

MIRRI Information System (MIRRI-IS)

How to compile the sheets

Introduction

This is a short guide on how to compile the MS Excel file for the submission of catalogue information to MIRRI for inclusion in the MIRRI Information System (MIRRI-IS). It complements the document “Guidelines for catalog uploading” and is a guide to the compilation of the Excel template file “MIRRI-IS_dataset_template_v20200601.xlsx”.

The template file is not intended to be filled in manually, of course. In principle, it could be compiled manually, but it is foreseen that many culture collections (CCs) will create the file programmatically, i.e. they will use some software to create it. However, we strongly suggest trying to compile the template file for some strains of the collection in order to achieve a better understanding of the desired format for submission of data, of allowed values for each data field and of all interrelations between data fields.

In the following paragraph, we suggest how to perform this exercise.

Sheets of the template

The template file includes various sheets having different aims and use. CC must understand the differences in order to appropriately fill them. Note that a light red background highlights all mandatory data. Specifically, the template files includes:

- Fixed reference sheets: these sheets include the complete list of allowed values for some of the data fields. They are fixed, i.e. cannot be changed by the CC, and usually hidden. As an example, consider the sheets for organism types and for forms of supply.
- Extendable reference sheets: these sheets include partial lists of allowed values for some of the data fields and can be updated by the CC in order to include some specific values for their needs. As an example, consider the sheets on sexual states and on markers that are not exhaustive.
- Collection specific reference sheets: these are aimed to include reference lists for each CC and must be filled in by the CC before inserting data in the main sheets. As an example, consider the sheets on growth media used by the CC, geographic origin of strains held by the CC and literature related to the CC strains.
- Catalogue sheets: these sheets are aimed to include the catalogue data, i.e. specific information on the strains included in the catalogue of the CC. These are the ‘Strains’ and ‘Genomic information’ sheets.

The following table report all sheets, their type and the relation to other sheets.

Sheet	Data	Sheet type	Needed by
Resource types	List of organism types	Fixed reference, hidden	Strains
Forms of supply	List of forms of supply	Fixed reference, hidden	Strains
Ploidy	List of allowed values for ploidy	Fixed reference, hidden	Strains
Growth media	List of growth media used by the CC	Collection specific reference	Strains
Geographic origin	List of origin localities for strains in the CC	Collection specific reference	Strains
Literature	List of publications for the strains in the CC, excluded all publications having a PMID or a DOI.	Collection specific reference	Strains
Sexual state	List of allowed values for the 'Sexual state' field in the 'Strains' sheet	Extendable reference	Strains
Strains	Data related to strains in the CC	Catalogue	Genomic information
Ontobiotope	List of ontobiotope terms for the relative field in the 'Strains' sheet	Fixed reference	Strains
Markers	List of marker names for the related field in the 'Genomic information' sheet	Extendable reference	Genomic information
Genomic information	Data related to sequences for the strains in the CC	Catalogue	-

Compilation steps

Due to the interrelation between sheets, for a proper compilation of the CC specific file the following steps should be performed in order.

1. Make a copy

This is needed to preserve the original template file and keep it as a reference.

There is not a standard naming convention, but we suggest that the name of the new file includes both the acronym of the collection and the date.

2. Complete the collection specific reference sheets

Collection specific reference sheets must be compiled before inserting data in the 'Strains' sheet because they are used as a list of allowed values for some of its data fields.

Such sheets are 'Growth media', 'Geographic origin' and 'Literature'. Read carefully the guidelines on how to insert your data into them. These sheets include some fictitious data meant to support the appropriate compilation of the sheet and that can be removed. Note that all growth media and origin localities needed to describe the strains must be included in the related sheets. Instead, only publications missing both PMID and DOI should be included. This is because PMIDs and DOIs can be directly included in the two literature related fields of the 'Strains' sheet.

Note that a light red background highlights all mandatory data.

3. Check and complete the extendable reference sheets

The extendable reference sheets include some of the possible values for the related data fields in the 'Strains' or 'Genomic information' sheets. CCs must check their contents and add all missing values that are needed for their strain descriptions. There are only two such sheets, 'Sexual state' and 'Markers'.

Note that a light red background highlights all mandatory data.

4. Complete the 'Strains' sheet

Once all reference sheets have been compiled, the 'Strains' sheet can be filled in. This is because some of its data fields take the list of allowed values from the reference sheets. On its turn, the 'Strains' sheet must be filled in before the 'Genomic information' sheet because the latter includes a data field for strain accession numbers that takes as allowed values the list of accession numbers of the 'Strains' sheet.

Note that a light red background highlights all mandatory data.

5. Complete the 'Genomic information' sheet

Note that a light red background highlights all mandatory data. In this case, the INSDC accession number should usually be submitted, but when it is not available, the sequence can be submitted instead. In any case, at least one of these two fields must be filled in.