

DPCC Data Standard Reference for Sequence Metadata v2.2						
Input Type	Project_Identifier	Contributing_Institution	Sample_Identifier	BioProject_Accession_Number	Embargo_End_Date	Provisional_Authors
Text Field	Text Field	Text Field	Text Field	Text Field	Date 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 PERFORMED THE ANALYSIS. Must be one of the CEIRS institution codes assigned by NIAID	Identifier initially assigned to each sample collected. If multiple samples are taken from the same host, each sample should have its own identifier.	The BioProject unique accession number(s) associated with the sequence submission	In case an embargo of the information is needed, the date that the information should be released to the public databases by the DPCC	The list of authors associated with the sequence submission. Default entry in absence of a Publication_PMID will be the authors listed by the user. If a Publication_PMID is provided, publication authors will be listed.
Format	Project_Code_XXXX Maximum length: 21 characters	Center three-letter code followed by three digits. Maximum length: 6 characters	Center-specific Allowed characters include alphanumeric, hyphen, and underscore: a-z, A-Z, 0-9, -, _ Maximum length: 50 characters	BioProject ID Maximum length: 100 characters	DD-Mon-YYYY DD-Mon-YY NA Maximum length: 11 characters	FirstName LastName FirstName M. LastName Maximum length: 500 characters
Value List	None	None	None	None	Date NA	Text NA
Duration	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.	The Sample_Identifier initially assigned to the surveillance sample must be provided.	The entry must be a valid BioProject accession number previously registered with the DPCC.	1. Leading 0 in DD is optional. 2. Month must match the first three letters of the month. Month is NOT case-sensitive. 3. Years may have two or four digits. 4. Date must conform to NIAID data release policies.	If a Publication_PMID is entered, use NA. The list of authors from the publication will be used.
Examples	SJCProj02_4001	SJC101	22258468	PRJNA37813	3-Mar-2011, 03-Mar-2011, 03-MAR-2011, 3-MAR-11, or NA	Indiana Jones, John M. Henry, or NA
Notes	None	None	Submissions without a corresponding surveillance submission will be listed as orphaned records. For non-surveillance submissions, enter a unique Sample_Identifier.	The sequence must be linked to a BioProject. If no relevant BioProject exists, one must be created.	An exact date must be provided. If Embargo_End_Date is NA, information will be released without delay. Embargo_End_Date cannot be more than 12 months after submission.	The DPCC will update the GenBank entry records if and when the authors provide a publication referenced after submission. For listing of multiple names, comma-separate the names maintaining the order of FirstName and LastName or FirstName M. LastName
Dependent Fields						Publication_PMID
Validation	Project_Identifier should be a valid project identifier.	Must be an Institution Code value registered with the DPCC	Validate field length	Must be a valid BioProject accession number.	Embargo_End_Date must be exact date and cannot be a past date.	Validate field length
Message Code	Error_9_PROJECT_NOT_FOUND	Error_1_INVALID_VALUE	Error_70_INVALID_FIELD_LENGTH	Error_104_INVALID_BIOPROJECT	Error_7_INVALID_EMBARGO_DATE	Error_70_INVALID_FIELD_LENGTH
Validation	Project_Identifier exists but user does not have permission to access or edit the project.		If Surveillance_Sample is N, Sample_Identifier should be unique across all DPCC data.	Validate field length	Embargo_End_Date cannot be more than 12 months after submission.	Multiple names must be entered as comma-separated values in the order of FirstName LastName or FirstName M. LastName.
Message Code	Error_4_DENIED_USER_ACCESS		Error_66_NON_UNIQUE_SAMPLE_ID	Error_70_INVALID_FIELD_LENGTH	Error_100_EMBARGO_DATE_12M	Error_76_INVALID_NAME_LIST
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						

	Strain_Name
Input Type	Text Field
Definition	The WHO or ICTV strain name of the virus that was sequenced
Format	<p>Influenza A virus: Antigenic Type/Host of Origin/Geographical Origin/Strain Number/Year of Isolation (Subtype)</p> <p>Influenza B, C, or D virus: Antigenic Type/Host of Origin/Geographical Origin/Strain Number/Year of Isolation</p> <p>SARS-CoV-2 and other viruses: Virus Name/Host of Origin/Geographical Origin/Strain Number/Year of Isolation</p> <p>Maximum length: 150 characters</p>
Value List	Text U
Curation	<p>For influenza viruses, strain names most follow WHO naming convention: Fields must be ordered as follows and separated with the "/" character:</p> <ol style="list-style-type: none"> 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)) <p>For SARS-CoV-2 viruses, strain names most follow ICTV naming convention: Fields must be ordered as follows and separated with the "/" character:</p> <ol style="list-style-type: none"> 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) <p>For all other viruses, please use the following convention: Fields must be ordered as follows and separated with the "/" character:</p> <ol style="list-style-type: none"> 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)
Examples	<p>For influenza: A/Hong Kong/1/1968 (H3N2), A/chicken/Fujian/4/2002 (H3N6), A/chicken/Fujian/4/2002 (H3Nx), 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/1987/2011, or D/swine/Oklahoma/1334/2011</p> <p>For SARS-CoV-2: SARS-CoV-2/human/USA/NY-PV08486/2020</p> <p>For other viruses: PHEV-CoV/swine/USA/15TOSU25049/2015</p>
Notes	<p>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</p> <p>(H3Nx) can be used in cases where a partial subtype has been determined (e.g., H5Nx, H3N2).</p> <p>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.</p>
Dependent Fields	
Validation	Validate field length
Message Code	Error_70_INVALID_FIELD_LENGTH
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	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	Brackets cannot be present if strain does not have rg_ prefix.
Message Code	Error_159_INVALID_NUMBER_OF_STRAIN_ELEMENTS
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	Brackets cannot be present if strain does not have rg_ prefix.
Message Code	Error_159_INVALID_NUMBER_OF_STRAIN_ELEMENTS

	Lab_Host	Parent_Strain_Name
Input Type	Text Field	Text Field
Definition	Description of the live animal host used for passaging the virus without additional information.	Name of the parental virus strain.
Format	Text Maximum length: 50 characters	Influenza A virus: Antigenic Type/Host of Origin/Geographical Origin/Strain Number/Year of Isolation (Subtype), or Influenza B, C, or D virus: Antigenic Type/Host of Origin/Geographical Origin/Strain Number/Year of Isolation, or SARS-CoV-2 and other viruses: Virus Name/Host of Origin/Geographical Origin/Strain Number/Year of Isolation, or Maximum length: 800 characters
Value List	Text NA	Text U 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 must 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)
Duration	None	For influenza viruses , strain names must 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 must 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)
Examples	ferret, or mouse	For influenza: A/Hong Kong/1/1968 (H3N2), A/chicken/Fujian/4/2002 (H3N6), A/chicken/Fujian/4/2002 (H5Nx), 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. For SARS-CoV-2: SARS-CoV-2/human/USANY-PV08486/2020 For other viruses: PHEV-CoV/swine/USA/15TOSU25049/2015
Notes	This field is strictly reserved for the common or scientific name of live animals associated with generation of the sequence. Enter passage history or experimental information in the Passage_History, Treatment, Transmission_Method, Severity, or Phenotype fields.	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	Validate field length	Validate field length
Message Code	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH
Validation		Influenza virus strain names must begin with the single letter A, B, C, or D, or be prefixed with rg-. Other strain names are not validated.
Message Code		QC_136_STRAIN_NAME_NOT_VALIDATED
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		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		Brackets cannot be present if strain does not have rg- prefix.
Message Code		Error_159_INVALID_NUMBER_OF_STRAIN_ELEMENTS

	Passage_History	Antigenic_Characterization	Treatment	Transmission_Method
Input Type	Text Field	Text Field	Text Field	Text Field
Definition	Description of the passage history of the virus strain	Any information about antigenic characterization	Description of any experimental treatments	Description of the experimental method for virus transmission
Format	Text Maximum length: 50 characters	Text Maximum length: 50 characters	Text Maximum length: 200 characters	Text Maximum length: 50 characters
Value List	Text ORI X	Text NA	Text NA	Text NA
Duration	To indicate a change in passage history at one institution, combine passage host and number: ESC2 To indicate a change in passage history for samples between users or institutions, separate passage host and number with a slash (?): ESC/2	None	None	None
Examples	E2	NA	NA	Aerosol contact
Notes	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 passage number is unknown, e.g., EX, CXE2. Enter X if passage history information is unknown. Enter ORI if virus was not passaged and the sequence was derived directly from the original sample material. Additional information relating to viral passaging can be entered in the Comments field.			
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				
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				

