

October 2016



Food and Agriculture
Organization of the
United Nations



The International Treaty
ON PLANT GENETIC RESOURCES
FOR FOOD AND AGRICULTURE

Item 5 of the Provisional Agenda

INTERNATIONAL TREATY ON PLANT GENETIC RESOURCES FOR FOOD AND AGRICULTURE

FIRST MEETING OF THE SCIENTIFIC ADVISORY COMMITTEE ON THE GLOBAL INFORMATION SYSTEM OF ARTICLE 17 OF THE TREATY

Rome, Italy, 24 – 25 November 2016

IMPLEMENTING DIGITAL OBJECT IDENTIFIERS (DOIs)

I. INTRODUCTION

1. This document describes the process that was adopted to select a suitable Permanent Unique Identifier (PUID) scheme to be adopted in the Global Information System (GLIS) of the Treaty. Furthermore, it provides an overview of the current status of the implementation of Digital Object Identifiers (DOIs).
2. The document also identifies and brings to the attention of the Committee some technical issues that need to be considered and addressed for the optimal further development and implementation of DOIs.

II. THE NEED FOR PERMANENT UNIQUE IDENTIFIERS

3. The accurate, permanent identification of the material is the cornerstone of any research activity on Plant Genetic Resources for Food and Agriculture (PGRFA). Information obtained during research and development of the material becomes valuable and reusable when there is no doubt about the identity of the material being studied. This critical requirement has been addressed at the local level by the community of users in the past but no general consensus has been reached on to which one, among the many available PUID schemes, to adopt.
4. In September 2014, the Treaty Secretariat received several requests from Contracting Parties and a regional PGRFA network to facilitate technical discussions on the matter and indicate what kind of PUID would be most appropriate for the community's needs.
5. During the following months, an extensive exploration of the most promising PUID technologies available was carried out by the Secretariat. The outcome of such work was presented at the "Consultation on the Global Information System on Plant Genetic Resources for Food and Agriculture" held in San Diego, USA, from 7 to 8 of January, 2015 through working document IT/COGIS-1/15/3.¹

¹ Available at: <http://www.fao.org/3/a-be643e.pdf>

6. The Consultation restricted the list of PUID schemes to be further investigated to Archival Resources Keys (ARKs), Digital Object Identifiers (DOIs) and Life Science Identifiers (LSIDs). Such further investigation resulted in a set of 20 indicators against which a group of 17 designated experts with a wide range of expertise ranked the three candidate PUID schemes. The results of the consultation, held between end of February and March 2015, were presented in the report² of the meeting and concluded that Digital Object Identifiers (DOI³) was the most suitable PUID technology for GLIS.

7. The outcome of the consultation was not unexpected. DOIs have, generally, been considered among the best PUID technologies available. However, the high cost associated to minting DOIs has considerably hampered their adoption. The next step, then, was to provide DOI minting services to the PGRFA community free of charge, effectively removing such hurdle. In recent years, the cost of DOI minting has been considerably reduced also thanks to the promotion strategy adopted by the DOI Foundation.

8. At the beginning of 2016, FAO concluded a service agreement with DataCite, a registration agency whose users are actively involved in the agency's strategy. This makes it easier for the Treaty to raise any particular issues of interest to the PGRFA community to be taken into account in the evolution of the services offered by DataCite. The current service agreement enables the Secretariat to mint an unlimited number of DOIs for a fixed annual fee.

9. DataCite focuses on the interactions among various aspects of the research activity and its results, and offers innovative services that will be quite relevant for GLIS. One example is the Event Data through which GLIS would be able to automatically identify the publications where its own DOIs are cited.

What is a Digital Object Identifier associated to?

10. The first critical decision that needed to be taken was with regard to what a DOI actually identifies. According to the DOI Handbook:

“A DOI name can be assigned to any object whenever there is a functional need to distinguish it from other objects.”⁴

“The choice of whether to assign a new DOI name to a changed item which already has a DOI name is one of functional granularity: it should be possible to identify an entity whenever it needs to be distinguished.”⁵

11. It was considered that a GLIS DOI will be associated to the physical PGRFA material and *not* to its description. The main reasons for this choice are:

- the description of the material may be incorrect. If the DOI is associated to the description, changes to it would require a new DOI to be minted. If the DOI is associated to the material, its description can be amended freely without minting a new DOI
- the description of the material may change over time, for instance, due to taxonomy changes.

12. Furthermore, as the plant genetic material is transferred, the different rights, obligations and quality control standards under which the material is put when it is incorporated in the Recipient's collection, warrant the assignation of a new DOI even if, at least in theory, the material is the same.

² See <http://www.fao.org/3/a-bq555e.pdf>

³ See <http://www.doi.org/>

⁴ See DOI Handbook, Chapter 2.3.1 Principles of assignment, <https://www.doi.org/hb.html>

⁵ See DOI Handbook, Chapter 2.3.3 Changed items, <https://www.doi.org/hb.html>

13. The Provider's DOI and the Recipient's DOI must be associated through appropriate relational operators provided for in the DOI metadata architecture facilitating the connection of information produced on the material across subsequent transfers and holders. This feature was favourably appraised by the institutions interested in assigning DOIs to their PGRFA.

III. DESCRIPTORS ASSOCIATED TO THE DIGITAL OBJECT IDENTIFIER

14. In order to assign a DOI to a PGRFA sample, it is necessary to provide some description of the material itself for two reasons:

- when a user searches for a DOI in the GLIS, a meaningful set of information should be returned describing the material to which the DOI is associated. This function is called *resolution*;
- on the other hand, if the user does not know the DOI but knows some of the attributes of the material being sought (e.g. the genus and species), he/she should be able to enter such information in a search form and obtain the list of matching records. This function is called *discovery*.

15. A set of descriptors to be provided when requesting the assignation of a DOI to some PGRFA material was identified. In order to minimize the friction of adopting the GLIS DOI Registration Service, it was considered best to divide the descriptors into three classes:

- mandatory, i.e. those descriptors essential to the identification of the material that must be provided in order to assign a DOI. An example of a mandatory descriptor is where the material is maintained (e.g. the name and address of the holding institution);
- highly recommended, i.e. those descriptors that should be provided if known in order to enrich the description of the material and facilitate the discovery function. An example of such a descriptor would be the biological status of the material;
- context descriptors, i.e. information that should be provided if available and that depends on the value of some other mandatory or highly recommended descriptor. For example, if the material was collected, some information on the collecting mission.

16. Recalling the wide adoption of the Multi-Crop Passport Descriptors in the genebank sector,⁶ this format was adopted as a reference model and expanded to take into account the needs of other user groups.

17. The resulting set of descriptors was then submitted to a group of over 200 experts from 61 countries nominated by Contracting Parties and stakeholders as well as invited for their recognized expertise in the field. The survey took place in July and August 2015 and was announced through a formal notification and widely circulated by email. The results were summarized in a report,⁷ which was circulated afterwards.⁸

18. Further work during 2016 led to the publication of two technical documents:⁹

*Data required for the assignation of Digital Object Identifiers in the Global Information System - v.1.*¹⁰ It describes the minimum data to be provided when registering samples in GLIS;

⁶ See <http://www.biodiversityinternational.org/e-library/publications/detail/faobiodiversity-multi-crop-passport-descriptors-v21-mcpd-v21/>

⁷ Report of Global survey on descriptors required to register material in the Global Information System on Plant Genetic Resources. Available at: <http://www.fao.org/3/a-bp470e.pdf>

⁸ NCP GB7-009 DOIs Consultation. Update and inputs for the implementation of the Global Information System. <http://www.fao.org/3/a-bp423e.pdf>

⁹ Both documents have been made available as information documents for this meeting.

*Guidelines for the optimal use of Digital Object Identifiers as permanent unique identifiers for germplasm samples - v.1.*¹¹ It outlines the main features and benefits of DOIs associated to samples of plant genetic resources for food and agriculture as well as a set of basic rules and useful advice for users to determine when to assign them. It also contains a few use cases.

19. Both documents will allow the user community to more precisely and permanently identify plant genetic resources for food and agriculture, especially the material available in the Multilateral System, as well as better model the relations between information referring to the same material when transferred. While the first version of the guidelines is a major orientation resource, particularly for genebanks, further user cases need to be identified. In this context, the collaboration with the Indonesian Center for Agricultural Biotechnology and Genetic Resources Research and Development (ICABIOGRAD) through the project funded by the Benefit-sharing Fund of the Treaty is expected to be helpful during the testing with breeders.¹²

IV. FURTHER IMPLEMENTATION OF DIGITAL OBJECT IDENTIFIERS (DOIS)

Mapping of GLIS descriptors to DataCite metadata

20. The Secretariat of the International Treaty has also been working on the mapping of the descriptors to the DOIs metadata structure. The mapping was done with the support of DataCite experts and is in compliance with the upcoming version 4 of the DataCite metadata structure.

21. Future changes in the GLIS descriptors may require changes in the mapping, for example necessitated as a result of collaboration with other users communities.

Accepting existing DOIs

22. Since the announcement of DOIs as the scheme of choice for GLIS, the Secretariat has become aware of the exploratory research work of other institutions and organizations about minting DOIs for PGRFA on their own as many research institutions already assign DOIs to publications and/or datasets. One example is the French National Institute for Agricultural Research (INRA), which, in order to ensure consistency, worked to align its own descriptors with the ones identified for GLIS.

23. The GLIS DOI Registration Service will be able to accept DOIs assigned to PGRFA samples by other organizations. However, to maximize the effectiveness and compatibility among all DOIs assigned to PGRFA, a major effort should be made by the community for the GLIS descriptors and the DOIs to be applied consistently.

The DOI Registration Service of GLIS

24. The GLIS DOI Registration Service is the first module that has been addressed for design and development, because of the clear and immediate needs expressed by the community. The effort is being carried out in collaboration with FAO's Information Technology Division that is providing architectural and technological advice in order to ensure corporate support for the system.

¹⁰ Available at: <http://www.fao.org/3/a-bp767e.pdf>.

¹¹ Available at: <http://www.fao.org/3/a-bq549e.pdf>.

¹² BSF-funded project W3B-PR-29-Indonesia "Multi-country construction of a test platform for the development and allocation of globally unique identifiers for rice germplasm". <http://www.fao.org/plant-treaty/areas-of-work/benefit-sharing-fund/projects-funded/bsf-details/en/c/359491/?iso3=IDN>

25. User accounts for the GLIS DOI Registration Service are the same as those managed by the PID Server component of Easy-SMTA that already contains user accounts for Providers and Recipients of PGRFA transfers done through the Treaty's Standard Material Transfer Agreement (SMTA). However, access to DOI registration will be granted to an existing Easy-SMTA account only upon request.

26. The Secretariat has received several expressions of interest from genebanks in Contracting Parties regarding the DOI registration service and the participation in the pilot phase as well as from some PGRFA networks and projects.

V. OTHER FUNCTIONS AND SERVICES

27. The GLIS DOI Registration Service will offer a web user interface through which it will be possible to register new material, update existing records, resolve DOIs and search material according to its descriptors.

28. The web user interface will be provided in the six official UN languages while provisions could be made to crowd-source the translation into even more languages. Availability of multiple languages is considered a crucial step towards making GLIS more attractive to as wide an audience as possible.

29. The manual registration of material offered by the web user interface is expected to be feasible only for very small collections. In order to accommodate the needs of larger organizations, an advanced WebServices registration Application Programming Interface (API) is being designed.

30. It is likely that a batch, file-based registration function would also be offered to support mid-sized collections that would find the manual registration impractical and the WebServices API too difficult to implement. This option, as already experienced with Easy-SMTA, is time-consuming and error prone and should be regarded only as a stopgap solution. The Connection Toolkit, described below, is regarded as an effective strategy to overcome this problem as well as provide additional value-added services to adopting institutions.

31. GLIS will also offer a query API, based on the REST model,¹³ with advanced functions such as rate limitation, content negotiation and response expansion. Adequate security measures are taken into account to prevent unauthorized use.

Further work on targets

32. Once a PGRFA is registered in GLIS, it becomes possible to associate links to resources on the Web where detailed information on the material can be found: this is what is called *targets*. Examples of targets are URLs to genomic data sources, original records regarding the field collecting missions of the material, characterization and evaluation studies and so on. Targets could also include DOIs of publications, links to pictures, to records describing indigenous knowledge and to any other resource considered valuable.

33. Additionally, when two or more PGRFA are related (e.g. because of a transfer, selection or some other genetic process), the system will display the list of targets associated to PGRFA related to the current one (we call these *related targets*). This would be very useful, for instance, to locate information made available by the recipient of material with the SMTA.

34. The wide variety of resources that can be made available suggests that targets should be classified using some controlled vocabulary, or preferably an ontology, describing what kind of information will be found in each one. The purpose of such classification would be to allow the

¹³ REST, Representational State Transfer. It is an architectural style and an approach to communications widely used in the development of Web services.

user (or the client) to filter the links presented in the GLIS page for the PGRFA material according to his/her needs: for instance, only show links with information on genomics. This facility is critical because there could be many links associated to a given PGRFA sample (and many more associated to all the related samples) and it would be impractical to visit them one by one to find out what kind of information each one offers.

The Connection Toolkit

35. While the GLIS APIs will be fully documented so that interested parties can proceed with their own implementation, to further reduce the cost that participating institutions will incur in integrating with GLIS, a lightweight and easily deployable Toolkit is proposed that will:

- offer a wide range of integration options such as simple file transfer, JDBC/ODBC access to the local database or WebServices;
- implement all necessary GLIS transactions;
- implement all necessary SMTA reporting transactions;
- allow to expose local datasets in a variety of Open Data formats.

36. The Connection Toolkit, to be developed in collaboration with the Information Technology Division (CIO) of FAO, will be deployed by participating institutions and easily configured to integrate with the local environment. The Toolkit will be maintained by the Secretariat to ensure compliance with any future version of the GLIS and Easy-SMTA system integration protocols.

37. The idea of the Toolkit has been very favourably received by small institutions (its primary target), as well as larger institutions unable or unwilling to invest resources in this activity. The first deployment of the Toolkit will be in the context of the Benefit-sharing Fund led by Indonesia. The experience acquired in that context will be used to validate and refine the GLIS API.

VI. ADVICE SOUGHT

38. The Scientific Advisory Committee is invited to take into consideration the information presented in this document and advise on:

- i) opportunities to discuss and validate the DOI descriptors and the guidelines, e.g. what organizations or institutions could be most representative of specific user communities;
- ii) standard formats and protocols that should be supported by GLIS;
- iii) further work on controlled vocabularies to be considered for adoption in the classification of targets and links to other resources;
- iv) challenges and opportunities to promote the adoption of DOIs with the PGRFA community; and
- v) any other technical issues that could facilitate the effective operation and wide implementation of Digital Objects Identifiers for PGRFA under the Global Information System.