
Curation	The entry must be one and only one member of the Value List.	The entry must be a valid PMID number: 7 or 8-digit number with no leading zeros. http://www.nlm.nih.gov/bsd/mms/medlineelements.html#pmid	None
Examples	Y	19465683, or NA	NA
Notes	Y = Yes N = No	If the reagent was referenced in multiple publications, comma-separate the individual PMIDs. Enter NA if the reagent was not referenced in a publication.	If there are no comments, enter NA.
Dependent Fields			
Validation	Field value should be one of valid values as in list.	Validate field length	Validate field length
Message Code	Error_1_INVALID_VALUE	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH
Validation		Must be a valid PMID.	
Message Code		Error_96_INVALID_PMID	
Validation			
Message Code			
Validation			
Message Code			
Validation			
Message Code			
Validation			
Message Code			
Validation			
Message Code			
Validation			
Message Code			
Validation			
Message Code			
Validation			
Message Code			
Validation			
Message Code			
Validation			
Message Code			