

## 5 Tips and tricks for batch registration

Version 1.1 - 21/10/2020

### Introduction

---

GLIS supports a batch registration procedure that we recommend for those institutions whose collections that are too large for manual registration using the web form but prefer not to adopt the XML protocol.

We recommend batch registration up to a few thousand PGRFA, i.e. a medium size collection. However, it can also be used for larger collections when DOIs are urgently needed or while the XML protocol is adopted.

Please keep in mind that batch registration does not support updating descriptors of DOIs that have been already assigned.

The documentation on batch registration is available at the following URL and consists of an Excel template and a document that explains in detail how to fill it in.

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

Also, it could be useful to refer to the document on the GLIS Descriptors available at

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

This document provides suggestions on how to best fill in the template based on the experience we have accumulated so far working with registrants.

### Excel template structure

---

The Excel template contains three sets of columns:

- **Mandatory descriptors:** These descriptors *must* be provided to get a DOI from GLIS. Their column headers are shaded in green,
- **Highly recommended descriptors:** These descriptors are not mandatory and can be left empty. However, we strongly recommend you provide as much information as possible because they allow GLIS to offer more advanced services to users. Their column headers are shaded in yellow,
- **Additional descriptors:** These descriptors depend on the context (e.g. information on collection applies only to those PGRFA that have been collected) and can be left empty but, again, we invite you to provide the information if readily available. Their column headers are shaded in blue.

In the following, suggestions to best filling in each set of columns will be provided.

### Mandatory descriptors

---

As explained above, mandatory descriptors must be provided. However, here are some considerations facilitating this.

- **M01: Organization or individual conserving the PGRFA** This descriptor is actually comprised of 5 columns. Of these, the really mandatory one is B because the Easy-SMTA PID is always available as the holder of the PGRFA must be registered in Easy-SMTA to be able to assign DOIs. All other columns under M01 can be derived from the Easy-SMTA PID. Also, the Easy-SMTA PID will be the same for all PGRFA being registered. If available, please provide also the WIEWS code in column A

- *M03: Date* Please make sure to use the format YYYY-MM-DD. In those cases where the exact date is unknown you can use YYYY-MM or just YYYY
- *M04: Method* Please use one of the codes listed in Table 1 of the document on batch registration available at the URL mentioned above. In most cases, for genebanks, the method will be Acquisition and therefore this column will contain acqu for all rows
- *M05: Scientific or common name* The descriptor's label already says that you can specify one of the two or both. For the scientific name, only the genus is enough and, of course, you cannot specify the specific epithet without providing the genus. For the common name, you can enter names in any language (separated by "|") but please make sure you provide at least the common name in English to facilitate user searches

## Highly recommended descriptors

---

These descriptors can be safely left empty, however, as explained below, they are very important and should be provided whenever possible. Despite their importance, some of these descriptors are not mandatory because registrants may not have the necessary information available and we rather have a PGRFA registered, albeit with incomplete information, than not registered at all.

The most important of such descriptors are:

- *R01: Link to associated information* This descriptor (and Link keyword codes) are fundamental to let GLIS users looking at your PGRFA's page access at least your own website's page on the same PGRFA; it is a great way to provide visibility to your work. Please refer to GLIS Technical Note 2: *Links to websites* for further details. It is not mandatory because not all holders have a web presence
- *R02: DOI(s) of progenitor(s)* This is another very important descriptor as it allows GLIS to build the genealogy tree of a PGRFA. It is not mandatory because the ancestor(s) of the PGRFA may be unknown or may not have a DOI assigned. We have dedicated GLIS Technical Note 3: *PGRFA-DOI relationships* to this and we invite you to read it
- *R03: Biological status* This is also quite important. As a MCPD descriptor, it is likely to be available, at least in genebanks. Please enter one of the codes listed in Table 3 in the document on batch registration
- *R05: Names* Please read the meaning of this descriptor. Here you should enter the name(s) assigned to the PGRFA as a "landrace, traditional variety or modern cultivar, or some other name or designation used to identify a breeder's selection or elite line or variety". Do not enter here the common name that you have already entered in M05: Scientific or common name or the identifier you entered in M02: PGRFA unique identifier

## Countries

---

Countries are represented in GLIS by ISO 3166 alpha-3 codes. Such codes are composed by 3 (usually uppercase) letters. For example: ITA for Italy, GBR for United Kingdom or DZA for Algeria, BWA for Botswana or KHM for Cambodia. As you can see, the code does not necessarily coincide with the first 3 letters of country in English! Please make sure to use the correct code, otherwise GLIS will report an error.

There are also country codes identifying countries that do not exist anymore, such as SUN for the Soviet Union or YUG for Yugoslavia. However, for PGRFA that have a long history, the associated information may only include such countries, without any further specification that would allow to specify one of the currently existing countries. GLIS accepts those historical country codes.

Beside the official country codes, the ISO 3166 standard supports of user-assigned code elements that can be used to identify countries, territories or other entities according to the meaning defined within the specific user community.

GLIS uses this option in two cases:

- geographic areas that are not better defined such as "Andes" (code XAN) or "Sahara" (code XSH)
- international institutions such as CGIAR centers like IRRRI (code XAM) or CIAT (code XAD)

In the case of international institutions, the introduction of such codes allows for more accurate representation of the provenance of the PGRFA. Material produced by a breeding program of CIMMYT, for instance, does not really originate from Mexico, or any other country where the

CIMMYT breeding program may be established, but rather from CIMMYT itself as an international institution. Therefore, the code to be used should be XAE rather than MEX.

Please refer to the document on batch registration or XML protocol for the complete list of codes accepted by GLIS<sup>1</sup>.

## Geographic coordinates

---

Geographic coordinates in GLIS can be expressed in any of the following representations:

- DDD°MM' SS" X (also known as DMS) where DDD are the degrees ( $\pm 90$  for latitude and  $\pm 180$  for longitude), MM are the minutes (00 to 59), SS are the seconds (00 to 59) and X is N/S for latitude or E/W for longitude
- $\pm$ DDD.XXXXX (also known as decimal coordinates) where DD are the degrees ( $\pm 90$  for latitude and  $\pm 180$  for longitude) and XXXXX is the fraction of degree (from 00000 to 99999) to be added. Positive latitudes are north of the equator, negative latitudes are south of the equator. Positive longitudes are east of the Prime meridian, negative longitudes are west of the Prime Meridian

GLIS prefers the decimal notation so, if you provide coordinates in the DMS notation, they will be converted into the decimal notation.

## MCPD codes

---

Most GLIS descriptors are derived from the Multi-Crop Passport Descriptors, a standard for information exchange on plant genetic resources originally developed in 1997 by FAO and Bioversity International (IPGRI at that time) and widely used by genebanks worldwide. GLIS adopted MCPD version 2.1 and will follow the development of the standard. When there is direct correspondence between a GLIS descriptor and a MCPD descriptor that uses a code, GLIS uses the same codes as the MCPD. Such descriptors are:

GLIS descriptor	MCPD descriptor
R03: Biological status	SAMPSTAT
R07: MLS status	MLSSTAT
A15: Collecting source	COLLSRC

Please note that descriptor R07: MLS status uses the same codes 0 and 1 defined for MLSSTAT but also includes sub-codes for the case Included (explaining in higher detail why the PGRFA is to be considered included in the MLS) that are not accepted by MLSSTAT.

## I am not fully clear on what needs to be done, please help!

---

Should you require further clarifications, please do not hesitate to send an email to PGRFA-Treaty@fao.org and we will be delighted to assist!

---

<sup>1</sup> See <http://www.fao.org/plant-treaty/areas-of-work/global-information-system/techdoc>