

DPCC Data Standard Reference for Sequencing Request v1.1

| | Project_Identifier | Sample_Identifier | Virus_Identifier | Sequencing_Study_Identifier |
|-------------------------|--|---|---|---|
| Input Type | Text Field | 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 | Identifier initially assigned to each sample collected. If multiple samples are taken from the same host, each sample should have its own identifier. | A unique laboratory identifier assigned to the virus by the collector or creator | The unique code associated with the sequencing study |
| Format | Project_Code_XXXX Maximum length: 21 characters | Center-specific Allowed characters include alphanumeric, hyphen, and underscore: a-z, A-Z, 0-9, -, _ Maximum length: 50 characters | Text Allowed characters include alphanumeric, hyphen, and underscore: a-z, A-Z, 0-9, -, _ Maximum length: 50 characters | Text Allowed characters include alphanumeric, hyphen, and underscore: a-z, A-Z, 0-9, -, _ Maximum length: 50 characters |
| Value List | None | None | Text NA | Text NA |
| Curation | The entry must be a Project Identifier value registered with the DPCC. | The Sample_Identifier initially assigned to the surveillance sample must be provided. | Values must match values previously submitted through the SWT Sequence Request Interface. | None |
| Examples | SJCProj02_4001 | 22258468 | SWN-9816 | SQN_Proj_01 |
| Notes | | Submissions without a corresponding surveillance submission will be listed as orphaned records. Submissions without a corresponding surveillance submission must complete all fields in the metadata template. | This field provides an additional layer of tracking if multiple viruses were isolated from the same surveillance sample. | If you require multiple samples to be sequenced by the same sequencing core, you may provide a Sequencing_Study_Identifier. All future submissions linked to the same Sequencing_Study_Identifier will automatically be sent to the same sequencing core. If you do not have this requirement, enter NA. |
| Dependent Fields | | | | |
| Validation | Project_Identifier should be a valid project identifier. | Validate field length | Validate field length | Validate field length |
| Message Code | Error_9_PROJECT_NOT_FOUND | Error_70_INVALID_FIELD_LENGTH | Error_70_INVALID_FIELD_LENGTH | Error_70_INVALID_FIELD_LENGTH |
| Validation | Project_Identifier exists but user does not have permission to access or edit the project. | | | |
| Message Code | Error_4_DENIED_USER_ACCESS | | | |
| Validation | | | | |
| Message Code | | | | |
| Validation | | | | |
| Message Code | | | | |

| |
|--------------|
| Validation |
| Message Code |
| Validation |
| Message Code |
| Validation |
| Message Code |
| Validation |
| Message Code |

| | Strain_Name | Influenza_Subtype | Host_Species |
|-------------------------|--|---|--|
| Input Type | Text Field | Text Field | Text Field |
| Definition | The WHO strain name of the virus that was isolated from the sample | The influenza subtype of the sample to be sequenced | Full scientific name of host genus and species, without abbreviations, as listed in the DPCC Species Dictionary |
| Format | Antigenic Type/Host of Origin/Geographical Origin/Strain Number/Year of Isolation (Subtype) Maximum length: 150 characters | Text Maximum length: 6 characters | Text Maximum length: 100 characters |
| Value List | Text U NA | Text | DPCC Species Dictionary ENV U |
| Curation | Must only be provided when there is a viral isolate, otherwise use U for samples that are PCR positive but there is no viral isolate. WHO strain 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)) | H and N are case-sensitive and must be entered in all-caps. If only one of the subtypes is known, use the form H5Nx or HxN1. | If the entry is not ENV or U, the host species name is validated against the DPCC Species Dictionary. |
| Examples | 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 (H3N2v), A/duck/Alberta/35/1976 (H1N1), B/Hong Kong/432/2014, or C/Texas/19876/2011 | H11N10, H5Nx, or mixed | Sus scrofa, or Anser albifrons |
| Notes | Reference: http://www.cdc.gov/flu/about/viruses/types.htm (HxNx) can be used in cases where a partial subtype has been determined (e.g., H5Nx). If there are mixed subtypes contained within a sample use A/chicken/Fujian/4/2002 (mixed) for Strain_Name | If there are mixed subtypes contained within a sample, use mixed | Please reference the DPCC Species Dictionary for allowed values. Use ENV(ironment) for samples taken from the environment (e.g., a water sample or feces picked up from the beach) when the source species is not known. If genus is known but species is unknown, then use the genus followed by 'sp.' (for example domestic duck would be reported as 'Anas sp.'). If both the genus and species are unknown, then use the scientific family name, without abbreviation (e.g., Charadriidae). Enter U for all other cases. |
| Dependent Fields | | | |
| Validation | Validate field length | Validate Influenza_Subtype name. Influenza_Subtype should follow pattern HxNx, x can be valid segment number or x if unknown segment. | Validate against DPCC Species Dictionary |
| Message Code | Error_70_INVALID_FIELD_LENGTH | Error_1_INVALID_VALUE | Error_1_INVALID_VALUE |
| 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. | Validate field length | |
| Message Code | QC_136_STRAIN_NAME_NOT_VALIDATED | Error_70_INVALID_FIELD_LENGTH | |
| 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 |

| | Lab_Host | Passage_History | Pathogenicity | Extract_Material |
|-------------------------|---|--|--|--|
| Input Type | Text Field | Text Field | Text Field | Text Field |
| Definition | Description of the live animal host used for passaging the virus without additional information | Description of the passage history of the virus strain | The pathogenicity of the virus strain | Type of material submitted for sequencing |
| Format | Text Maximum length: 50 characters | Text Maximum length: 50 characters | Text Maximum length: 4 characters | Text Maximum length: 30 characters |
| Value List | Text NA | Text ORI X | HPAI LPAI U NA | cDNA DNA PCR fragment Viral RNA |
| Curation | None | 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 | The entry must be one and only one member of the Value List. If the virus is not an avian influenza strain, enter NA. | The entry must be one and only one member of the Value List. |
| Examples | ferret, or mouse | E2 | LPAI | Viral RNA |
| Notes | This field is strictly reserved for the common or scientific name of live animals associated with generation of the sequence. | 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 will be derived directly from the original sample material. Additional information relating to viral passaging can be entered in the Comments field. | HPAI = Highly pathogenic avian influenza LPAI = Low pathogenic avian influenza | |
| Dependent Fields | | | | |
| Validation | Validate field length | Validate field length | Validate field length | Field value should be one of valid values as in list. |
| Message Code | Error_70_INVALID_FIELD_LENGTH | Error_70_INVALID_FIELD_LENGTH | Error_70_INVALID_FIELD_LENGTH | Error_1_INVALID_VALUE |
| Validation | | | Field value should be one of valid values as in list. | Validate field length |
| Message Code | | | Error_1_INVALID_VALUE | Error_70_INVALID_FIELD_LENGTH |
| Validation | | | | |
| Message Code | | | | |
| Validation | | | | |
| Message Code | | | | |

| |
|--------------|
| Validation |
| Message Code |
| Validation |
| Message Code |
| Validation |
| Message Code |
| Validation |
| Message Code |

| | Volume | Concentration | Concentration_Determined_By | Sequencing_Technology |
|-------------------------|---|--|---|---|
| Input Type | Number Field | Number Field | Text Field | Text Field |
| Definition | Volume of the extract in microliters (µl) | Concentration of the extract in nanograms per microliter (ng/µl) | Type of instrument used to determine the concentration or titer | The name of the sequencing technology used to obtain the submitted sequences |
| Format | Number Maximum length: 10 characters | Number Maximum length: 10 characters | Text Maximum length: 30 characters | Text Maximum length: 250 characters |
| Value List | None | None | Nanodrop Qbit Spectrometry | Illumina MiSeq Illumina HiSeq Illumina NextSeq Life Tech Ion-Proton Oxford MiniON PacBio RS No Preference |
| Curation | The entry must be a single numerical value. | The entry must be a single numerical value. | The entry must be one and only one member of the Value List. | Enter No Preference if a specific sequencing technology is not needed for this request. The entry must be one or more members of the Value List. |
| Examples | 250.5 | 25.5 | Nanodrop | Life Tech Ion-Proton |
| Notes | | | | If more than one Sequencing_Technology is requested, comma-separate individual technologies. |
| Dependent Fields | | | | |
| Validation | Validate field length | Validate field length | Validate field length | Field value should be one of valid values as in list. |
| Message Code | Error_70_INVALID_FIELD_LENGTH | Error_70_INVALID_FIELD_LENGTH | Error_70_INVALID_FIELD_LENGTH | Error_1_INVALID_VALUE |
| Validation | Must be a valid number | Must be a valid number | | Field length including values from the Value List and free text following 'OTH-' must be less than 250 characters. |
| Message Code | Error_142_INVALID_NUMBER_ONLY | Error_142_INVALID_NUMBER_ONLY | | Error_75_INVALID_FIELD_LENGTH_ONLY |
| Validation | | | | |
| Message Code | | | | |
| Validation | | | | |
| Message Code | | | | |

| |
|--------------|
| Validation |
| Message Code |
| Validation |
| Message Code |
| Validation |
| Message Code |
| Validation |
| Message Code |

| | Analysis_Type | Raw_Sequences | Comments |
|-------------------------|--|--|---|
| Input Type | Text Field | Text Field | Text Field |
| Definition | Types of analyses you wish to receive | Would you like to receive the raw sequence reads? | Text describing anything else of interest related to the submission |
| Format | Text Maximum length: 150 characters | Text Maximum length: 1 character | Text Maximum length: 2000 characters |
| Value List | Amino acid analysis Annotation Consensus Sequence Geographic Mapping Lineage tracking Phylogenetics Variant Analysis | Y N | Text NA |
| Curation | The entry must be one or more members of the Value List. | The entry must be one and only one member of the Value List. | None |
| Examples | Consensus Sequence,Phylogenetics, or Consensus Sequence | Y | NA |
| Notes | Consensus Sequence is required for all submissions. If more than one Analysis_Type is requested, comma-separated individual values. | | If there are no comments, enter NA. |
| Dependent Fields | | | |
| Validation | 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 |
| Validation | | Field value should be one of valid values as in list. | |
| Message Code | | Error_1_INVALID_VALUE | |
| Validation | | | |
| Message Code | | | |
| Validation | | | |
| Message Code | | | |

| |
|--------------|
| Validation |
| Message Code |
| Validation |
| Message Code |
| Validation |
| Message Code |
| Validation |
| Message Code |