Format for batch registration of DOIs in the Global Information System

History of changes

This document is available at

      http://www.fao.org/plant-treaty/areas-of-work/global-information-system/techdoc/en/

Please refer to the above URL for later versions of this document.

Version	Date	Description
1.7	15/10/2019	Renamed Method code "Observation - Inherited" to "Inherited" to harmonize it with the latest version of the Descriptors. Renamed the Excel template
1.6	14/02/2019	Increased max length of georeferencing method. Edited the wording on MLS status in Table 5 to match the Descriptors document
1.5	10/11/2018	Added n/a (Other) to identifier types. Added new user-defined codes to ISO-3166 alpha-3 to designate CGIAR centers. Explained that registration to Easy-SMTA is mandatory and therefore the Easy-SMTA PID is always available
1.4	09/02/2018	Clarified description of Historical. Harmonized with version 2.1 of document "Data required for the assignation of Digital Object Identifiers in the Global Information System". Changed crop name to common name to prepare for non-crop related PGRFA.
1.3	21/07/2017	Modified column Exist into Historical.
1.2	20/07/2017	Updated to new version of descriptors. Species (specific epithet) is now a mandatory descriptor but, if not provided, sp. is assumed. Physical existence of the material is no longer assumed. The new column Exist is introduced with values y or n. If empty or not present, y is assumed.
1.1	11/05/2017	Update target keywords table
1.0	29/03/2017	Removed geographic coordinates from M01
0.1	03/02/2017	Initial draft

Introduction

This document describes the batch registration procedure available in the Global Information System (GLIS) of the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA).

The procedure is based on an Excel document that you populate with information about the PGRFA you wish to register in GLIS. A template of the Excel document (GLIS data template 1.7.xlsx) is available together with this document. We strongly advise you to use it.

Once you have completed filling out the template, send it to us (plant-treaty@fao.org); after having validated it, we will upload it to GLIS and report about the success or failure of the upload.

This document refers to the latest version of "Data required for the assignation of Digital Object Identifiers in the Global Information System" available at

      http://www.fao.org/plant-treaty/areas-of-work/global-information-system/descriptors/en/

Please refer to that document for details on the meaning of the descriptors.

The compilation of the Excel table, being usually done by hand, is quite time consuming and error-prone and should be adopted only for small collections or as a temporary solution. If possible, we recommend the use of the XML-based protocol that GLIS offers. Details on such system-to-system integration protocol can be obtained by contacting the Treaty Secretariat.

Please note that quite a few columns will likely contain constant values, for example:

Column	Descriptor	

A	M01	The FAO WIEWS code of the organization holding the PGRFA being registered. When an entire collection is registered, this value, if available, is likely to be the same for all rows. When the WIEWS code is entered in column A, columns C to E can be left empty
B	M01	In order to obtain DOIs, the holder of the PGRFA must be registered on Easy-SMTA and have a PID, such PID will be the same for all rows. When the Easy-SMTA PID is entered in column B, columns A and C to E can be left empty

Similar considerations can be applied to other columns so that the compilation of the Excel table may be simpler than expected.

General formatting rules

All cells in the Excel document must be formatted as text; it is particularly important the Excel does not format dates and numerical values.

Some descriptors are mapped to two or more columns in the Excel document because they contain separate pieces of information.

Some descriptors allow for multiple values to be specified. In such cases, unless otherwise specified, subsequent values are to be separated using the character "|" (vertical pipe) without spaces, e.g.

<div align="center">

`value1|value2|value3`

</div>

Never enter spaces anywhere in the Excel table unless they are part of the actual information. Even if the content of a cell looks misaligned or garbled, GLIS will interpret it correctly if you scrupulously follow the format specifications provided in this document.

Dates must always be entered in the `YYYY-MM-DD` format. Please note that date fragments are also accepted (e.g. `YYYY-MM` or `YYYY`).

In addresses, multiple lines can be entered using the "|" (vertical pipe) as line separator, e.g.

`123, Some St.|10123 Somecity`   will be interpreted as
```
123, Some St.
10123 Somecity
```

Do not enter the country name in the address as the country is provided in a separate element.

Countries must be specified using the ISO-3166 alpha-3 country codes available at https://en.wikipedia.org/wiki/ISO_3166-1_alpha-3. For historical reasons, some custom country codes are also made available as described in Table 7 below.

Links, i.e. URL of web pages where additional information about the PGRFA can be found, must be associated to one or more keywords describing what kind of information the user would find at that page. Keywords are represented by codes as described in Table 2 below.

For example, let us imagine that the URL http://www.test.org/123 will provide characterization and evaluation data about a specific PGRFA and URL http://www.imagebank.org/photos/456 will show pictures of the same PGRFA sample. The correct way to enter these two links in column L of the Excel template is:

<div align="center">

`http://www.test.org/123|http://www.imagebank.org/photos/456`

</div>

and the corresponding keyword codes (column M of the Excel template) would be

<div align="center">

`2,3|5`

</div>

it is critical to enter keyword code lists in the same order as the links as they will be associated to the links in positional order.

For column T, Other identifiers, the type of each identifier must be specified and multiple identifiers can be separated by "|" as in the following example:

<div align="center">

`type,identifier|type,identifier|…`

</div>

The `type` is the code taken from Table 4 below. For example, assuming that the following identifiers are to be provided:

Type	Value
ARK	ark:/12148/bpt6k107371t
Genesys UUID	55e94ae9-aff7-4fbb-88d1-1de17c5e683f

the correct content of column T would be:

```
ark,ark:/12148/bpt6k107371t|genesysuuid:55e94ae9-aff7-4fbb-88d1-1de17c5e683f
```

In the following tables, the column contents are:

Column	Description
Col	Column in the Excel template
Desc.	Descriptor in document "GLIS descriptors v2", see the Introduction of this document
Label	The label of the column in the Excel template
Type	Data type of the column values
Len/Frm	Maximum length or format of the values to be entered in the column. For columns that may contain multiple values, an asterisk is present (e.g. 128*) to indicate that the length applies to each item of the list
Description	A brief description of the column's content

Mandatory descriptors

Mandatory descriptors are located in columns **A** to **K** of the Excel template as in the following table. Please note that it is not necessary to provide all elements in the M01 descriptor if one between column A and column B is provided as it is enough to fully identify the location (organization, legal entity or individual) where the PGRFA is conserved. However, when neither the FAO WIEWS institute code nor the Easy-SMTA PID is available, name, address and country should be indicated.

Col	Desc	Label	Type	Len/Format	Description
A	M01	FAO WIEWS code	String	16	FAO WIEWS code of the organization where the PGRFA is conserved, if available. If this value is provided, GLIS will use it to obtain all other elements of the M01 descriptor
B	M01	Easy-SMTA PID	String	16	Easy-SMTA PID of the organization, legal entity or individual conserving the PGRFA. This value must be provided, and, if no WIEWS code is provided in column A, GLIS will use it to obtain all other elements of the M01 descriptor
C	M01	Name	String	128	Surname and name of individuals or organization name of the organization or legal entity conserving the PGRFA. This column is provided for compatibility, but its content is ignored. It can be left empty
D	M01	Address	String	128	Address of the organization, legal entity or individual conserving the PGRFA. This column is provided for compatibility, but its content is ignored. It can be left empty
E	M01	Country	String	3	ISO-3166 alpha-3 code of the country of the organization, legal entity or individual conserving the PGRFA. This column is provided for compatibility, but its content is ignored. It can be left empty
F	M02	PGRFA unique identifier	String	128	A string that *uniquely* identifies the PGRFA that is being registered in holder's collection. This value will be returned by GLIS in the response message and is assumed to be used to associate the DOI to the corresponding record in your local database
G	M03	Date	Date	YYYY-MM-DD	Date in which PGRFA became part of the collection. Date fragments (YYYY-MM and YYYY) are also accepted. Make sure that Excel does not format this

Col	Desc	Label	Type	Len/Format	Description
					column as "date" but as "text"
H	M04	Method	String	4	Method through which the PGRFA has been obtained. See Table 1 below for the codes accepted by this element
I	M05	Genus	String	64	The taxon of the genus for the PGRFA. At least one between column I and column K must be provided
J	M05	Species	String	128	Specific epithet of the PGRFA scientific name
K	M05	Common name(s)	String	128*	Common name of the PGRFA. At least one between column I and column K must be provided. Multiple names must be separated with "|", you can specify names in any language, e.g. rice|riz|arroz. Please make sure to provide at least the name in English

Highly recommended descriptors

Highly recommended descriptors are located in columns **L** to **W** of the Excel template as in the following table. They are not mandatory but should be provided if available to allow GLIS to provide advanced search functions on your PGRFA.

Col	Desc	Label	Type	Len/Format	Description
L	R01	Links to associated information	String	256*	URL of the web page where additional information on the PGRFA can be found. Any number links can be provided but must be separated using "|"
M	R01	Link keyword codes	String	16*	Each link must be associated to one or more keyword codes taken from Table 2 below and separated by a comma (','). Each code list is separated from the next by "|". A code list in position n is associated to the link in position n. Please see under "General formatting rules" above for details
N	R02	DOI(s) of progenitor(s)	String	256*	DOI of the PGRFA(s) from which the current PGRFA was derived. The DOI must be registered in GLIS. Multiple DOIs must be separated using "|". Please note that the number of allowed DOIs depends on the method as follows: Acquisition: 1 DOI In-house copy: 1 DOI In-house variant: 1 DOI Novel distinct PGRFA: 1 or more DOIs Observation - Natural: 0 DOI Inherited: 0 DOI
O	R03	Biological status	String	3	Biological status of the PGRFA. See Table 3 below
P	R04	Species authority	String	64	Authority for the specific epithet
Q	R04	Subtaxa	String	128	Any additional infra-specific taxon such as subspecies, variety, form, Group and so on
R	R04	Subtaxon authority	String	64	Authority for the subtaxon at the most detailed level provided
S	R05	Names	String	128*	Other names or designations of the PGRFA. Any number of names can be provided but must be separated by "|"
T	R06	Other identifiers	String	128*	Types and values of other identifiers associated to the PGRFA as indicated in "General formatting rules" above
U	R07	MLS status	String	2	Code that identifies the status of the PGRFA with regard to the MLS. See Table 5 below
V	R08	Historical PGRFA	String	y/n	Indicates whether the PGRFA no longer exists. Allowed value is y/n with y=material is no longer available and n=material available

Additional descriptors

Additional descriptors are located in columns **X** to **AZ** of the Excel template as in the following table. They are not mandatory but should be provided if available to allow GLIS to provide advanced search functions on your PGRFA.

Col	Desc	Label	Type	Len/Format	Description
W	A01	FAO WIEWS code	String	16	FAO WIEWS code of the organization providing the PGRFA, if available. If this value is provided, GLIS will

					use it to obtain all other elements of the A01 descriptor
X	A01	Easy-SMTA PID	String	16	Easy-SMTA PID of the organization, legal entity or individual providing the PGRFA, if available. If this value is provided, GLIS will use it to obtain all other elements of the A01 descriptor
Y	A01	Name	String	128	Surname and name for individuals or organization or legal entity name of the provider
Z	A01	Address	String	128	Address of the organization, legal entity or individual providing the PGRFA. Separate multiple lines with "|"
AA	A01	Country	String	3	ISO-3166 alpha-3 code of the country of the organization, legal entity or individual providing the PGRFA. See Table 7 for details
AB	A02	PGRFA unique ID	String	128	Unique identifier for the PGRFA in the provider's management. Please note that this is not the DOI of the PGRFA assigned by the provider (that, if available, should be entered in column N, R02, DOI(s) of progenitor(s)), but is rather the identifier that the provider assigned to the PGRFA locally in his collection (e.g. the provider's accession number for genebanks)
AC	A03	Country of provenance	String	3	ISO-3166 alpha-3 code of the country in which the PGRFA material was either collected or bred or selected, or the first country in the known history of the PGRFA. For historical reasons, this column can also accept the extended codes listed in Table 7 below
AD	A04	FAO WIEWS code	String	16	FAO WIEWS code of the organization collecting the PGRFA, if available. If this value is provided, GLIS will use it to obtain all other elements of the A04 descriptor
AE	A04	Easy-SMTA PID	String	16	Easy-SMTA PID of the organization, legal entity or individual collecting the PGRFA, if available. If this value is provided, GLIS will use it to obtain all other elements of the A04 descriptor
AF	A04	Name	String	128	Surname and name for individuals or organization or legal entity name of the collector
AG	A04	Address	String	128	Address of the organization, legal entity or individual collecting the PGRFA. Separate multiple lines with "|"
AH	A04	Country	String	3	ISO-3166 alpha-3 code of the country of the organization, legal entity or individual collecting the PGRFA. See Table 7 for details
AI	A05	Collector's PGRFA unique ID	String	128	Identifier assigned by the collector(s) to the PGRFA collected
AJ	A06	Collecting mission ID	String	128	Identifier of the collecting mission
AK	A07	Location where PGRFA was collected	String	128	Description of the location where the PGRFA was collected
AL	A08	Latitude	String	10	Latitude where the PGRFA was collected in either DD°MM'SS"X (where X is N or S) format or DDD.XXXXX (up to 5 decimals, preceded by minus sign for S) format. No spaces are allowed
AM	A09	Longitude	String	10	Longitude where the PGRFA was collected in either DD°MM'SS"X (where X is E or W) format or DDD.XXXXX (up to 5 decimals, preceded by minus sign for W) format. No spaces are allowed
AN	A10	Uncertainty	String	16	Uncertainty of lat/lon coordinates, in metres
AO	A11	Geodetic *datum*	String	16	Geodetic *datum* of the lat/lon coordinates
AP	A12	Georeferencing method	String	128	Georeferencing method
AQ	A13	Elevation	Integer	16	Elevation of collecting site in metres above sea level
AR	A14	Collecting date	Date	YYYY-MM-DD	Date on which the PGRFA was collected. Date fragments (YYYY-MM and YYYY) are also accepted
AS	A15	Collecting source	String	2	Code of the nature of the location where the PGRFA was collected. See Table 6 below
AT	A16	FAO WIEWS code	String	16	FAO/WIEWS code of the breeding organization, if available. If this value is provided, GLIS will use it to obtain all other elements associated to the A16 descriptor
AU	A16	Easy-SMTA PID	String	16	Easy-SMTA PID of the breeding organization, legal entity or individual, if available. If this value is provided, GLIS will use it to obtain all other elements of the

					A16descriptor
AV	A16	Name	String	128	Surname and name for individuals or organization name of the breeder
AW	A16	Address	String	128	Address of the breeding organization, legal entity or individual. Separate multiple lines with "|"
AX	A16	Country	String	3	ISO-3166 alpha-3 code of the breeding organization, legal entity or individual. See Table 7 for details
AY	A16	Ancestry	String	32.767	Pedigree or other description of the ancestry of the PGRFA and how it was bred. Please note that this column should not contain DOIs of ancestors (that, if available, should be entered in column N, R02, DOI(s) of progenitor(s)), but is rather a list of identifiers assigned to the PGRFA locally in your or somebody else's collection (e.g. the provider's accession number for genebanks)

Response table

When no error is found in the Excel table, GLIS will produce a response table with the following columns:

Column	Label	Description
A	WIEWS	The FAO/WIEWS code of the organization registering the PGRFA. This column will contain the same values you provided in column A of the Excel table
B	PID	The Easy-SMAT PID of the organization, legal entity or individual registering the PGRFA. This column will contain the same values you provided in column B of the Excel table
C	Genus	The genus of the PGRFA. This column will contain the same values you provided in column I of the Excel table
D	PGRFA ID	The identifier used in your system to identify the PGRFA. This column will contain the same values you provided in column F of the Excel table
E	DOI	The DOI assigned to the PGRFA

The table contains both the genus and the identifier you provided in the Excel table to allow you to uniquely identify the PGRFA in order to record the corresponding DOI.

Errors

Should any error be found in the Excel table that you send us, we will provide exhaustive explanations and suggestions for fixing it.

In some cases, when there is no doubt about the correct value of a column, we may adjust it and proceed with the upload, for instance, when dates are not properly formatted. In such cases, we will inform you of the changes we have made.

Coding tables

This section lists the codes used in the tables above unless a link is provided to some external reference site in the column description. Codes must be entered exactly as shown.
Some of the tables are derived from FAO/Bioversity MCPD v.2.1, 2015.

Table 1: M04 - Method

Code	Description
acqu	Acquisition
ihcp	In-house copy
ihva	In-house variant
nodi	Novel distinct PGRFA
obna	Observation - Natural
obin	Inherited

Table 2: R01 – Target keyword codes

Code	Description
1	Passport data
1.1	Genealogy
1.2	Collection documents
2	Characterization
3	Evaluation
3.1	Chemical analysis
3.2	Abiotic stress
3.3	Biotic stress
3.4	Biochemical markers
3.5	Molecular markers
3.6	Cytological characters
3.7	Genomics
3.8	Phenomics
4	Environment
5	Multimedia

Table 3: R03 – Biological status

Code	Description
100	Wild
110	Natural
120	Semi-natural/wild
130	Semi-natural/sown
200	Weedy
300	Traditional cultivar/landrace
400	Breeding/research material
410	Breeder's line
411	Synthetic population
412	Hybrid
413	Founder stock/base population
414	Inbred line (parent of hybrid cultivar)
415	Segregating population
416	Clonal selection
420	Genetic stock
421	Mutant
422	Cytogenetic stocks
423	Other genetic stocks
500	Advanced or improved cultivar
600	GMO
999	Other

Table 4: R06 – Other identifier's type

Code	Description
ark	ARK
genesysid	Genesys ID
genesysuuid	Genesys UUID. The prefix urn:uuid: can be omitted
gmsid	GMS ID
lsid	LSID
purl	PURL
sgsvid	Global Seed Vault ID
n/a	Other

Table 5: R07 – MLS status

Code	Description
0	Not available under the MLS
1	Available under the MLS
11	The PGRFA is of a crop listed in Annex I and is under the management and control of a Contracting Party to the Treaty and in the public domain
12	The PGRFA is in an international collection under Article 15 of the Treaty
13	The holder received the PGRFA with an SMTA
14	The holder has voluntarily placed the PGRFA in the MLS
15	The PGRFA is derived from, and distinct from, material previously received from the MLS, is still under development and not yet ready for commercialization, and may be made available at the discretion of the developer, with an SMTA

Table 6: A15 – Collecting source

Code	Description
10	Wild habitat
11	Forest or woodland
12	Shrubland
13	Grassland
14	Desert or tundra
15	Acquatic habitat
20	Farm or cultivated habitat
21	Field
22	Orchard
23	Backyard, kitchen or home garden (urban, peri-urban or rural)
24	Fallow land
25	Pasture
26	Farm store
27	Threshing floor
28	Park
30	Market or shop
40	Institute, Experimental station, Research organization, Genebank
50	Seed company
60	Weedy, disturbed or ruderal habitat
61	Roadside
62	Field margin
99	Other

Table 7: Extensions to the ISO-3166 alpha-3 list of country codes

To address the historical or approximate nature of country information available in some cases, according to the ISO-3166 standard, the following user-defined and historical codes are also available in addition to the standard ones.

Name	Code
Alpine	XAL
Andes	XAN
Arabia	XAR
Australasia	XAA
Australia & New Zealand	XAZ
Benelux	XBE
Bengal	XBN
Cape Horn	XCH
Caribbean	XCR
Caspian	XCP
Central Africa	XCF
Central America	XCA
Czechoslovakia	CSK
East Africa	XEF
Eastern Europe	XEE
Far East	XFE
German Democratic Rep.	DDR
Germany, Federal Rep. of	BRD

Himalaya	XHM
Iberia	XIB
Mediterranean	XMD
Middle East	XME
New Hebrides	NHB
North Africa	XNF
North America	XNA
North-East Asia	XNE
Pacific Ocean	XPO
Panama Canal Zone	PCZ
Sahara	XSH
Scandinavia	XSC
Sea of Japan	XSJ
Serbia and Montenegro	SCG
South America	XSA
South East Asia	XAS
Southern Africa	XSF
Southern Europe	XSE
Union of Soviet Soc. Rep.	SUN
Upper Volta	HVO
West Africa	XWF
Western Europe	XWE
Yemen, Democratic	YMD
Yugoslavia	YUG

Additionally, to designate International Institutions such as the CGIAR as provenance of the PGRFA being registered, the following codes can also be used

Name	Code
Africa Rice	XAB
Bioversity International	XAC
CIAT	XAD
CIMMYT	XAE
CIP	XAF
ICARDA	XAG
ICRAF	XAH
ICRISAT	XAI
IITA	XAJ
ILRI	XAK
IRRI	XAM