

## 3 PGRFA-DOI relationships

Version 1.1 - 21/10/2020

### Introduction

During the life of a PGRFA, events can occur that somehow transform it or change its status or produce new PGRFA based on it. One of the main purposes of GLIS, thanks to the adoption of a DOIs, is to record the relationships that are established among PGRFA (and therefore among the DOIs that identify them) during these operations.

The full conceptual framework on which the GLIS model is based is described in the Guidelines<sup>1</sup>, in this document we will provide a synthetic recollection of some of the concepts and focus on the technical details of how these relationships are represented and what can be done with them.

### How PGRFA (and their DOIs) are related

Depending on what is done with a PGRFA (to which a DOI is assigned), another PGRFA may result that can be assigned another DOI; the two DOIs are related in a way that is recorded in GLIS. Descriptor<sup>2</sup> M04: Method tells us how the subsequent PGRFA is obtained.

GLIS currently provides 6 different methods that can be briefly described as follows. Please note that, as genetic transformations may be involved as well as just transfers or collection, we will use the terms "Predecessor" to identify the PGRFA that existed before the event and "Successor" the PGRFA that exists after the event.

Each method involves zero, one or more Predecessors as follows

Method	Code	Number of Predecessors
Acquisition	acqu	0 (e.g. for collection) or 1 (e.g. for transfers)
In-house copy	ihcp	1
In-house variant	ihva	1
Novel distinct PGRFA	nodi	1 or more
Observation - Natural	obna	0
Inherited	obin	0

Table 1: Number of Predecessors according to the Method

Whenever a new PGRFA is registered in GLIS, using the appropriate method and providing the Predecessors, a relationship among such Predecessors and the Successor PGRFA (and therefore among the corresponding DOIs) is established.

Please note that the last two Methods do not really establish a relationship as there is no Predecessor and no Successor; likewise for Acquisition without a Predecessor, as in the case of collection or acquisition from *in situ* conditions.

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

<sup>2</sup> See <http://www.fao.org/plant-treaty/areas-of-work/global-information-system/descriptors>

## How PGRFA (or DOI) relationships are modeled in GLIS

---

The relationships among PGRFAs (and their DOIs) constitute a *graph* where the nodes are the PGRFA/DOIs and the arcs are the actions connecting them together, represented by the Methods. The graph is also *directed*, i.e. each arc has a direction leading from the Predecessor to the Successor.

In Figure 1 below, an example is provided with two Predecessors that are combined through Method "Novel distinct PGRFA" (ND in the diagram) to obtain a Successor.

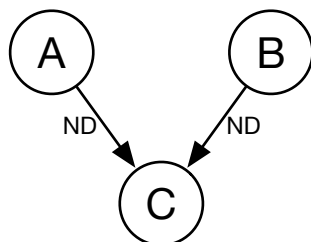


Figure 1: PGRFA C is obtained by crossing A and B

This is a directed graph with three nodes (A, B and C) and two arcs (A-C and B-C); arcs must be represented with the correct direction from Predecessor to Successor.

While graphs can be managed using regular SQL databases, they are best handled using specialized tools. GLIS uses Neo4J<sup>3</sup>, a leading graph database manager, exclusively to manage the relationship graph.

## What can be done with the relationship graph

---

The first and most obvious purpose of the relationship graph is, starting from a specific PGRFA, to find its ancestors to understand how that material came into existence.

Going in the other direction, is it easy to find all PGRFA that have been somehow obtained from the current one, i.e. its descendants. These queries can be limited to any number of "generations" (or "hops" in graph terminology) up or down the genealogy.

As the arcs are labelled with the Method, it also possible to look at specific modalities through which ancestors and/or descendants have been obtained.

Remember that at each PGRFA node one or more links to external websites, publications and datasets can be associated; therefore, by navigating the graph, it would be possible to access information associated to ancestors or descendants. It will be the user's responsibility to decide if and at what extent such information may be relevant to the PGRFA she started from. For example, if the Method is *Acquisition* or *In-house copy*, it is likely that results associated to immediate neighbors will be largely applicable to the PGRFA at hand. For other Methods, implying genetic modifications, some of the results may be not so relevant.

An interesting query for a holder would be to find out what has been done with the PGRFA distributed. The graph would provide all descendants and it would be possible to list the links to external websites, the publications and datasets associated to each one.

Similarly, in an aggregate fashion, it would be possible to establish the "most popular" PGRFA, i.e. those that are most frequently transferred or used for selection, variants or genetic modifications.

The graph could also be used to establish potential duplications, a problem that has been plaguing genebanks for decades, leading to a more efficient distribution of PGRFA.

The Secretariat plans to implement an interactive graphic browser allowing GLIS users to run queries such as

```
show me the PGRFA within 3 hops from the current one that have  
been obtained by Acquisition and have links to Genomic information
```

The user would then be able to navigate through the graph and explore the neighboring PGRFA.

---

<sup>3</sup> See <https://www.neo4j.com>

Of course, all this is possible if registrants dutifully provide the DOI(s) of Predecessor(s). This information is not required to register a PGRFA, but it is clearly one of the most important ones and should be provided whenever available.

## How are relationships established in GLIS?

There are several events leading to the "creation" of a new PGRFA (with its own, new DOI associated) as explained in the Guidelines document recalled above. In practice, the relationship is established by filling-in descriptor R02: DOI(s) of progenitor(s). This descriptor is available:

- in the web form (see Figure 2 below),
- in the Excel template for batch registration (see Figure 3 below), and
- in the XML document (see Figure 4 below)

Figure 2: Descriptor R02: GLIS DOI(s) of progenitor(s) in the web form

ord codes	DOI(s) of progenitor(s)	Biolog
1	R02	
	1.18730/JEBGP 10.18730/SBBDZ	
	10.18730/ABGZC	

Figure 3: Descriptor R02: DOI(s) of progenitor(s) in column N of the Excel template

```

...
<progdoi>
  <doi>10.18730/35F1</doi>
  <doi>10.18730/FE23A</doi>
</progdoi>
...

```

Figure 4: Fragment of the XML message with progenitors' DOIs indicated

The DOI(s) provided must:

- be already registered in GLIS. This means that you must register the Predecessors before you can register the Successors and connect them to the corresponding Predecessors,
- be in the correct number depending on the Method indicated as per Table 1 above; GLIS will flag an error otherwise

Please note that, if you use the XML protocol or the Web form that support updates, and as the R02 descriptor is recommended but not mandatory, you can register all PGRFA at once and then update the Successors with the DOIs of the corresponding Predecessors. Batch registration through Excel tables does not support updates at this time.

## What can GLIS do if you do not provide the DOI(s) of Predecessor(s)

Maintaining and enriching the relationship graph is one of the main advantages offered by GLIS. If you do not provide the DOI(s) of Predecessor(s), part of the graph will be missing, potentially

limiting other stakeholders who will, for instance, be unable to properly identify the ancestry of your material.

Therefore, it is very important that you provide the DOI(s) of Predecessor(s) whenever available. Should this not be the case, and when you have received the PGRFA from another collection, we strongly recommend that you accurately report (if available, of course) the provider and the unique identifier<sup>4</sup> assigned to the PGRFA in the provider's collection. In the future, GLIS may use this information to locate the candidate Predecessor and establish the relationship using its DOI.

We also invite registrants to periodically revisit their Successors and check whether some Predecessor has now been given a DOI so that the relationship can be established.

**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>4</sup> See descriptors A01 and A02