Vademecum for DOI Registrants

1 What is this document about?

The adoption of Digital Object Identifiers[1] (DOI) in the Global Information System[2] (GLIS) of the International Treaty offers a powerful tool for accurately and permanently identifying Plant Genetic Resources for Food and Agriculture (PGRFA). However, the use of DOIs in this context is quite new and some practical information may be useful.

This document does not deal with how to register DOIs[3], rather it focuses on what you can do with DOIs once they have been assigned to your collection, briefly describing how DOIs can be used in a variety of real-world cases.

To keep this document short and focused, inevitably, we had to simplify some concepts. If you are interested in more details, please refer to the GLIS page on the Treaty website at

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

or send us an email at pgrfa-treaty@fao.org, we will be happy to provide all the clarifications you may need.

2 DOIs in SMTAs

The core purpose of the International Treaty is to facilitate access to PGRFA through the Standard Material Transfer Agreement[4] (SMTA), a standard legal contract between Provider and Recipient that makes it very easy to obtain material under the Treaty's framework.

One of the stipulations of the SMTA is that the Provider, together with the material, must make available to the Recipient "All available passport data and, subject to applicable law, any other associated available non-confidential descriptive information".

A very simple and effective way to be in compliance with this requirement is just to indicate, in the Annex 1 to the SMTA, the DOIs of the PGRFA you are shipping! Why? Because the Recipient only needs to enter the DOIs in the GLIS home page[5] to access the detail page on the DOI. This page contains:

---

[1] See https://www.doi.org for details
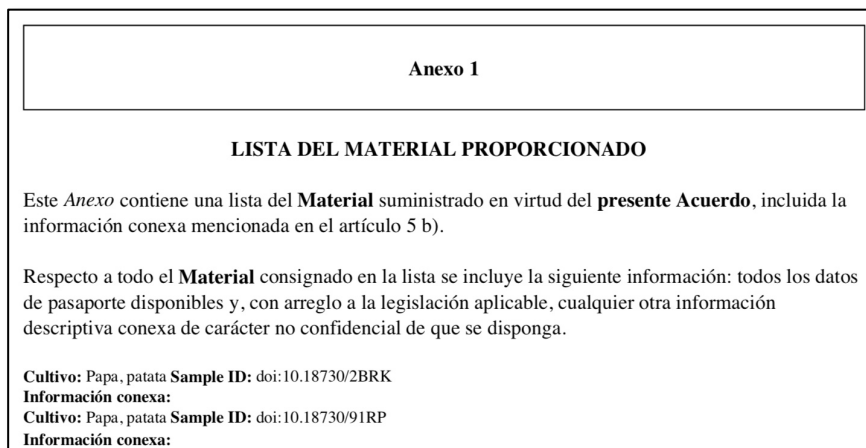
[2] See http://www.fao.org/plant-treaty/areas-of-work/global-information-system

[3] For this, just send an email to pgrfa-treaty@fao.org and we will be happy to provide all the assistance you may require

[4] See http://www.fao.org/plant-treaty/areas-of-work/the-multilateral-system/overview

[5] https://glis.fao.org/glis. Please note that the DOI can be entered in any DOI resolution service (e.g. https://www.doi.org itself) to get to the detail page in GLIS

- all the passport information you have provided when you registered the PGRFA
- all links to external systems provided by you (or others, such as Genesys, WIEWS or EURISCO) where additional information on the PGRFA can be found
- links to publications and/or datasets related to the PGRFA (more on this below)

An example of a SMTA with DOIs in Annex 1 is shown in the following Figure 1. Note that, to make it very clear that the Sample ID is a DOI, the prefix "doi:" is used.

Anexo 1

LISTA DEL MATERIAL PROPORCIONADO

Este *Anexo* contiene una lista del **Material** suministrado en virtud del **presente Acuerdo**, incluida la información conexa mencionada en el artículo 5 b).

Respecto a todo el **Material** consignado en la lista se incluye la siguiente información: todos los datos de pasaporte disponibles y, con arreglo a la legislación aplicable, cualquier otra información descriptiva conexa de carácter no confidencial de que se disponga.

Cultivo: Papa, patata **Sample ID:** doi:10.18730/2BRK
Información conexa:
Cultivo: Papa, patata **Sample ID:** doi:10.18730/91RP
Información conexa:

Figure 1: SMTA Annex 1 with DOIs listed as Sample ID

3 Citing DOIs in publications

One of the uses of PGRFA that the International Treaty most actively promotes is *scientific research*. Scientific research typically results in the dissemination of the findings through papers published in journals or books.

When a paper is published, it is good practice to provide an accurate and detailed list of the PGRFA that have been studied[6] so that readers can search the Web for additional information. Now, from what we have said in chapter 2 above, what better than a DOI can provide super easy access to the information associated to a particular PGRFA?

This alone is a very good reason to list the DOIs of the PGRFA the paper refers to, but there is more.

As you probably know, most publications are assigned a DOI by their publishers. The DOI System now offers a service that records the *citations* of DOIs by other DOIs. In other words, there is a way for GLIS to automatically identify publications citing its own DOIs and list them at the bottom of the page as shown in Figure 2 below!

---

[6] This is a particular instance of *data citation,* i.e. accurate, consistent and standardized referencing for datasets just as bibliographic citations are provided for other published sources like research articles

Figure 2: Publications and datasets citing this PGRFA (bottom of the page)

The underlying technology[7] is quite new and there are still a few glitches: for example, to make this work, it is currently necessary to list the GLIS DOIs in the publication *among other bibliographic references* as shown in Figure 3

Figure 3: GLIS DOIs among other bibliographic references

Sometimes, editors and publishers find this odd and might create difficulties, we are available to assist your authors in the discussion to get this through. In the meantime, we are actively working with the DOI System and publishers to identify alternative, more intuitive solutions to streamline this very useful service. We even offer the option of entering data into GLIS manually to streamline the process until it gets more widely spread. Just contact us at pgrfa-treay@fao.org!

---

[7] A blog post by Crossref and DataCite addressed to publishers explains how cited DOI can be passed for this to work: https://doi.org/10.5438/9456-sb49

4 Citing DOIs in datasets

Scientific research often involves processing data organized in *datasets*. Datasets usually are the foundation of the research, but they can also be its output, or some of its outputs. For instance, a DNA sequence of particular PGRFA is a dataset that is studied, and the outcomes of this investigation are published. The dataset itself should be assigned a DOI to become a valuable piece of information that can be easily found and maybe reused in other experiments. Datasets are published by depositing them in repositories[8] from which they can be accessed by interested users.

As for publications, the dataset's DOI can be related to the DOIs of the PGRFA whose information is organized in the dataset and this relationship can be identified by GLIS and made available to its users.

For this to happen, much like for publications, it is necessary that the GLIS DOIs of the PGRFA to which the dataset is related are provided when the DOI of the dataset is obtained. This is normally accomplished by filling in the repository's dataset deposit form that offers a suitable field in which the GLIS DOIs can be entered.

5 Receiving PGRFA that already has a DOI assigned

When you receive a PGRFA from another collection, maybe a DOI has been already assigned to it by the Provider. It is important to note that the DOI identifies the Provider's material and, among other information, reports the holder of the material (in this case, the Provider).

As PGRFA are living organisms, they can evolve and differentiate, especially when moved from one collection to another. Also, when a PGRFA is incorporated into a new collection, the Recipient becomes the holder of the received material. We recommend that a new DOI is assigned to the received material and such new DOI is related to the Provider's DOI, because:

- as said, the material may change over time and become progressively different from the original. Imagine a PGRFA that is transferred, say, 10 times from one Recipient to the next. It is not possible to guarantee that, after 20 years, the material at the end of the transfer chain is still precisely the same as the original. Therefore, to be able to properly identify the chain of transfers that led to the final material, it is necessary that a new DOI is assigned at each exchange and is related to its predecessor
- one of the mandatory descriptors associated to the DOI is the holder of the material. If a PGRFA is moved from collection A to collection B and from there to collection C, the only way in which we can know that C received the "version" of the PGRFA that was provided by B is to assign a new DOI at each step. Otherwise, it would appear that C received the material from A, i.e. the intermediary role of B would be completely obfuscated!
- if a Recipient publishes results on the material received, she should cite her own DOI, not the Provider's. Otherwise, it would appear that the study was carried out on the Provider's material that, as we have just seen, could be different from the original

Please note that Recipients are free not to register new DOIs for the material[9] they receive (and therefore use the Provider's DOI) or register a new DOI without referencing the Provider's DOI as ancestor, but this would lead to the issues described above.

---

[8] Example of public repositories are Zenodo (https://zenodo.org) or Figshare (https://figshare.com), but many institutions publish their datasets on their own websites

[9] Document "Guidelines for the optimal use of Digital Object Identifiers as permanent unique identifiers for germplasm samples" (http://www.fao.org/plant-treaty/areas-of-work/global-information-system/guidelines) explicitly lists this option (not assigning a new DOI to the material but rather use the DOI registered by the Provider) as the "default preferred option for many general users" (see paragraph 10 Implementation for recipients of PGRFA). However, it also lists the issues that this choice would potentially imply.

Establishing the chain of relationships is also very useful for Providers as they will be able to see what is being done with their material, e.g. where it is transferred to and what publications and/or datasets are being produced studying it.

How can you reference the Provider's DOI when registering a new DOI for the material you just received? The answer is very simple, just indicate the Provider's DOI in descriptor `R02: DOI(s) of progenitor(s)` and specify `Acquisition` in descriptor `M04: Method`.

When proper relationships are established between a PGRFA DOI and the DOI(s) of its progenitor(s), GLIS is able to build a graph, such as the tiny example shown in Figure 4, that can be navigated to look for ancestors or descendant of the current PGRFA.

In the figure, we show two initial PGRFA (A and B, they are colored differently because we imagine them being conserved by different institutions). A and B are then transferred to other institutions (and new DOIs, C and D, are assigned and related to their predecessors A and B). C is further transferred to the final institution (green nodes) and is associated a new DOI, E. In the meantime, the same institution performs a selection on D obtaining a new PGRFA, identified by DOI F. Then E and F are crossed to create a new PGRFA that will be identified by DOI G. As you can see, how G was obtained is fully documented, something very valuable, especially for breeders and scientists.
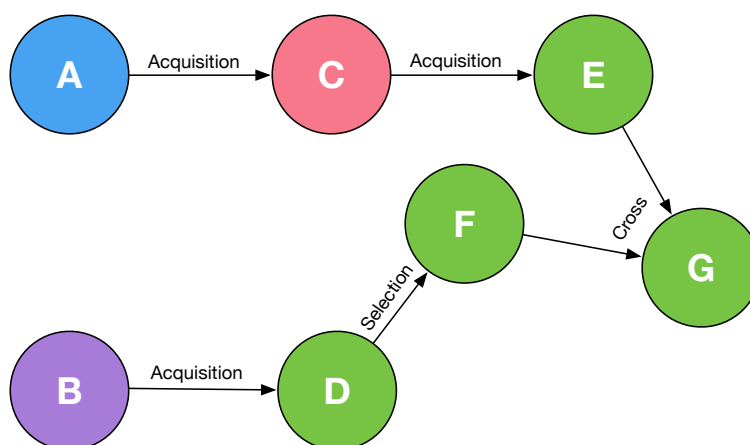

Figure 4: PGRFA ancestry as DOI graph

6 Creating new PGRFA

Genebanks and breeders routinely process PGRFA in ways resulting in material that is genetically different from its predecessor(s) or that somehow needs to be discriminated from its ancestry. Assigning a new DOI to this material (and relating it to the predecessor(s)) is a very powerful yet convenient way of tackling these cases. The new DOI ensures that the new material is identified as different from the predecessor(s) yet it is related to it (or them) so that the ancestry of the material can be easily obtained.

The Guidelines[10] and the Descriptors[11] documents describe in detail the real-world cases and the applications of each `Method` (descriptor `M04`). Depending on your specific case, when registering the DOI for the new material, indicate that `Method` in descriptor `R04` and list the DOI(s) of progenitor(s) in descriptor `R02` to correctly relate the new DOI to its ancestry.

Please refer to Figure 4 above for an illustration of this in relation to how F and G are obtained.

---

[10] See http://www.fao.org/plant-treaty/areas-of-work/global-information-system/guidelines

[11] See http://www.fao.org/plant-treaty/areas-of-work/global-information-system/descriptors

7 Conclusion

The adoption of DOIs in GLIS has opened a whole lot of new possibilities, such as:

- *improved data quality* Documentation systems are being updated to accommodate the DOIs and new releases of popular Genebank Management Systems, such as GRIN-Global, now support DOI registration to GLIS. Also, information on PGRFA is being validated and amended as required in view of DOI registration
- *facilitated system interconnection* We now have a single and permanent identifier that can be reliably used across systems to point to the same material
- *promoted accumulation of information* Information scattered across different systems can now be reliably associated to the same material and becomes available to users
- *increased visibility of your work* GLIS promotes access to information *where it is created and maintained* through links to web sites, publication and datasets. This means more visibility for your collection, your website, and your scientific research