Food and Agriculture Organization of the United Nations

The International Treaty ON PLANT GENETIC RESOURCES FOR FOOD AND AGRICULTURE

# Guidelines for the optimal use of Digital Object Identifiers as permanent unique identifiers for Plant Genetic Resources for Food and Agriculture - v.2

20 July 2017

## 1. Introduction

These guidelines are based on a broad consultative process and describe the main features and benefits of Digital Object Identifiers (DOIs) associated to Plant Genetic Resources for Food and Agriculture (PGRFA) and a set of basic principles for users to determine when to assign them.

This document along with *Data required for the assignation of DOIs in GLIS* [http://www.fao.org/plant-treaty/areas-of-work/global-information-system/descriptors/en/] serve as reference guides for the effective use of DOIs.

## 2. Background

Several communities[1] have highlighted the importance of creating and adopting Permanent Unique Identifiers for improved identification of PGRFA[2]. The reasons cited include the difficulty of collaboration on conservation, research and breeding without a common standard for identification, and the difficulty of finding information associated with the material. Following broad consultation, DOIs were selected as the most appropriate, web-resolvable digital identifiers.

In accordance with the requirements set out in the International Treaty on Plant Genetic Resources for Food and Agriculture (the Treaty) for the Global Information System (GLIS), the system under construction will (1) build on and facilitate linkage between existing systems and (2) allow for registration of DOIs applicable to all types of PGRFA. In addition, (3) GLIS will not replace existing systems or duplicate their functionality but provide new services needed by the user community and missing from existing systems, (4) DOIs will be easy to implement, (5) GLIS will also accommodate DOIs created by other systems, and (6) use of the system will be voluntary. Registration of DOIs for PGRFA will be voluntary,

[1] Including the genebank community, genomics community, plant breeders and journal editors.
[2] The Treaty defines PGRFA as "any genetic material of plant origin of actual or potential value for food and agriculture". This broad definition encompasses not only accessions conserved in genebanks and PGRFA conserved *in situ*, but also breeding lines, research materials, and protected modern varieties.

and, except for a small number of essential metadata descriptors, most descriptors are voluntary.

It is hoped that DOIs will become the global standard for public identification of PGRFA, facilitating linkage between the material and diverse sources of information associated with the material.

## 3. Bringing new opportunities

The functionality provided by the DOI system brings the following new opportunities for users:

- It exposes the material to the public or collaborators in a format that can be resolved by humans as well as computers.
- It enables information on the material to be harvested by robots searching publications and online databases that refer to the PGRFA by its DOI, and thus to be made more readily available.
- It facilitates access to the information about the PGRFA and related PGRFA by pointing to websites and systems where detailed information is created, maintained and made available to the public.
- It provides a simple way for recipients of material under an SMTA to comply with their obligations under the provisions of Article 6.9 [3] of the SMTA, simply by using the DOI in their publications and online datasets to refer to the material received.
- It helps developers of PGRFA to maintain their records and comply with their obligations under SMTA article 6.5b[4].
- It enables families of related PGRFA to be identified and thus jointly searched. For example, it enables a genebank manager to easily find all publications and online datasets created by recipients of accessions from the genebank.
- It provides a simple reliable mechanism to identify accessions that are duplicated across genebanks.
- It facilitates interoperability between databases, by providing a single common standard for sample identification used by all communities.
- It enables collaborating laboratories, should they wish, to track samples between them with any appropriate degree of precision while each laboratory continues to use its own in-house sample tracking system, providing assurance that they are working on the same material.

The precision that a holder of PGRFA needs for the above functions is a primary criterion in acquiring DOIs for PGRFA.

---

[3] SMTA article 6.9 states "The Recipient shall make available to the Multilateral System, through the information system provided for in Article 17 of the Treaty, all non-confidential information that results from research and development carried out on the Material"

[4] "In the case that the Recipient transfers a Plant Genetic Resource for Food and Agriculture under Development to another person or entity, the Recipient shall […] (b) identify, in Annex 1 to the new material transfer agreement, the Material received from the Multilateral System […]".

## 4. What is identified?

DOIs can be used to identify PGRFA held by any individual or organization[5], including genebanks, plant breeders, geneticists, other plant scientists, extension officers, seed companies, plant variety protection offices, gardeners, farmers, landowners, and land managers.[6]

At the discretion of the holder, within the guidelines set out here, the material identified can be any entity recognized as such by the holder. It may be a single DNA sample extracted from a plant, or a single seed or plant or plantlet, or a seed lot contained within a single packet or set of plantlets in one tissue culture tube, or the whole seed lot or set of clonal material harvested from a plot or field, or even multiple generations. The material can be an F1 hybrid, a segregating population, a pure line selected from a mixture or from a segregating population, a mixture of pure lines, or any other genetically homogenous or heterogeneous entity. It may be a landrace or other genetically heterogeneous variety, a modern released cultivar, a genebank accession. It may be formally conserved, for example as in a genebank, or have a transitory existence.

It is therefore essential to include with the DOI information on the nature or category of the PGRFA being identified. The key aspect of this categorization is the real world event that resulted in the PGRFA becoming an entity managed by the holder, such as the collection of a sample from *in situ* conditions, the accession of a sample into a genebank collection, the creation (harvest of seed) of a genetically distinct sample by breeding, the registration of a cultivar in a country, or the first documentation of presence of the PGRFA in a natural habitat. This event is known in the DOI metadata as the method of acquiring or creating the PGRFA, one of the few mandatory descriptors.

The DOI identifies the material itself, not the associated data. This is an important distinction. *Inter alia*, it means that if the data associated with the material change, the holder should correct the data without changing the DOI.

## 5. How is the material identified?

GLIS is not intended to replace existing information systems, and therefore does not replace existing systems for identifying PGRFA. Existing identifiers will continue to be used. In a publication or online article, the first reference to the material would include both its DOI and the local identifier normally used by the holder; subsequent references within a single publication may specify only the local identifier.

---

[5] The identity of the "holder" is not necessarily clear. For example, in the case of material managed *in situ* or on farm, the holder may be an individual land owner or manager, or the local community jointly responsible for the land, or an organization that owns or manages the land. The decision will depend on the local context, and the DOI system accommodates any of these possibilities.

[6] It follows immediately that the DOI identifies PGRFA within the context of the individual or organization that holds it, and thus one DOI maps to the combination of (1) the identity of holder of the PGRFA with (2) how the holder identifies the physical PGRFA material among all other PGRFA held by the holder.

However, when PGRFAs are transferred across organizations, locally assigned identifiers have been insufficient to uniquely identify the material. A globally unique, persistent identifier such as a DOI is preferable to maintain consistency over time, provide for proper recognition of rights and obligations, and facilitate access to research outcomes contributed by subsequent recipients of the material.

The assigned DOI should be used to identify the material publicly, especially in electronic media that can be searched online.

## 6. The PGRFA holder's commitment

A holder of PGRFA who obtains a DOI for a sample of PGRFA makes a commitment to associate that DOI permanently with the material, and not to use the same DOI for any other PGRFA.

Obtaining a DOI does not require or imply any commitment by the holder to maintain the PGRFA alive, and does not change any commitment the holder may or may not already have. If the PGRFA dies or is lost, the DOI persists as an historical record and will not be reused for other PGRFA. This way, any information accrued when the material was available can still be accessed.

Obtaining a DOI does not require or imply any commitment by the holder to make the PGRFA or associated data available to others, and does not change any commitment the holder may or may not already have about the material or the associated information.

## 7. Relationship with existing systems

Many holders of PGRFA have some form of inventory management system and/or workflow system, with provision for quality control, tracking, data collection and data management. For genebanks this requires the documentation of accessions and their origins, maintaining records of viability, health, genetic integrity and quantity of seeds or clones, tracking progress through viability tests, characterization, and growing out to rejuvenate or multiply stocks, e.g. GRIN-Global. For plant breeders it requires identification and tracking progress through crossing, selection, multiplication, evaluation and release, e.g. Breeding for Results (B4R) being promoted by the CGIAR Excellence in Breeding Platform. For the seed industry it involves tracking progress through seed maintenance from breeders' seed to seed sold to farmers, with appropriate verification of genetic identity. These systems are the primary sources of information on PGRFA, and would normally be the primary sources of data for GLIS.

In addition, some communities have developed portals to expose data on the material they hold and to enable users to search those data. These are typically data warehouses, or secondary data sources, populated by the holders of PGRFA from their primary sources. An example is Genesys (www.genesys-pgr.org), enabling the public to search for accessions available in participating genebanks across the world. Another is the FAO's World

Information and Early Warning System on Plant Genetic Resources for Food and Agriculture (WIEWS: http://www.fao.org/wiews).

GLIS is not designed to replace any of these systems, and hence does not duplicate their functionality. Genebanks, breeders and others lacking such functionality may acquire it through relevant capacity-building initiatives.

In order to link existing systems, GLIS needs to maintain in a central repository the data required to identify the corresponding records in the corresponding system. These are the mandatory descriptors (holder, local identifier, scientific name or crop name, method, date), which should be uploaded by the holder of the material from a primary data source.

The DOI system makes no assumption about the nature of the documentation system of a holder of PGRFA. It assumes only that the holder can identify the PGRFA held with sufficient precision and permanence to meet the holder's commitments as described above.

## 8. When to obtain a DOI

For holders who have opted to use DOIs to publicly identify the PGRFA that they hold, the basic rule for when to obtain a DOI for a PGRFA is simple:

***assign a DOI to any PGRFA that you manage and whose present or past existence you wish or need to make publicly known, at the level of detail you choose***

and, as a corollary

*do **not** assign a DOI to any PGRFA whose present or past existence you do **not** wish or need to make publicly known*

The DOI system provides for a pre-publication embargo period, during which a DOI is assigned but not published. This opens the possibility of assigning DOIs to PGRFA without immediately making their existence publicly known.

Establishment of more specific rules requires consideration of the reasons, or use cases, for making known the existence of the PGRFA. Many different use cases have been identified, but many are specific to particular stakeholder groups. For simplicity, the following sections suggest standards for implementation by specific groups. These may be refined and self-enforced as appropriate to each group.

## 9. Implementation for genebank managers

The primary objectives of genebank managers are to conserve PGRFA as accessions and provide samples of those accessions to users. They need to conserve rationally (avoiding unnecessary duplication and undesirable gaps), and to facilitate rational use (facilitating the searching of associated information to find the best match between the material available for distribution and material needed by users). They need to inform potential users about the existence of their accessions, and they need to provide descriptive information about their accessions, preferably including information collected by their users. Depending on the crop,

one accession may be genetically more or less heterogeneous, and precise monitoring of genetic identity correspondingly difficult.

Hence for genebanks one DOI would typically correspond to one available accession. Standard basic practice for genebank managers would therefore be:

- Obtain a DOI for each accession that is available to users. If known and not confidential, specify also the DOI of the progenitor of the accession. If the accession was acquired from another genebank, the progenitor would be the DOI of the accession in the provider's genebank: hence there could be two DOIs for samples that are intended to be the same genetic material, but conserved by different genebanks. Alternatively, if the providing genebank's accession has no DOI or its DOI is not known, the provider's accession identifier may be specified.
- If the genetic composition of an accession is changed, whether deliberately or unintentionally, such that the new material would be registered as a new accession, obtain a new DOI for the new material. This could include separating a mixed accession into its components, selecting a pure line out of an accession, or discovering a mislabelled sample.
- Different samples of an accession held by the same genebank would normally share the same DOI, unless the genebank manager has a special need to identify the specific samples publicly. The progenitor DOI must be specified, to record that the sample is part of the accession.
- Remind recipients that SMTA Art 6.9 obliges them to make available, through the GLIS, all non-confidential information that results from research and development carried out on the material received, and inform recipients that they can[7] fulfil this obligation by using DOIs registered in GLIS in all their publications and public datasets. This will help the genebank manager associate users' results to the provider's material.

## 10. Implementation for recipients of PGRFA

Recipients of PGRFA have widely differing needs and capacities. For some users, precise monitoring of genetic identity of the variants can be critically important; the intrinsic genetic variability of genebank accessions can pose significant challenges, and they may need more precise identification than just an accession ID. For other users, this genetic variability may be desirable or unimportant. Some users have their own advanced computerised systems for managing and identifying PGRFA and associated information; others may have nothing.

In the case that a recipient receives material for which the provider has already obtained a DOI, GLIS offers the recipient three options to identify the material received:

1. Use the DOI for the material as registered by the provider;

[7] Under SMTA Art 5a, the provider cannot oblige recipients to make their results available to the provider. SMTA Art 6.9 does not state how recipients should make their results available through the GLIS. Hence the provider can do no more than offer DOIs as a mechanism to comply with the recipient's obligations.

2. Obtain and use a new DOI to identify the recipient's sample as a distinct entity from the provider's material, and specify that the DOI of the provider's sample is the progenitor of the new DOI;
3. Obtain and use a new DOI to identify the recipient's sample, without specifying the DOI of the progenitor.

Option 1 could be considered the preferred default option for many general users. It would enable associated information to be collated through GLIS without regard for the source of data. However, the principle is similar to asking recipients to provide reference to the provider's accession ID, which has met with limited success in the past. Moreover, it comes with a number of consequences which may be undesirable in some cases:

- Any online data or online publications by the recipient will be associated directly to the provider's material.
- Attribution of the data to the user will be possible only if an independent mechanism is developed to identify the source of data.
- Similarly, separating data created under the control of the provider from other data will be possible only if an independent mechanism is developed to identify the source of data.
- Similarly, if a problem is detected with quality control or standards or genetic integrity, it not will be possible to identify the extent of the problem unless an independent mechanism is developed to do so.
- If the user obtains multiple copies of the same accession from the same provider, it will not be possible to distinguish between the copies unless an independent mechanism is developed.
- If the user distributes the material to a third party using the original provider's DOI, and the third party assigns a new DOI to the material received, it will be linked to the original provider's DOI, and the role of the original recipient as intermediary will not be public.

Option 2 is preferred when any of the above consequences of option 1 would cause problems. It would be preferred by recipients that want or need separate public identification for material under their management, or recognition through the DOIs of the associated information that they publish. It would be preferred in partnerships where provider and recipient use independent data management systems and need to track exchanges of samples, and wish to do so through DOIs. The optional link to the provider's DOI would enable full flexibility over the scope of searches for information associated with the genetic material: searches could be limited only to the provider's DOI or only to the recipient's DOI, or could include both.

Option 3 would be for recipients who want to make public the existence of their sample and/or provide associated information, but who do not want to publicly disclose the source of their material.

## 11. Relationships between DOIs

The core novel function of GLIS through the implementation of the DOI System will be to point to various existing databases as a stable and unique reference. The DOI Module of GLIS has the capacity to establish relationships between DOIs, thus connecting records across systems. They are of two kinds:

- A DOI for a PGRFA is related to DOIs for resources containing information about the PGRFA. The latter include online datasets and publications containing data and information about the PGRFA. The user may explicitly declare them within GLIS as "links to associated information" (see the descriptor R01). In addition, GLIS will systematically trawl the web for resources containing references to the DOI for the PGRFA, and will automatically add these to the DOI's links to associated information. This will enable users to easily discover online data and information associated with the PGRFA.
- A DOI for a PGRFA is related to its progenitor(s), which, being themselves PGRFA, may also have a DOI (see the descriptor R02 DOIs of progenitors). The genetic relationship between a PGRFA and its progenitors can be one of several types depending on how the PGRFA came into existence (see descriptor M04 Method): the DOI may be a genetic copy of, or a variant of, or novel PGRFA incorporating, its progenitor(s). This will enable users to search whole sets of PGRFA: for example, a set of PGRFA that are at least intended to be copies, or the set of PGRFA that are variants of one specified PGRFA, or the set of PGRFA that incorporates one specific progenitor.

  In the case that a provider transfers a PGRFA to a recipient and the recipient chooses to obtain a separate DOI using the GLIS tools, the provider's DOI will be automatically identified as the progenitor of the recipient's DOI. This will help assuring the accurate documentation of transfers of PGRFA between providers and recipients.

  In the case that a holder of PGRFA changes the genetic composition of a PGRFA, or wishes to ensure against possible changes in genetic composition (see the next section), it will ultimately be the user's responsibility to ensure that each DOI is correctly associated with its progenitor(s), although GLIS will provide tools to help the user.

## 12. Managing and using DOIs

- Germplasm holders will prepare to adopt DOIs by adding an additional field in their database that will receive the DOI assigned to each eligible material.
- The Treaty Secretariat will provide an easy-to-use software toolkit that will facilitate the assignment of DOIs and the uploading and correction of associated data.
- If a holder loses a sample for which a DOI has been assigned, the status of the DOI can be changed on the GLIS server to "historic".
- The holder is encouraged to use the DOI in all publications and online articles and databases containing data collected on the germplasm. In a publication or online

article, the first reference to the germplasm should include both its DOI and the local identifier normally used by the holder; subsequent references within a single publication may specify only the local identifier.

Further information on the Global Information System and DOIs can be found in FAQs: http://www.fao.org/plant-treaty/areas-of-work/global-information-system/faq/en/