

## Exchange formats to submit NGS data

Project.csv	<b>mandatory file</b>	
Contact.csv	<b>mandatory file</b>	
Reference.csv	optional file	It should be fill if you refer to a reference in « Project.csv »
Experiment.csv	<b>mandatory file</b>	
Hardware.csv	<b>mandatory file</b>	
Sample.csv	<b>mandatory file</b>	
Library.csv	optional file	It should be fill if you refer to a library in « Sample.csv »
Run.csv	<b>mandatory file</b>	
SubRun.csv	<b>mandatory file</b>	
SubmitFiles.csv	optional file	It should be fill if you refer to a result file in « Analysis.csv » or « SubRun.csv »
Analysis.csv	optional file	It should be fill if you refer to an analysis in other files
Software.csv	optional file	It should be fill if you fill « Analysis.csv »

**Project.csv:** information about the global project

**Detail of the fields :**

Field name	Mandatory?	Description of the field	Regex / Authorised values	Length max
<b>ProjectCode:</b>	Y	Project code (explicit). Example: <b>GrapeReSeq</b>	This field had to be filled with a string composed only by letters and numbers (accentuate letters, space character... are not allowed). This field is used as key ( <a href="#">see more below</a> )	30 characters.
<b>ProjectName:</b>	Y	Project name. Example: <b>GrapeReSeq</b>		100 characters.
<b>Title:</b>	Y	Project title. Example: <b>Large scale re-sequencing in the Vitis genus for identification of resistance genes, SNP discovery and high throughput genotyping</b>		255 characters.
<b>ProjectCoordinator:</b>	Y	Coordinator contact number. This field can be multiple. If you have several ProjectCoordinator, separate them by ', '.  Example(s): <b>2</b> <b>2, 16</b>	This field is used as key ( <a href="#">see more below</a> )	
<b>BioinformaticsContact:</b>	Y	Bioinformatics contact number. This field can be multiple. If you have several BioinformaticsContact, separate them by ', '.  Example(s): <b>10</b> <b>7, 15, 20</b>	This field is used as key ( <a href="#">see more below</a> )	
<b>PartnerContact:</b>	N	Partner contact number. This field can be multiple. If you have several PartnerContact, separate them by ', '.  Example(s): <b>13</b> <b>20, 23</b>	This field is used as key ( <a href="#">see more below</a> )	
<b>BeginningDate:</b>	N	Date of the beginning of the project. Example: <b>01/04/2009</b>	This field had to be filled with a date format. Convention is dd/MM/yyyy.	
<b>Comment:</b>	N	Comments on the project.		1000 characters.

<b>Affiliation:</b>	Y	Affiliation of the project. Example: <b>genoplante</b>	This field is used as key ( <a href="#">see more below</a> )	
<b>ReferenceNumber:</b>	N	Reference number. This field can be multiple. If you have several References, separate them by ', '.  Example(s): <b>2</b> <b>1, 4</b>	This field is used as key ( <a href="#">see more below</a> )	

Keys informations:

- Values of ProjectCoordinator, BioinformaticsContact and PartnerContact have to be found in [Contact.csv](#) (which has been submitted in the same batch).
- Values of ReferenceNumber have to be found in [Reference.csv](#) (which has been submitted in the same batch).
- Values of Affiliation have to be found in the [database](#). If you need other value than the ones already in the DB, contact us.
- Values of ProjectCode defined keys for other submission file ([Experiment.csv](#) and [Sample.csv](#)) so their values have to be found in these files with a strict respect of the lower/upper case and hyphen/underscore.
  - ProjectCode can be filled only by one value.

Some of fields have to be unique in the whole file: ProjectCode / ProjectName

**Contact.csv:** information concerning the different persons involved in a project

**Detail of the fields :**

Field name	Mandatory?	Description of the field	Regex / Authorised values	Length max
<b>ContactNumber:</b>	y	Contact number (positive integer). Example: 15	This field is used as key ( <a href="#">see more below</a> )	
<b>ContactName:</b>	y	Contact last name, in upper case. Example: DUPONT	This field had to be filled with a word/name.	100 characters.
<b>FirstName:</b>	y	Contact first name, first letter in upper case. Example: Jacques	This field had to be filled with a word/name.	50 characters.
<b>Fax:</b>	n	Following international fax number conventions. Example: +33 (0) 1.60.87.37.99	This field had to be filled with a phone format.	40 characters.
<b>Telephone:</b>	n	Following international phone number conventions. Example: +33 (0) 1.60.87.37.00	This field had to be filled with a phone format.	40 characters.
<b>Email:</b>	y	Electronic mail. Example: Jacques.Dupont@email.com	This field had to be filled with an email address.	255 characters.
<b>Laboratory:</b>	n	Laboratory name. Example: Plant genetic laboratory		100 characters.
<b>Institution:</b>	y	Institution. Example: INRA		120 characters.
<b>Address:</b>	y	Street address. Example: 2, rue Gaston Crémieux		255 characters.
<b>City:</b>	y	City, in upper case (with cedex if needed). Example: EVRY cedex		60 characters.
<b>State:</b>	n	State. Example: Texas		60 characters.
<b>ZipCode:</b>	y	Zip code. Example: 91057		60 characters.
<b>Country:</b>	y	Country, in upper case. Example: FRANCE		60 characters.
<b>WebSite</b>	n	Website		

Keys informations:

- Values of ContactNumber defined keys for other submission file ([Sample.csv](#) and [Analysis.csv](#)) so their values have to be found in these files.
- ProjectCode can be filled only by one value.

Some of fields have to be unique in the whole file: ContactName / FirstName / Institution

**Reference.csv:** bibliographic references

**Detail of the fields :**

Field name	Mandatory?	Description of the field	Regex / Authorised values	Length max
<b>ReferenceNumber:</b>	y	Reference number (positive integer). Example: 2	This field is used as key ( <a href="#">see more below</a> )	
<b>Authors:</b>	y	Publication title. Example: Edwards D, Murray JA, Smith AG		1000 characters.
<b>Title:</b>	y	Publication title. Example: dbSNP: the NCBI database of genetic variation		600 characters.
<b>Journal:</b>	y	Journal name. Example: Plant physiology		500 characters.
<b>Volume:</b>	y	Journal volume. Example: 13		10 characters.
<b>Subvolume:</b>	n	Journal subvolume.		10 characters.
<b>Pages:</b>	y	Journal pages. Example: 1-13		20 characters.
<b>Year:</b>	y	Year: yyyy. Example: 2000		4 characters.
<b>Medline number:</b>	n	Medline reference.		20 characters.

Keys informations:

- Values of ReferenceNumber defined keys for [Project.csv](#) so their values have to be found in this file.
  - ReferenceNumber can be filled only by one value.

Some of fields have to be unique in the whole file: Title

**Experiment.csv:** information about an experiment.

**Detail of the fields :**

Field name	Mandatory?	Description of the field	Regex / Authorised values	Length max
<b>ExperimentNumber:</b>	y	Experiment number (positive integer). Example: 3	This field is used as key ( <a href="#">see more below</a> )	
<b>ExperimentName:</b>	y	Experiment name. Example: <b>Vitaceae resequencing</b>		100 characters.
<b>StudyType:</b>	y	The aim of the experiment. Example: <b>resequencing</b>	This field is used as key ( <a href="#">see more below</a> )	
<b>SequencingType:</b>	y	Is it single or paired end sequencing?	This field had to be filled with « single end » or « paired end » term.	
<b>SequencingCenter:</b>	y	Sequencing center Example: <b>CNG-EPGV</b>		100 characters.
<b>HardwareNumber:</b>	y	Hardware number. Example: 2	This field is used as key ( <a href="#">see more below</a> )	
<b>SampleNumber:</b>	y	Sample number . This field can be multiple. If you have several samples, separate them by ' , ' .  Examples : 4 5, 9	This field is used as key ( <a href="#">see more below</a> )	
<b>Abstract:</b>	n	Abstract		255 characters.
<b>Description:</b>	n	Description		255 characters.
<b>ProjectCode:</b>	y	Project code Example: <b>GrapeReseq</b>	This field is used as key ( <a href="#">see more below</a> )	

Keys informations:

- Values of HardwareNumber have to be found in [Hardware.csv](#) (which has been submitted in the same batch).
- Values of SampleNumber have to be found in [Sample.csv](#) (which has been submitted in the same batch).
- Values of ProjectCode have to be found in [Project.csv](#) (which has been submitted in the same batch).
- Values of StudyType have to be found in the [database](#). If you need other value than the ones already in the DB, contact us.
- Values of ExperimentNumber defined keys for other submission file ([Sample.csv](#) and [SubRun.csv](#)) so their values have to be found in these files.
  - ExperimentNumber can be filled only by one value.

Some of fields have to be unique in the whole file: ExperimentName

**Hardware.csv:** hardware used for the sequencing.

**Detail of the fields :**

Field name	Mandatory?	Description of the field	Regex / Authorised values	Length max
<b>HardwareNumber:</b>	y	Hardware number (positive integer). Example: 3	This field is used as key ( <a href="#">see more below</a> )	
<b>Name:</b>	y	Hardware name. Example: Illumina		100 characters.
<b>Model:</b>	n	Hardware model. Example: GA II		100 characters.
<b>Make:</b>	n	Hardware maker. Example: Illumina, Inc		255 characters.
<b>Type:</b>	y	Type of hardware. Example(s): Sequencing Hardware	This field is used as key ( <a href="#">see more below</a> )	
<b>Description:</b>	n	Hardware description.		255 characters.

Keys informations:

- Values of Type have to be found in the [database](#). If you need other value than the ones already in the DB, contact us.
- Values of HardwareNumber defined keys for [Experiment.csv](#) so their values have to be found in this file.
  - HardwareNumber can be filled only by one value.

Some of fields have to be unique in the whole file: Name / Model

**Sample.csv:** information on the sample used for the sequencing.

**Detail of the fields :**

Field name	Mandatory?	Description of the field	Regex / Authorised values	Length max
<b>SampleNumber:</b>	y	Sample number (positive integer). Example: 5	This field is used as key ( <a href="#">see more below</a> )	
<b>SampleName:</b>	y	Sample name. Example: PN40024		100 characters.
<b>SampleType:</b>	n	Sample type (gDNA, mRNA, ...). Example: gDNA	This field is used as key ( <a href="#">see more below</a> )	
<b>Tissue:</b>	n	Tissue type.		100 characters.
<b>Taxon:</b>	y	Taxon of the sample. Example: <i>Vitis vinifera</i> L.	This field is used as key ( <a href="#">see more below</a> )	
<b>Accession:</b>	n	Accession of the sample. Example:	This field is used as key ( <a href="#">see more below</a> )	
<b>Lines:</b>	n	Line of the sample. Example: PN40024	This field is used as key ( <a href="#">see more below</a> )	
<b>Description:</b>	n	Description.		255 characters.
<b>LibNumber:</b>	n	Library number. Example: 10	This field is used as key ( <a href="#">see more below</a> )	
<b>ExperimentNumber:</b>	y	Experiment number. This field can be multiple. If you have several experiments, separate them by ', '.  Examples : 3 6, 9, 10	This field is used as key ( <a href="#">see more below</a> )	
<b>ContactNumber:</b>	y	Contact number. Example: 5	This field is used as key ( <a href="#">see more below</a> )	
<b>ProjectCode:</b>	y	Project code. Example: 15	This field is used as key ( <a href="#">see more below</a> )	

Keys informations:

- Values of SampleType, Taxon, Accession and Line have to be found in the [database](#). If you need other value than the ones already in the DB, contact us.  
NB: Be aware that Accession and Line, if fielded, must have the same taxon as Taxon in this file. If not, they won't be inserting in database.



- Values of LibNumber have to be found in [Library.csv](#) (which has been submitted in the same batch).
- Values of ExperimentNumber have to be found in [Experiment.csv](#) (which has been submitted in the same batch).
- Values of ContactNumber have to be found in [Contact.csv](#) (which has been submitted in the same batch).
- Values of ProjectCode have to be found in [Project.csv](#) (which has been submitted in the same batch).
- Values of SampleNumber defined keys for other submission file ([Experiment.csv](#) and [SubRun.csv](#)) so their values have to be found in these files.
  - SampleNumber can be filled only by one value.

Some of fields have to be unique in the whole file: SampleName

**Library.csv:** information on the library.

**Detail of the fields:**

Field name	Mandatory?	Description of the field	Regex / Authorised values	Length max
<b>LibNumber:</b>	y	Library number (positive integer). Example: 10	This field is used as key ( <a href="#">see more below</a> )	
<b>LibName:</b>	y	Library name. Example: PN40024 sample		100 characters.
<b>FragmentLength:</b>	n	Fragment length before sequencing (positive integer in bp). Example: 400		
<b>Strategy:</b>	n	Sequencing technique intended for this library.  Example(s) : PoolClone RNA-Seq		255 characters.
<b>Source:</b>	n	Sequencing materiel source.  Example(s) : Genomic Transcriptomic		255 characters.
<b>Selection:</b>	n	Method used to select and/or enrich the sequenced material.  Example(s) : PCR ChIP		255 characters.
<b>ProtocolDescription:</b>	n	Description.		255 characters.

Keys informations:

- Values of LibNumber defined keys for [Sample.csv](#) so their values have to be found in this file.
  - LibNumber can be filled only by one value.

Some of fields have to be unique in the whole file: LibName

**Run.csv:** information about the run(s).

**Detail of the fields :**

Field name	Mandatory?	Description of the field	Regexp / Authorised values	Length max
<b>RunNumber:</b>	y	Run number (positive integer). Example: 10	This field is used as key ( <a href="#">see more below</a> )	
<b>RunName:</b>	y	Run name. Example: 100331_GA10-EAS1698_0002_61EW4AAXX		100 characters.
<b>RunDate:</b>	n	Following date conventions, production date of the run. Example: 08/06/2010	This field had to be filled with a date format. Convention is dd/mm/yyyy.	
<b>Description:</b>	n	Description.		255 characters.

Keys informations:

- Values of RunNumber defined keys for [SubRun.csv](#) so their values have to be found in this file.
- RunNumber can be filled only by one value.

Some of fields have to be unique in the whole file: RunName

**SubRun.csv:** information about the sub-run(s).

**Detail of the fields :**

Field name	Mandatory?	Description of the field	Regex / Authorised values	Length max
<b>SubRunName:</b>	y	Subrun name (exemple: runName_canalNumber). Example: 100331_GA10-EAS1698_0002_61EW4AAXX_s1 PN40024_run1_s_1	This field is used as key ( <a href="#">see more below</a> )	
<b>RunNumber:</b>	y	Run number. Example: 9	This field is used as key ( <a href="#">see more below</a> )	
<b>SubRunType:</b>	y	Sub-run type. Example: lane	This field is used as key ( <a href="#">see more below</a> )	
<b>ReadsNumber:</b>	n	Reads number. Example: 69067936		positive integer
<b>ReadLength:</b>	n	Average length of the reads. Example: 76		positive integer
<b>Description:</b>	n	Description.		255 characters.
<b>ExperimentNumber:</b>	y	Experiment number. This field can be multiple. If you have several Experiments, separate them by ', '. Example(s): 7, 15, 20	This field is used as key ( <a href="#">see more below</a> )	
<b>SampleNumber:</b>	y	Sample number. Example: 3	This field is used as key ( <a href="#">see more below</a> )	
<b>FileNumber:</b>	n	Experiment files number. This field can be multiple. If you have several Files, separate them by ', '. Example(s): 5 9, 18	This field is used as key ( <a href="#">see more below</a> )	
<b>AnalysisNumber:</b>	n	Analysis number. This field can be multiple. If you have several Analyses, separate them by ', '. Example(s): 9 7, 14	This field is used as key ( <a href="#">see more below</a> )	

Keys informations:

- Values of SubRunType have to be found in the [database](#). If you need other value than the ones already in the DB, contact us.
- Values of RunNumber have to be found in [Run.csv](#) (which has been submitted in the same batch).
- Values of ExperimentNumber have to be found in [Experiment.csv](#) (which has been submitted in the same batch).
- Values of SampleNumber have to be found in [Sample.csv](#) (which has been submitted in the same batch).
- Values of FileName have to be found in [SubmitFiles.csv](#) (which has been submitted in the same batch).
- Values of AnalysisNumber have to be found in [Analysis.csv](#) (which has been submitted in the same batch).

Some of fields have to be unique in the whole file: SubRunName

**SubmiFiles.csv:** information on the submit result files.

**Detail of the fields :**

Field name	Mandatory?	Description of the field	Regex / Authorised values	Length max
<b>FileNumber:</b>	y	File number (positive integer). Example: 23	This field is used as key ( <a href="#">see more below</a> )	
<b>FileName:</b>	y	File name. Example(s) : PN40024_run1_s_1_1.fastq run2_PN40024.bam		100 characters.
<b>RepertoryName:</b>	n	Directory where the file(s) can be found if it is already at URGI.		
<b>Link:</b>	n	URL where the file(s) can be found (file repository) Example : <a href="http://urqi.versailles.inra.fr/Species/Wheat/Sequence-Repository">http://urqi.versailles.inra.fr/Species/Wheat/Sequence-Repository</a>		
<b>FileType:</b>	n	File type. Example(s) : fastq bam	This field is used as key ( <a href="#">see more below</a> )	
<b>FileSize:</b>	n	File size.		positive integer
<b>Description:</b>	n	Description.		255 characters.
<b>Downloadable:</b>	y	Is the file downloadable ?	This field had to be filled with « yes » or « no » term.	

Keys informations:

- Values of FileNumber defined keys for other submission file ([Analysis.csv](#) and [SubRun.csv](#)) so their values have to be found in these files.
  - FileNumber can be filled only by one value.
- Values of FileType have to be found in the [database](#). If you need other value than the ones already in the DB, contact us.

Some of fields have to be unique in the whole file: FileName / FileType

**Analysis.csv:** information on the analysis.

**Detail of the fields :**

Field name	Mandatory?	Description of the field	Regex / Authorised values	Length max
<b>AnalysisNumber:</b>	y	Analysis number (positive integer). Example: 13	This field is used as key ( <a href="#">see more below</a> )	
<b>AnalysisName:</b>	y	Analyse name. Example: <i>Détection de SNP chez PN40024</i>		255 characters.
<b>AnalysisType:</b>	y	Analysis type. Example: <i>mapping software</i>	This field is used as key ( <a href="#">see more below</a> )	
<b>AnalysisDate:</b>	n	Following date conventions, date of the analysis. Example: <i>25/01/2011</i>	This field had to be filled with a date format. Convention is dd/mm/yyyy.	
<b>GenomeVersion:</b>	y	When a reference is used, it is the reference genome version. Otherwise, it is the genome version. Example: <i>Vitis vinifera 12X</i>		100 characters.
<b>SoftwareNumber:</b>	y	Software number. Example: 15	This field is used as key ( <a href="#">see more below</a> )	
<b>CommandLine:</b>	n	Command to launch the analysis. Example: <i>default options</i>		255 characters.
<b>Comments:</b>	n	Comments on the analyse.		1000 characters.
<b>FileNumber:</b>	n	File number. This field can be multiple. If you have several File, separate them by ', '. Example(s): <i>9</i> <i>7, 14</i>	This field is used as key ( <a href="#">see more below</a> )	
<b>ContactNumber:</b>	y	Contact number. Example: 5	This field is used as key ( <a href="#">see more below</a> )	
<b>ProjectCode:</b>	n	Project code. Example: 15	This field is used as key ( <a href="#">see more below</a> )	

Keys informations:

- Values of AnalysisType have to be found in the [database](#). If you need other value than the ones already in the DB, contact us.
- Values of SoftwareNumber have to be found in [Software.csv](#) (which has been submitted in the same batch).
- Values of FileNumber have to be found in [SubmitFiles.csv](#) (which has been submitted in the same batch).
- Values of ContactNumber have to be found in [Contact.csv](#) (which has been submitted in the same batch).

- Values of ProjectCode have to be found in [Project.csv](#) (which has been submitted in the same batch).
- Values of AnalysisNumber defined keys for [SubRun.csv](#) so their values have to be found in this file.
  - AnalysisNumber can be filled only by one value.

Some of fields have to be unique in the whole file: AnalysisName



**Software.csv:** software used for an analysis.

**Detail of the fields :**

Field name	Mandatory?	Description of the field	Regex / Authorised values	Length max
<b>SoftwareNumber:</b>	y	Software number (positive integer). Example: 10	This field is used as key ( <a href="#">see more below</a> )	
<b>Name:</b>	y	Software name. Example: Genalys		100 characters.
<b>Version:</b>	n	Version number for the software. Example: 2.0 beta		20 characters.
<b>Type:</b>	y	Type of software. Example: Multiple sequence analysis/SNP detection	This field is used as key ( <a href="#">see more below</a> )	
<b>URL_Software:</b>	n	URL to locate the software. Example: <a href="http://software.cng.fr/">http://software.cng.fr/</a>		120 characters.
<b>URL_Doc:</b>	n	URL for the documentation. Example: <a href="http://software.cng.fr/docs/doc_frames.html">http://software.cng.fr/docs/doc_frames.html</a>		120 characters.
<b>FTP_Site:</b>	n	Downloading site. Example: <a href="http://software.cng.fr">http://software.cng.fr</a>		120 characters.
<b>Local:</b>	n	Software location on URGI server.		50 characters.
<b>Free:</b>	y	Is it a freeware ?	This field had to be filled with « yes » or « no » term.	
<b>Description:</b>	n	Software description.		255 characters.

Keys informations:

- Values of Type have to be found in the [database](#). If you need other value than the ones already in the DB, contact us.
- Values of SoftwareNumber defined keys for [Analysis.csv](#) so their values have to be found in this file.
  - SoftwareNumber can be filled only by one value.

Some of fields have to be unique in the whole file: Name / Version