

DPCC Data Standard Reference for BioProject Registration v1.3			
	Project_Identifier	Contributing_Institution	Additional_Umbrella_Project
<b>Input Type</b>	Text Field	Text Field	Text Field
<b>Definition</b>	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 or more of the CEIRS institution codes assigned by NIAID	The accession number(s) of additional CEIRS or non-CEIRS Umbrella Project(s) associated with the BioProject.
<b>Format</b>	Project_Code_XXXX  Maximum length: 21 characters	Center three-letter code followed by three digits.  Maximum length: 70 characters	BioProject ID  Maximum length: 45 characters
<b>Value List</b>	None	None	Text NA
<b>Curation</b>	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.	Must be one or more valid NCBI umbrella project accession numbers.
<b>Examples</b>	SJCProj02_4001	SJC101	PRJNA46305
<b>Notes</b>	None	Multiple Contributing_Institutions can be entered as comma-separated values.	The BioProject will automatically be associated with the umbrella project of the Contributing_Institution. Enter NA if there are no other associations.  Enter any additional CEIRS or non-CEIRS umbrella projects this BioProject should be associated with as comma-separated values.
<b>Dependent Fields</b>			
<b>Validation</b>	Project_Identifier should be a valid project identifier.	Must be an Institution Code value registered with the DPCC	Must be a valid umbrella project
<b>Message Code</b>	Error_9_PROJECT_NOT_FOUND	Error_1_INVALID_VALUE	Error_104_INVALID_BIOPROJECT
<b>Validation</b>	Project_Identifier exists but user does not have permission to access or edit the project.		
<b>Message Code</b>	Error_4_DENIED_USER_ACCESS		

	BioProject_Owner	BioProject_Owner_Email	BioProject_Title
<b>Input Type</b>	Text Field	Text Field	Text Field
<b>Definition</b>	The name of the primary owner of the BioProject.	The email address of the primary owner of the BioProject.	Provide a brief title for public display.
<b>Format</b>	FirstName LastName FirstName MI. LastName  Maximum length: 50 characters	Text@domain.com  Maximum length: 50 characters	Text  Maximum length: 250 characters
<b>Value List</b>	None	None	None
<b>Curation</b>	FirstName LastName FirstName MI. LastName	The entry must be a valid email address.	None
<b>Examples</b>	Indiana Jones, or John M. Henry	jsmith@email.edu	Infection of quail with avian influenza H5N1 and H5N2 viruses
<b>Notes</b>			
<b>Dependent Fields</b>			
<b>Validation</b>	Validate field length	Validate field length	Validate field length
<b>Message Code</b>	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH
<b>Validation</b>		Must be a valid email address.	
<b>Message Code</b>		Error_114_INVALID_EMAIL	

	BioProject_Description	Consortium	Project_Data_Type
<b>Input Type</b>	Text Field	Text Field	Text Field
<b>Definition</b>	Provide a description of the project goals and relevance.	The name of any additional consortium or other group that is involved in the sequencing. The default entry in absence of a Consortium will be "NIAID Centers of Excellence for Influenza Research and Surveillance (CEIRS)".	General label indicating the project goal
<b>Format</b>	Text Maximum length: 1000 characters	Text Maximum length: 250 characters	Text Maximum length: 500 characters
<b>Value List</b>	None	Text NA	Assembly Clone ends Epigenomics Exome Genome sequencing Genome sequencing and assembly Map Metagenome Metagenomic assembly Phenotype or Genotype Proteome Random survey Raw sequence reads Targeted loci cultured Targeted loci environmental Targeted locus (Loci) Transcriptome or gene expression Variation OTH-
<b>Curation</b>	None	Enter NA if no additional Consortium is applicable	The entry must be one or more members of the Value List.
<b>Examples</b>	Transcriptional profiling was carried out on lung and ileum samples at 1dpi and 3dpi from quail infected with either low pathogenic (H5N2) or highly pathogenic (H5N1) avian influenza. Infected birds were compared to control birds at each time point.	National Veterinary Services Laboratories, USDA Swine Surveillance, or NA	Transcriptosome or gene expression
<b>Notes</b>		Enter a non-CEIRS consortium. Do not list departmental or institution affiliations. If more than one consortium or group was involved, semicolon-separate individual consortia.	
<b>Dependent Fields</b>			
<b>Validation</b>	Validate field length	Validate field length	Validate field length
<b>Message Code</b>	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH	Error_75_INVALID_FIELD_LENGTH_OTH
<b>Validation</b>			Field value should be one of valid values as in list.
<b>Message Code</b>			Error_1_INVALID_VALUE

	Sample Scope	Relevance	Organism
<b>Input Type</b>	Text Field	Text Field	Text Field
<b>Definition</b>	Indicates the scope and purity of the biological sample used for the study	The primary general relevance of the study	The most descriptive organism name for the study
<b>Format</b>	Text  Maximum length: 30 characters	Text  Maximum length: 30 characters	Text  Maximum length: 100 characters
<b>Value List</b>	Environment Monoisolate Multiisolate Multispecies Synthetic OTH-	Agricultural Environmental Evolution Industrial Medical Model organism OTH-	None
<b>Curation</b>	The entry must be one and only one member of the Value List.	The entry must be one and only one member of the Value List.	None
<b>Examples</b>	Monoisolate	Medical	Influenza A virus
<b>Notes</b>	Select Multiisolate when the research goal is to compare multiple individuals or strains of the same species, e.g., a Variation or Genome sequencing and assembly Project_Data_Type.  Select Multispecies when different species are being examined.  Select Monoisolate if the research goal is to make a single genome or transcriptome assembly, even if more than one individual was the source of the genetic material.		
<b>Dependent Fields</b>			
<b>Validation</b>	Field value should be one of valid values as in list. NOTE: User can enter other value by prefixing 'OTH-'	Field value should be one of valid values as in list. NOTE: User can enter other value by prefixing 'OTH-'	Validate field length
<b>Message Code</b>	Error_1_INVALID_VALUE	Error_1_INVALID_VALUE	Error_70_INVALID_FIELD_LENGTH
<b>Validation</b>	Validate field length	Validate field length	
<b>Message Code</b>	Error_75_INVALID_FIELD_LENGTH_OTH	Error_70_INVALID_FIELD_LENGTH	

	Internal Field Only		
	Publication PMID	Comments	BioProject Accession Number
<b>Input Type</b>	Text Field	Text Field	Text Field
<b>Definition</b>	The PubMed Unique Identifier (PMID) for the publication in which the sequence was published	Text describing anything else of interest related to the submission	<b>Internal Field Only</b> Populated with the BioProject Accession Number after successful registration.
<b>Format</b>	PMID  Maximum length: 50 characters	Text  Maximum length: 2000 characters	Text  Maximum length: 15 characters
<b>Value List</b>	Text NA	Text NA	Text
<b>Curation</b>	The entry must be a valid PMID number: 7 or 8-digit number with no leading zeros. <a href="http://www.nlm.nih.gov/bsd/mms/medlineelements.html#pmid">http://www.nlm.nih.gov/bsd/mms/medlineelements.html#pmid</a>	None	None
<b>Examples</b>	19465683, or NA	NA	PRJNA561570
<b>Notes</b>	If BioProject is linked to multiple publications, comma-separate individual PMIDs.  Enter NA if BioProject is not related to a publication.	If there are no comments, enter NA.	
<b>Dependent Fields</b>			
<b>Validation</b>	Validate field length	Validate field length	Validate field length
<b>Message Code</b>	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH
<b>Validation</b>	Must be a valid PMID		
<b>Message Code</b>	Error_96_INVALID_PMIID		

	Internal Field Only	Internal Field Only	Internal Field Only
	Locus_Tag_Prefix	Grant_ID	Grant_Title
<b>Input Type</b>	Text Field	Text Field	Text Field
<b>Definition</b>	<b>Internal Field Only</b> Populated with the BioProject locus tag identifier after successful registration.	<b>Internal Field Only</b> Populated with the Center-specific contract number after successful registration.	<b>Internal Field Only</b> Populated with the Center-specific Center name after successful registration.
<b>Format</b>	Text  Maximum length: 100 characters	Text  Maximum length: 50 characters	Text  Maximum length: 250 characters
<b>Value List</b>	Text	HHSN266200700005C HHSN266200700006C HHSN266200700007C HHSN266200700008C HHSN266200700009C HHSN2662007000010C HHSN272201400004C HHSN272201400005C HHSN272201400006C HHSN272201400007C HHSN272201400008C HHSN272201400026C	Center for Rapid Influenza Surveillance and Research (CRISAR) Center for Research on Influenza Pathogenesis (CRIP) Data Processing and Coordinating Center (DPCC) Emory-UGA Center of Excellence for Influenza Research and Surveillance (CEIRS) Johns Hopkins Center of Excellence for Influenza Research and Surveillance (CEIRS) Minnesota Center of Excellence for Influenza Research and Surveillance (MCEIRS) New York Influenza Center of Excellence (NYICE) St. Jude Center of Excellence for Influenza Research and Surveillance (CEIRS)
<b>Curation</b>	None	None	None
<b>Examples</b>	FYC12	HHSN272201400008C	Center for Research on Influenza Pathogenesis (CRIP)
<b>Notes</b>			
<b>Dependent Fields</b>			
<b>Validation</b>	Validate field length	Validate field length	Validate field length
<b>Message Code</b>	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH	Error_70_INVALID_FIELD_LENGTH
<b>Validation</b>			
<b>Message Code</b>			