

DPCC Data Standard Reference for SRA Analysis v1.1

| | Project_Identifier | Contributing_Institution | Embargo_End_Date |
|-------------------------|--|--|--|
| Input Type | Text Field | Text Field | Date 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 | In case an embargo of the information is needed, the date that the information should be released to the public databases by the DPCC |
| Format | Project_Code_XXXX Maximum length: 21 characters | Center three-letter code followed by three digits. Maximum length: 6 characters | DD-Mon-YYYY DD-Mon-YY NA Maximum length: 11 characters |
| Value List | None | None | Date NA |
| 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. | 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. |
| Examples | SJCProj02_4001 | SJC101 | 3-Mar-2011, 03-Mar-2011, 03-MAR-2011, 3-MAR-11, or NA |
| Notes | | A different Contributing_Institution may have performed the original collection of the surveillance sample that was analyzed. | 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 45 days after submission. |
| Dependent Fields | | | |
| Validation | Project_Identifier should be a valid project identifier. | Must be an Institution Code value registered with the DPCC | Embargo_End_Date must be exact date and cannot be a past date. |
| Message Code | Error_9_PROJECT_NOT_FOUND | Error_1_INVALID_VALUE | Error_7_INVALID_EMBARGO_DATE |
| Validation | Project_Identifier exists but user does not have permission to access or edit the project. | | Embargo_End_Date cannot be more than 45 days after submission. |
| Message Code | Error_4_DENIED_USER_ACCESS | | Error_99_EMBARGO_DATEError_45D |
| Validation | | | |
| Message Code | | | |

| | SRA_Study_Accession_Numbers | Analysis_Title |
|-------------------------|--|---|
| Input Type | Text Field | Text Field |
| Definition | Identifier assigned to an SRA Study. Allows the analysis files to be linked to the correct SRA submission | Title of the analysis object which will be displayed in short form in the Analysis browser and in database search results |
| Format | Text Maximum length: 200 characters | Text Maximum length: 100 characters |
| Value List | Text NA | Text |
| Curation | None | None |
| Examples | SRA00057 | Quasispecies analysis of Influenza A virus (A/AA/Huston/1945(H1N1)) |
| Notes | If the analysis references sequences previously submitted to the SRA, provide the SRA_Study_Accession_Numbers. Enter NA if the SRA Run and SRA Analysis information are being submitted concurrently. | |
| Dependent Fields | | |
| Validation | Validate field length | Validate field length |
| Message Code | Error_70_INVALID_FIELD_LENGTH | Error_70_INVALID_FIELD_LENGTH |
| Validation | If not NA, must be a valid SRA Study accession. | |
| Message Code | QC_106_INVALID_SRA_STUDY | |
| Validation | | |
| Message Code | | |

| | Analysis_Type |
|-------------------------|---|
| Input Type | Text Field |
| Definition | Type of analysis performed |
| Format | Text Maximum length: 21 characters |
| Value List | Abundance measurement De novo assembly Reference alignment Sequence annotation |
| Curation | The entry must be one and only one member of the Value List. |
| Examples | Sequence annotation |
| Notes | |
| Dependent Fields | |
| Validation | Field value should be one of valid values as in list. |
| Message Code | Error_1_INVALID_VALUE |
| Validation | |
| Message Code | |
| Validation | |
| Message Code | |

| | Analysis_Program | Analysis_Program_Version |
|-------------------------|--|--------------------------------------|
| Input Type | Text Field | Text Field |
| Definition | Identify the processing tools used to produce the analysis | The analysis program version number |
| Format | Text Maximum length: 50 characters | Text Maximum length: 5 characters |
| Value List | Text NA | Number NA |
| Curation | None | None |
| Examples | Euler-SR | 2.54 |
| Notes | | |
| Dependent Fields | | |
| Validation | Validate field length | Validate field length |
| Message Code | Error_70_INVALID_FIELD_LENGTH | Error_70_INVALID_FIELD_LENGTH |
| Validation | | |
| Message Code | | |
| Validation | | |
| Message Code | | |

| | File_Name | Comments |
|-------------------------|--|---|
| Input Type | Text Field | Text Field |
| Definition | The name(s) of the analysis data file(s). | Text describing anything else of interest related to the submission |
| Format | Text Maximum length: 2000 characters | Text Maximum length: 2000 characters |
| Value List | Text | Text NA |
| Curation | The file or files included in the submission must have one of the following extensions: ACE BED CSV GFF MAF TSV VCF WIG | None |
| Examples | analysis1_15_11_15.vcf | NA |
| Notes | The entry must be the full file name, with extension (e.g., sequence_data.vcf). Users can submit multiple comma-separated files. File name should be unique in the zip file. ACE = Multiple alignment file format BED = Tab-delimited assembly format CSV = Comma-separated values GFF = General feature format MAF = Mutation annotation format TSV = Tab-separated value format VCF = Variant call format WIG = Wiggle format | If there are no comments, enter NA. |
| Dependent Fields | | |
| Validation | Validate field length | Validate field length |
| Message Code | Error_70_INVALID_FIELD_LENGTH | Error_70_INVALID_FIELD_LENGTH |
| Validation | File is missing or is empty | |
| Message Code | Error_137_FILE_MISSING_OR_EMPTY | |
| Validation | Validate file extension | |
| Message Code | Error_133_SRA_ANALYSIS_UNSUPPORTED_FILE_EXT | |

| | Internal Field Only |
|-------------------------|--|
| | SRA_Accession_Numbers |
| Input Type | Text Field |
| Definition | Internal Field Only Accession numbers associated with the SRA Analysis submission. |
| Format | Text Maximum length: 200 characters |
| Value List | Text |
| Curation | None |
| Examples | |
| Notes | |
| Dependent Fields | |
| Validation | |
| Message Code | |
| Validation | |
| Message Code | |
| Validation | |
| Message Code | |