MIRRI Information System (MIRRI-IS)

MIRRI-IS dataset

(Version 5.1, delivered on March 24th, 2023)

Index of contents

Name	MIRRI Accession number
Short name	mirriAccessionNumber
Description	Unique identifier of the strain in the MIRRI-IS. It will be created on the first submission of a strain in the MIRRI-IS in a one-to-one connection with the Accession number of the strain in the CC. It is meant as a reference within the MIRRI-IS and as a unique reference for interoperability with other Life Science tools. This information will be returned to CCs in association with the relative strain accession number. CCs are invited to include it in their catalogue and return it to MIRRI-IS at every following submission.
Syntax	The MIRRI Accession number will be composed by the 'MIRRI' prefix followed by a numeric code of seven digits
Values	On first submission, this field should not be compiled, since it will be automatically assigned by the MIRRI-IS. At following submissions, the accession number returned by MIRRI-IS should be included.
Validation	When missing, check whether the strain accession number was already included. If not, a new value will be created and returned to the collection. If the strain accession number was already submitted in the past, MIRRI-IS will retrieve the related values and assign it to the strain again. When a value is submitted by the collection, MIRRI-IS will check that the correct syntax is used and that the MIRRI and the strain accession numbers are properly associated. If not, inform the collection.
Examples	MIRRI1220234

Name	Accession number (MANDATORY)
Short name	accessionNumber
Description	Unique identifier of the strain in the CC.
Syntax	It should include the CC acronym followed by a space character and a number or code. The code should not include spaces. In case the CC changes the accession number of the strain, the previous number should be included in the "Other culture collection numbers" field.
Values	Free text, according to defined syntax.
Validation	Check that the correct syntax is used.
Examples	LMG 25 CBS 1546.1 TUCC00000110

Name	Other culture collection numbers
Short name	otherCollectionNumbers
Description	Accession numbers of the same strain in other CCs, when known.

Syntax	Accession numbers formatted as above specified and separated by a semicolon character. Should not include accession numbers that do not follow the relative syntax. As an exception, Herbarium numbers can be included here.
Values	Free text, according to defined syntax.
Validation	For accession numbers of strains of CCs available in MIRRI-IS: • control that taxon names or synonyms are identical and if the assigned name is not the current name warn the original CC; • if the name is incorrect warn all the CCs having the strain.
Examples	CBS 316.51; NRRL 1944; QM 191; MUCL 9645

Name	**Restrictions on use (MANDATORY)**
Short name	useRestrictions
Description	Report if the strain can be used for commercial development or not.
Syntax	One of the allowed values.
Values	One of the following values: 1 (no known restrictions apply), 2 (only for non-commercial purposes), 3 (for commercial development a special agreement is requested).
Validation	Check that one and only one of the allowed values is used. Report errors to the CC.
Examples	1

Name	**Nagoya protocol restrictions and compliance conditions (MANDATORY)**
Short name	nagoyaConditions
Description	Situation of the strain in relation to the Nagoya protocol.
Syntax	One of the allowed values.
Values	One of the following: 1 ("No known restrictions under the Nagoya protocol"), 2 ("Documents providing proof of legal access and terms of use available at the collection"), 3 ("Strain probably in scope, please contact the culture collection").
Validation	Check that one and only one of the allowed values is used. Report errors to the CC.
Examples	1

Name	**ABS related files**
Short name	absFile
Description	Uniform Resource Locator (URL) of the Internationally Recognized Certificates of Compliance (IRCC) providing evidence that the strain was accessed in accordance with Prior Informed Consent (PIC) and Mutually Agreed Terms (MAT).
Syntax	The field must include one or more titles and properly formatted URLs (including a scheme, such as http, https or ftp, a hostname, possibly a path, and a file name). For details see https://url.spec.whatwg.org/#urls . Titles can include spaces and will be shown to users instead of the URLs.

	When importing multiple links, provide the first title followed by a semi-colon and then the corresponding URL. For the following links, add a semi-colon before the title. For example: First title;URL1;Second title;URL2;Third title;URL3 When only one URL is provided, the title may be omitted. In this case, the URL will be shown in clear to users.
Values	Free texts for title(s) followed by a semi-colon and a valid and complete URLs.
Validation	Check URLs. Report errors to the CC.
Examples	ABS file;http://www.domain.tld/path/abs_file.pdf

Name	MTA file
Short name	mtaFile
Description	Strain specific Material Transfer Agreement (MTA) document
Syntax	Working URL of the strain specific MTA document, if any. The MTA document must be made available and accessible on-line. To this aim, CCs must provide the URL. Links to MTA files will then be created by MIRRI-IS according to the user requests.
Values	Valid URL.
Validation	Check that a file with the given URL is available on-line. Report errors to the CC.
Examples	http://www.domain.tld/path/mta_strainA.pdf

Name	Strain from a Registered Collection
Short name	registeredCollection
Description	Strain included in the registered CC according to the EU Regulation 511/2014. Unregistered CCs can omit this information.
Syntax	One of the allowed values.
Values	One of the following values: 1 (for No), 2 (for Yes)
Validation	Check that one and only one of the allowed values is used. Report errors to the CC.
Examples	2

Name	Risk Group (MANDATORY)
Short name	riskGroup
Description	Risk group according to EU Directive 2000/54/EC and its amendments and corrections.
Syntax	One of the allowed values.
Values	Allowed values: 1, 2, 3, 4.
Validation	Check that one and only one of the allowed values is used. Report errors to the CC.
Examples	3

Name	Dual use
Short name	dualUse
Description	Specify whether the strain has the potential for a harmful use according to <u>EU Council Regulation 2000/1334/CE</u> and its amendments and corrections.
Syntax	One of the allowed values.
Values	One of the following values: 1 (for No), 2 (for Yes).
Validation	Check that one and only one of the allowed values is used. Report errors to the CC.
Examples	2

Name	Quarantine in Europe
Short name	euQuarantine
Description	Specify whether the strain is subject to quarantine according to <u>European Directive 2000/29/CE</u> and its amendments and corrections. The list of quarantine organisms is available in the <u>Commission Implementing Regulation (EU) 2019/2072</u>.
Syntax	One of the allowed values.
Values	One of the following values: 1 (for No), 2 (for Yes)
Validation	Check that one and only one of the allowed values is used. Report errors to the CC.
Examples	2

Name	Organism type (MANDATORY)
Short name	organismType
Description	The type of the resource.
Syntax	One of the allowed values. Alternatively, in special cases, both Filamentous Fungi and Yeast can be specified, separated by a ";".
Values	One of the following terms: Algae, Archaea, Bacteria, Cyanobacteria, Filamentous Fungi, Yeast, Microalgae.
Validation	Check that one and only one of the allowed values is used, but for the special case above. Report errors to the CC.
Examples	Archaea

Name	Taxon name (MANDATORY)
Short name	speciesName
Description	Taxon name including genus, species and variant names, as taken from the authoritative nomenclature reference: Mycobank for fungi and yeasts, the Prokaryotic Nomenclature Up-to-date for bacteria and archaea, and AlgaeBase for algae, microalgae and cyanobacteria.
Syntax	According to the appropriate nomenclature. For Archaea, Bacteria, Filamentous Fungi and Yeasts, genus name followed by species name and by the subspecies and variant names, when appropriate. The

	subspecies name must be preceded by "subsp.". The variant name must be preceded by "var.". When the species name is not available, do not include "sp.". When the genus name is not available, specify the family name instead.
	In order to cope with delays in nomenclature updates, the most updated taxon name can be used, even when it is missing from the current version of the reference. In this case, a remark must be included in the 'Comment on taxonomy' data field.
	For hybrid strains, more than one taxon name can be specified. The semicolon ";" must be used as a separation character.
Values	All taxon names included in the authoritative nomenclature references, reported according to the given syntax.
Validation	Check for the correct syntax and the existence of the taxon name(s) in the reference nomenclatures. Report errors to the CC.
Examples	Candidaceae
	Candida
	Candida albicans
	Candida albicans var. claussenii
	Actinomyces globisporus subsp. Flaveolus

Name	Infrasubspecific names
Short name	infrasubspecificNames
Description	Infrasubspecific names including biovar, chemovar, cultivar, morphovar, pathovar, phagovar, serovar, forma specialis, phase.
Syntax	The infrasubspecific name, usually preceded by a short specification of its type, e.g. "pv." for pathovar and "sv." for serovar.
Values	Free text
Validation	None
Examples	pv. lachrymans
	sv. Typhi

Name	Comment on taxonomy
Short name	taxonomyComments
Description	Any comment and/or note on the taxonomy of the strain that can be of interest for the users. It may be used, e.g., for information on new species or revised nomenclatures. It must be used when the Taxon name data field includes a name that is not present in the nomenclature reference. It can also include remarks, e.g. related to morphology, that may suggest a different classification. A comparison with a second taxon with which there could be a confusion, along with the reasons why it is not that, may also be included here, as well as the taxonomic name which was used by the depositor.
	It should not be used for information of exclusive interest for the collection and therefore not useful/of interest for the users.
Syntax	None

Values	Free text
Validation	None
Examples	This strain was labeled as Leptolyngbya antarctica ANT.GENTNER2.5 and studied by Taton et al. (2006). However, it is closely related to strains ULC017, ULC031, ULC03636, ULC037 and ULC043. As ULC031 was designated as reference strain for Shackletoniella antarctica, this strain probably belongs to that new genus and species. This strain was deposited as the type strain of Pediococcus inopinatus, but partial 16S rRNA sequence analysis and API 50 CHL tests question its identity at species level, being closely related to species Pediococcus acidilactici. Closely related to C. griseum, and belongs to the C. griseum clade sensu Zare and Gams (cf. Mycol. Prog. (2016): DOI 10.1007/s11557-016-1214-8)

Name	Type
Short name	type
Description	Specify whether the strain is a type strain.
Syntax	Specify one of the allowed values.
Values	One of the following values: 1 (for No), 2 (for Yes)
Validation	Check that one of the allowed values is present.
Examples	2

Name	Status
Short name	status
Description	For type strains, specify their type(s) (type, neotype, holotype, epitype, etc) and the related species name(s). A list of allowed values is not defined and this information must be provided as free text. Future improvements of the dataset will likely foresee a list of values for the kind of type.
Syntax	None
Values	Free text
Validation	Check for the kind of type
Examples	Holotype of Candida albicans Type of Candida albicans; Epitype of Something else

Name	History of deposit
Short name	depositHistory
Description	Transfers of the strain between isolation and deposit in the CC.
Syntax	The field includes entries separated by "<" meaning "received from". Entries may include persons or CCs. The name of the CC should be followed by the month, when available, and year of the acquisition. Between parentheses, the strain designation or CC numbers and/or a name can also be entered when a name change has occurred.

Values	Free text, according to above syntax
Validation	Check for the validity of the format. Report errors to the CC.
Examples	CECT, 1995 < CBS, 1990 < ATCC, 1989
	NCTC, Nov. 1973 (Bacillus loehnisii) < T. Gibson, 1935 < Kral Collection (Bacillus probatus)

Name	Depositor
Short name	depositor
Description	Name, institute and town / country of the depositor.
Syntax	None
Values	Free text
Validation	None
Examples	M. Sebald, Inst. Pasteur, Paris, France
	P. Hirsch, Inst. Allg. Mikrobiol. Univ. Kiel, Germany

Name	Date of deposit
Short name	depositDate
Description	Date when the strain was deposited at the CC
Syntax	Should include a full date in the ISO 8601 format.
	YYYY-MM-DD for full dates, YYYY-MM for year and month only, YYYY for year only.
	See https://en.wikipedia.org/wiki/ISO_8601 for a quick introduction.
Values	A valid date in one of the above formats
Validation	Check for the validity of the format.
	Report errors to the CC.
Examples	1999-02-20

Name	Collected by
Short name	collector
Description	Name, institute and town / country of the collector.
Syntax	None
Values	Free text
Validation	None
Examples	J. Fraser, Moredun Res. Inst., Edinburgh, UK

Name	Date of collection (MANDATORY)
Short name	collectionDate
Description	Date when the sample was collected.
Syntax	Must include a full date in the ISO 8601 format.
	YYYY-MM-DD for full dates, YYYY-MM for year and month only, YYYY for year only. When collected before 2014-10-12 and the date is unknown, put 0001-01-01.

	See https://en.wikipedia.org/wiki/ISO_8601 for a quick introduction.
Values	A valid date in one of the above formats
Validation	Check for the validity of the format. Report errors to the CC.
	When transferring data to the MIRRI-IS, the fictious data 0001-01-01 will be substituted by the text "Before 2014-10-12".
Examples	1999-11-27

Name	**Isolated by**
Short name	isolator
Description	Name, institute and town / country of the isolator.
Syntax	None
Values	Free text
Validation	None
Examples	I. Orskov, Ser. Inst., Copenhagen, Denmark
	D. Haas, Inst. Pasteur, Paris, France

Name	**Date of isolation**
Short name	isolationDate
Description	Date when the strain was isolated from the sample.
Syntax	Should include a full date in the ISO 8601 format.
	YYYY-MM-DD for full dates, YYYY-MM for year and month only, YYYY for year only. See https://en.wikipedia.org/wiki/ISO_8601 for a quick introduction.
Values	A valid date in one of the above formats
Validation	Check for the validity of the format. Report errors to the CC.
Examples	2019-08-17

Name	**Date of inclusion in the catalogue**
Short name	accessionDate
Description	Date when the strain was included in the catalog and/or an accession number was assigned to it.
Syntax	Should include a full date in the ISO 8601 format.
	YYYY-MM-DD for full dates, YYYY-MM for year and month only, YYYY for year only. See https://en.wikipedia.org/wiki/ISO_8601 for a quick introduction.
Values	A valid date in one of the above formats
Validation	Check for the validity of the format. Report errors to the CC.
Examples	1996-12-13

Name	**Tested temperature growth range**
Short name	temperatureGrowthRange
Description	The lowest and the highest temperature at which the strain was tested for growing.

Syntax	Temperatures are expressed as decimal numbers in Celsius degrees and must be separated by a semicolon. The symbol ° and the letter C should not be included.
Values	Decimal numbers
Validation	Check for the validity of the format. Report errors to the CC.
Examples	15;35

Name	Recommended growth temperature (MANDATORY)
Short name	recommendedTemperature
Description	The recommended growing temperature for the strain.
Syntax	The temperature is expressed as decimal number in Celsius degrees. The symbol ° and the letter C should not be included.
Values	Decimal number
Validation	Check for the validity of the format. Report errors to the CC.
Examples	24

Name	Recommended medium for growth (MANDATORY)
Short name	recommendedMedium
Description	The medium that is recommended for growing the strain. Only one medium can be specified for a strain.
Syntax	A textual reference, usually an acronym, to the appropriate growth medium in a table provided by the CC.
Values	CCs are invited to submit a table including a list of the growth media they use. The table should include at least an acronym and a description for each growth medium. If the CC also have a number associated to the medium, this should also be included. A full description of the recipe is also welcome. It may also include an HTML or XML or JSON definition of the medium. All descriptions must be in English.
Validation	Check for the presence of the textual reference in the provided table of growth media. Report errors to the CC.
Examples	AGA GYA

Name	Form of supply (MANDATORY)
Short name	supplyForms
Description	The forms of supply of the strain to users.
Syntax	One or several of the allowed values. If more values are includes, they must be listed in alphabetical order and separated by a ";".
Values	Allowed values: Agar, Cryo, Dry Ice, Liquid Culture Medium, Lyo, Oil, Water.
Validation	Check for the validity of the format. Report errors to the CC.
Examples	Cryo Agar; Lyo

Name	Other denomination
Short name	otherDenomination
Description	Unofficial names that are often used for the strain, e.g. in publications, or a name given to the strain by the isolator before its deposit at the collection.
Syntax	None
Values	Free text
Validation	None
Examples	S288c; AB1157

Name	Coordinates of geographic origin
Short name	geographicCoordinates
Description	The geographic coordinates and altitude of the location where the sample was collected.
Syntax	Latitude, longitude, precision and altitude. Values are separated by semicolons. Latitude and longitude are expressed in decimal degrees. Cardinal directions North and West are implicit and must not be reported. Precision must be expressed in kilometres as decimal number. Altitude must be expressed in meters above sea as decimal number. Both precision and altitude can be omitted. See examples. Conversion of latitude and longitude values from the sexagesimal format (as in 40° 26′ 46″) to the decimal format (40.446) can easily be achieved as follows: decimal degree = sexagesimal degree + (sexagesimal minutes/60) + (sexagesimal seconds/3600)
Values	Decimal numbers from -180 to 180 for longitude and -90 to 90 latitude. Decimal numbers for precision and altitude. Put a question mark for missing precision or altitude when one of them is missing. Leave values empty when they are both missing. See examples.
Validation	Check for the validity of the format and values. Report errors to the CC. When the information is missing, MIRRI-IS will try to determine it through the geographical database GeoNames (see http://geonames.org/) using the geographic origin information and will return it to the CC.
Examples	44.4128;8.8900;0.2;100 (all four values are available) 44.4128;8.8900;0.2;? (precision is available, but altitude is missing) 44.4128;8.8900;?;100 (precision is missing, but altitude is available) 44.4128;8.8900 (both precision and altitude are missing)

Name	Country (MANDATORY)
Short name	country
Description	The country where the sample was collected. This information should be provided with reference to the name of the country where the locality is located now, if possible. Former country names should be specified only when the locality or its current country are unknown. The Antarctic region and High Seas may also be specified.

	It is mandatory only for strains isolated after October 12th, 2014.
Syntax	This information must be expressed by using the ISO-3166 standard for country codes. The preferred set is ISO 3166-1 alpha-2 (two letters code), but ISO 3166-1 alpha-3 (three letters code) is also accepted. Former country codes must follow standard's part three ISO 3166-3 (four letters code). Only one code can be included. For High Seas, the code XX should be used.
Values	All codes included in ISO 3166-1 alpha-2 (two letters code), ISO 3166-1 alpha-3 (three letters code) and ISO 3166-3 (four letters code). In the examples, IT is the two letters code for Italy, GBR is the three letters code for United Kingdom, and CSHH is the four letters code for the former country Czechoslovakia.
Validation	Check for the validity of codes. Return errors to CCs.
Examples	IT GBR CSHH

Name	**Geographic origin (MANDATORY)**
Short name	geographicOrigin
Description	The locality where the sample was collected, defined with the highest possible precision. It is mandatory only for strains isolated after October 12th, 2014.
Syntax	Reference to a unique identifier in a separate table, which includes all localities where at least one strain was collected. For organisms constructed in a lab, use the address of the depositor.
Values	The geographic location should be defined in the separate table with the highest possible precision, but unambiguously. It should include locality, city, province, region, country. The use of English names is strongly recommended, at least for the most common geographic names. For old strains for which the geographic origin is not known, refer to the special locality 'Unknown'. Avoid specifying countries and continents only. The origin table can include either separate fields for the geographic details or one single text including all details. The first format is preferred over the second. In order to improve the description of the location, you can check if it is described in GeoNames (see http://geonames.org/) and use its 'Administrative hierarchy' to include further rows with information missing in the table, e.g. administrative commune and region, until you find the country. NB! While querying GeoNames, you may also recover geographic coordinates and altitude of the locality. E.g., in order to insert Altafjorden, look at GeoNames. You will find it associated to the GeoNames record n. 780944, whose administrative hierarchy reports Norway as country, Troms og Finnmark as adm1 and Alta as adm2. You will also retrieve 70.05765, 23.08293 for geographic coordinates.

	Altitude is not specified since this is a fiord.
	In your separate table for geographic origins you should include, either in separate cells or in a unique description, Altafjorden, Alta, Troms og Finnmark, Norway.
Validation	Check for the presence of the reference in the table of localities. Report errors to the CC.
Examples	1

Name	GMO
Short name	gmo
Description	Specify whether the strain is a Genetically Modified Organism (GMO).
Syntax	One of the allowed values.
Values	One of the following values: 1 (for No), 2 (for Yes).
Validation	Check that one and only one of the allowed values is used. Report errors to the CC.
Examples	1

Name	GMO construction information
Short name	gmoConstruction
Description	Information on the construction of the GMO. By now, this information can be provided as free text. Future improvements of the dataset will likely foresee some syntactical rules and/or list of values.
Syntax	None
Values	Free text
Validation	None

Name	Mutant information
Short name	mutant
Description	Information on mutant strains. By now, this information can be provided as free text. Future improvements of the dataset will likely foresee some syntactical rules and/or list of values.
Syntax	None
Values	Free text
Validation	None
Examples	X-ray mutant of NRRL 1951.B25 Glutamine auxotroph of strain 74-A

Name	Genotype
Short name	genotype
Description	Information on the genotype of the strain.

	By now, this information can be provided as free text. Some syntactical rules and/or list of values are foreseen in the next version of the MIRRI-IS dataset.
Syntax	None
Values	Free text
Validation	None
Examples	leu2-3 leu2-112 his4-519 can1 gln-1b

Name	Literature
Short name	identificationLiterature
Description	Information on literature linked to the identification and properties of the strain. Does not include literature related to the sequence of the strain, which should be included in the field "Literature linked to the sequence/genome". All data must be included in a distinct table. For publications indexed by Pubmed or having an official DOI number, collections should provide the relative identifiers, respectively PMIDs and DOIs. All usual bibliographic fields used for citing a paper, a book, a patent, or a document available on-line, including, e.g., authors, title, journal, volume, issue, pages, editors, publishers, etc… should be submitted as separate fields. Identifiers linking to the separate table must be included in the strain table. More papers can be included for a single strain by listing more identifiers separated by ";".
Syntax	Numeric identifiers separated by a semicolon ";".
Values	Reference numbers of the literature sheet.
Validation	None
Examples	12; 18; 24

Name	Sexual state
Short name	sexualState
Description	Information on strain sexual state / mating type, for relevant resource types.
Syntax	One of the allowed values. More can be added on-demand by CCs.
Values	Mata Matalpha Mata/Matalpha Mata Matb Mata/Matb MTLa MTLalpha MTLa/MTLalpha MAT1-1 MAT1-2 MAT1

	MAT2
	MT+
	MT-
Validation	Check that one and only one of the allowed values is used.
	Report errors to the CC.
Examples	Mata
	MTLa/MTLalpha

Name	**Ploidy**
Short name	ploidy
Description	Information on the ploidy level of the strain.
Syntax	One of the allowed values.
Values	One of the following values: 1 (for Haploid), 2 (for Diploid), 3 (for Triploid), 4 (for Tetraploid), 5 (for Polyploid (over 4n)), 9 (for Aneuploid).
Validation	Check that one and only one of the allowed values is used.
	Report errors to the CC.
Examples	2

Name	**Interspecific hybrid**
Short name	hybrid
Description	This field reports whether the strain is an interspecific hybrid.
Syntax	One of the allowed values.
Values	One of the following values: 1 (for No), 2 (for Yes).
Validation	Check that one and only one of the allowed values is used.
	Report errors to the CC.
Examples	2

Name	**Pathogenicity**
Short name	pathogenicity
Description	Information about pathogenicity of the strain for plants, humans and animals.
	Can include specification for the Belgian plant pathogenicity code.
Syntax	None
Values	Free text
Validation	None
Examples	Pathogenic to Agaricus bisporus.
	Transmissible murine colonic hyperplasmia.

Name	**Enzyme production**
Short name	enzymes
Description	Information about enzyme production by the strain. The CCs should also include the Enzyme Commission (EC) number of the enzyme after its name, included in

	parentheses. See, e.g. <u>ExplorEnz: Search the Official IUBMB Enzyme List (enzyme-database.org)</u>. By now, this information should be provided as free text. Future improvements of the dataset will likely foresee some syntactical rules and/or list of values.
Syntax	None
Values	Free text
Validation	None
Examples	Decarboxylase (4.1.1), Isomerase, Pectinase.

Name	**Production of metabolites**
Short name	metabolites
Description	Information about metabolite production by the strain. By now, this information should be provided as free text. Future improvements of the dataset will likely foresee some syntactical rules and/or list of values.
Syntax	None
Values	Free text
Validation	None
Examples	Capreomycin; oxytetracyclin.

Name	**Applications**
Short name	applications
Description	Information about applications of the strain. By now, this information should be provided as free text. Future improvements of the dataset will likely foresee some syntactical rules and/or list of values.
Syntax	None
Values	Free text
Validation	None
Examples	Biomass electricity generation Studies of pathway of beta-phenylpropionic acid metabolism Environmental restoration

Name	**Remarks**
Short name	remarks
Description	Any further note that is not present in the other fields.
Syntax	None
Values	Free text
Validation	None
Examples	Two stable colony types giving identical gel electrophoretic protein profiles. Strain was preserved after several local lesion passages with Nicotiana tabacum cv. Java as host plant.

Name	**Plasmids**

Short name	plasmids
Description	Information about plasmids in the strain.
	It may include plasmid name and type (original plasmid, cloning vehicle, recombinant plasmid), restriction sites, relevant genes (e.g., origin of replication, transposons, promoters, terminators, structural genes).
	By now, this information should be provided as free text. Future improvements of the dataset will likely foresee some syntactical rules and/or list of values.
Syntax	None
Values	Free text
Validation	None
Examples	pUZ8
	PO100 of HfrR4
	Plasmid free

Name	Plasmids collections fields
Short name	plasmidCollections
Description	Information about availability of strain plasmids in CCs of plasmids.
Syntax	It should include the name of the plasmid followed by the CC number in parentheses. More than one plasmid can be reported, separated by ";".
Values	Plasmid names should be provided as free text.
	CC numbers should be composed by the CC acronym followed by a number separated by a space.
Validation	Check the syntax of the information.
	Report errors to the CC.
Examples	pUZ8 (LMBP 8011)

Name	Substrate/host of isolation
Short name	substrate
Description	Information about the substrate and the host of isolation of the strain. It may include the detailed substrate from which the strain was isolated and the name of host plant/animal.
	By now, this information should be provided as free text. Future improvements of the dataset will likely foresee some syntactical rules and/or list of values.
Syntax	None
Values	Free text
Validation	None
Examples	Soil under Pinus sylvestris.
	Flowering plant of Helleborus foetidus.
	Arachis hypogaea.

Name	Isolation habitat
Short name	isolationHabitat
Description	Information about the biotope where the species was found.

	It should include environmental physical factors, such as humidity, range of temperature, pH and light intensity, as well as biotic factors, such as the availability of food and the presence or absence of predators. When some Ontobiotope identifiers are included in the related field, this field should include the corresponding preferred terms. It may also include information already specified in the related fields Geographic origin, Geographic origin coordinates and Altitude. By now, this information should be provided as free text. Future improvements of the dataset will likely foresee some syntactical rules and/or list of values.
Syntax	None
Values	Free text
Validation	None
Examples	Tropical rain forest. Salt marsh, Salicorna habitat. Forest litter, radioactivity 1.5x10(4) Bq/kg.

Name	Ontobiotope term for the isolation habitat
Short name	ontobiotopeTerms
Description	Information about the habitat where the species was found provided by using the most specific instance(s) of the Ontobiotope ontology of microorganism habitats. Note that this ontology is mainly for bacteria.
Syntax	The id(s) of the term(s) should be provided. Only use ids, not preferred names. Ontobiotope ids include the prefix "OBT:" followed by an integer of six digits, as in "OBT:001119". When submitting more ids, they must be separated by a semicolon ";".
Values	Any valid ids from the Ontobiotope. See the Ontobiotope browser.
Validation	Check for validity of ids. Check for the syntax.
Examples	OBT:001119; OBT:002941

Name	Genomic sequences and accession numbers
Short name	sequences
Description	Known genomic sequences and related INSDC accession numbers of the strain. According to the resource type, these include, but are not limited to, the nuclear ribosomal Internal Transcribed Spacer (ITS), the nuclear ribosomal Large SubUnit (LSU) and the 16S rRNA gene. Any further gene or marker that is considered of relevance by the CC, such as Calmodulin (CaM) and β-Tubulin, can be included. These data must be submitted in the Excel sheet "Genomic Information", including, in distinct fields, the following information: accession number of the strain in the CC, marker name, INSDC accession number of the marker sequence, sequence.
Syntax	Fields in the table follow different syntaxes:

	Strain accession number: as defined in the related field of the MIRRI-IS dataset.
	Marker name: the short name of the marker.
	INSDC accession number: An INSDC accession number is an alphanumeric code made by a fixed number of letters followed by a fixed number of digits, without any separation. For sequences, the code is currently made of two letters followed by six numbers.
	Sequence: Any valid genomic sequence.
Values	Values of fields in the table are as follows:
	Strain accession number: any accession number in the CC.
	Marker name: any common marker designation.
	INSDC accession number: Any valid INSDC accession number.
	Sequence: Genomic sequence, any format, any length.
Validation	Check for the validity of the syntaxes, formats and values. Check that the sequence in INSDC actually relates to the named gene sequence of the given strain. Report errors and discrepancies to the CC.
Examples	See attached table.

Name	Literature linked to the sequence/genome
Short name	sequenceLiterature
Description	Information on literature linked to the sequences or genome of the strain. Do not include here literature linked to the identification and properties of the strain. Include here identifiers linking to a separate literature sheet in the same file. All data must be included in a distinct table. For publications indexed by Pubmed or having an official DOI number, collections should provide the relative identifiers, respectively PMIDs and DOIs. When neither a PMID nor a DOI are available, all usual bibliographic fields used for citing a paper, a book, a patent, or a document available on-line, including, e.g., authors, title, journal, volume, issue, pages, editors, publishers, etc… must be submitted as separate fields. Identifiers linking to the separate table must be included in the resource table. Multiple papers can be included for a single strain just by reporting more identifiers separated by ";".
Syntax	Numeric identifiers separated by a semicolon ";"
Values	Reference numbers of the literature sheet.
Validation	Any errors and inconsistencies will also be reported to the CCs.
Examples	12; 26; 52

Name	Link to other sites
Short name	siteLinks

Description	URLs of extended descriptions of the strain in other information systems, such as GCM genomes, INSDC, UNIPROT. The site name will be used as a title for an active link in the MIRRI-IS for users willing to navigate to the site.
Syntax	Site name and working URL (including method, server and full path), joined by a semicolon character ";". The site name must be consistent in the CC, i.e. a given site must always be associated to a unique site name. If providing more than one site, separate pairs by a semicolon character ";".
Values	One or more pairs site name - valid URL.
Validation	Check that the URLs are working. Report errors to the CC.
Examples	Site 1;URL1;Site 2;URL2

Name	QPS
Short name	qps
Description	Specify whether the strain is from a species qualified as QPS (Qualified Presumption of Safety), according to the "Updated list of QPS-recommended biological agents for safety risk assessments carried out by EFSA" periodically delivered by the European Food Safety Agency (EFSA).
Syntax	One of the allowed values.
Values	One of the following values: 1 (for No), 2 (for Yes)
Validation	Check that one and only one of the allowed values is used. Report errors to the CC.
Examples	2

Name	Axenic culture
Short name	axenicCulture
Description	Specify whether the strain is an axenic culture, that is entirely free of all other contaminating organisms.
Syntax	One of the allowed values.
Values	One of the following values: "Axenic", "Not axenic".
Validation	Check that one and only one of the allowed values is used. Report errors to the CC.
Examples	Not axenic