



Food and Agriculture
Organization of the
United Nations



The International Treaty
ON PLANT GENETIC RESOURCES
FOR FOOD AND AGRICULTURE

Summary report of the Task Force on Permanent Global and Unique Identifiers for PGRFA in the context of the Global Information System of Article 17

Background and Terms of Reference of the Consultation

The electronic Task Force on the Permanent Global and Unique Identifiers (PUIDs) was a recommendation of the Expert Consultation on the Global Information System (COGIS-1) which was organized by the International Treaty on Plant genetic Resources for Food and Agriculture in San Diego on 7 and 8 January 2015. The experts in San Diego, nominated by the regions and the Secretariat of the International Treaty, reviewed the paper *Technical Options to Facilitate the Establishment of Data Links in the Field of Plant Genetic Resources for Food and Agriculture: Permanent Unique Identifiers*¹ and advised on additional research to be undertaken by the Secretariat in consultation with appropriate parties and relevant stakeholders for the elaboration of standards.

In its Report², the Consultation recognized the urgency of adopting some sort of Permanent Unique Identifiers (PUIDs), as well as best practices and methodologies for their deployment as an essential element for the Global Information System (GLIS) on Plant Genetic Resources for Food and Agriculture.

In January and February the Secretariat undertook additional research on the candidate identifiers discussed at the Expert Consultation, namely Archival Resource Key (ARK), Digital Object Identifier (DOI) and Life Science Identifier (LSID), and called for an online Task Force with the experts nominated by the participants of the Consultation.

Organization of the Task Force

The members of the Task Force represented a diverse group with a mix of a diverse and expertise in the areas of documentation of plant genetic resources for food and agriculture, including genebanks, ontology projects, data management organizations in the areas of biodiversity, universities, data standards, international CGIAR Centers, etc. The list of invited participants is attached as in *Appendix 1*.

The Secretariat circulated a note to the participants in advance with background information on the work of the Expert Consultation, the paper prepared and other useful information, including the usefulness of the adoption of PUIDs for the sharing of a non-confidential information associated with PGRFA in the context of the Multilateral System of Access and Benefit-

¹ <http://www.planttreaty.org/sites/default/files/cogis1w3.pdf>

² <http://www.planttreaty.org/sites/default/files/COGIS1re.pdf>

sharing. The work of the Task Force was organized as 3 one-week online discussions focusing on:

- Week 1: Review of the conditions and requirements for PUIDs for PGRFA material;
- Week 2: Analysis of the three PUID candidates identified by the Expert Consultation: ARKs, DOIs and LSDs;
- Week 3: Discussion of the minimum metadata structure and formats to be associated to each PUID.

Each week's discussion started with a message from the Secretariat describing that week's focus and offering some initial elements for the discussion. After each week, a summary report of the discussion was also circulated containing the major outcomes.

Summary of Discussion

Week 1- Conditions and requirements

During Week 1, the discussion focused on the definition of the requirements and concepts to be used in Week 2 to score the three candidate PUIDs. Based on the research undertaken by the Secretariat, a preliminary list of 21 items was circulated to open the discussion. The participants pointed out the importance of clearly understanding and agreeing on what would be identified through a PUID. It was agreed, following the proposal of the Secretariat that the focus should be, at least during the initial phase, only PGRFA material as defined in the Standard Material Transfer Agreement (SMTA), including both the material in genebanks or in other type of collections.

Another aspect that drew the participant's attention was the presence of other identifiers that are already assigned to the PGRFA that, once registered in GLIS, will receive a PUID. There was wide consensus on the idea of not replacing or redefining at all Accession Numbers or Collecting Numbers, but rather to capture all existing identifiers in the metadata structure so that they could be used as keys to locate the PGRFA of interest and thus allowing the Global System to provide a valuable mapping function on other existing identifiers assigned by various players to the same PGRFA.

Cost concerns were also expressed, especially in relation to DOIs, during the first week. Based on the research and interviews already undertaken, the Secretariat explained that the cost, traditionally associated to DOIs, had been largely minimized also thanks to the introduction of third party agents that offer registration services at very low prices. The Secretariat was asked to further minimize or eliminate the cost associated with the initial assignation of permanent unique identifiers to facilitate their adoption. The Task Force indicated that the necessity of taking into account all the costs involved in running the GLIS service for the extended period of time required, that go well beyond the simple registration fee but include the technical infrastructure of the system as well as appropriate human resources for its design, development and maintenance.

The detailed list of requirements, with their definitions, resulting from the first round of discussions is contained in *Appendix 2*.

Week 2- Analysis of the three PUID candidates

During Week 2 participants used the list of requirements obtained from Week 1 to analyse, score and compare the three PUID candidates. Given the technical nature of the questions, some participants provided more general comments on the preferred PUID and others replied with a partially compiled scoring for their best known PUID candidates. The Secretariat also provided a table with scores for all three candidate PUIDs taken from an extensive study of available literature, technical specification documents and discussion groups.

Despite some slight differences in some of the indicators, DOIs emerged as the recommended identifier type and standard for the development of the GLIS. See *Appendix 3* for a summary table aggregating the final scores assigned by the Task Force to the three PUID types compared.

The Secretariat initiated additional research to identify the most cost efficient arrangements to offer DOI registration free of charge to PGRFA users within the Global Information System.

Week 3- Metadata Structure and Formats

Week 3 focused on the metadata description to be associated to the DOIs. Some of the intended uses and functions of the metadata description were described in the opening email of the week's discussion as a) provide accurate identification of the PGRFA material; b) provide enough information to both users and applications; c) provide searchable descriptors for discovering the DOI associated to the PGRFA; d) model interactions between PGRFA using a set of relational operators, e) allow for the registration of multiple targets (i.e. web pages where additional details can be found on the PGRFA), f) accommodate existing identifiers allowing them to be used as search keys and g) can be mapped to some widely adopted format such as Resource Description Framework (RDF).

Adoption of RDF as the preferred format was recommended. An initial list of metadata fields was circulated using the list of Multi-Crop Passport Descriptors V.2 (MCPD). A lively discussion ensued where the list was further refined resulting in the one provided as Annex 4. However, it was felt that collaborating with some ongoing ontology initiatives proposed by the participants would be beneficial. Also, the list resulting from Week 3 should be considered as the starting point for further work in this area.

It was also pointed out that it may be unlikely that any existing ontology would be able to provide all fields required by the Global System and that the evaluation of several other ontologies may be in order to identify the set of fields to be adopted and in order to avoid a new ontology to be created.

The Task Force paid particular attention to some concepts and ideas, among them:

- the necessity of identifying a set of metadata fields that must be compiled when a PGRFA is registered;
- the availability of the information required to populate the mandatory metadata fields;
- a set of optional metadata fields that should be compiled if known and applicable;
- ownership of data is with data curators who are responsible for maintaining the metadata set associated to PGRFA they hold.

Moreover, it was considered important that the Global System would require minimal effort from data curators to facilitate its adoption, particularly in relation to acquisition costs for DOIs and development of infrastructure. It was also recognized that some sort of central services should be put in place.

Recommendations and Follow-up Actions

After extensive discussion of about 200 messages, the Task Force provided the following recommendations and suggestions for possible follow-up actions in the near future:

- 1) The digital Object Identifiers (DOIs) were considered the most appropriate identifier type to be proposed as a standard for the development of the Global Information System;
- 2) The Secretariat should make arrangements to facilitate the registration of DOIs free of charge for the PGRFA community;
- 3) The list of metadata fields, expressed in MCPD format, provided as *Appendix 4* is to be further refined through collaboration with leading ontology initiatives.
- 4) DOIs should be tested in a pilot phase involving PGRFA holders for the elaboration of best practices and methodologies for their deployment.

List of Participants

Name	Email	Institution
Mr Stefan Weise	weise@ipk-gatersleben.de	IPK Gatersleben
Mr. Dag Terje Endresen	dag.endresen@nhm.uio.no	Geo-Ecology Research Group GBIF Norway, Natural History Museum University of Oslo
Ms Ramona Walls	rlwalls2008@gmail.com	iPlant Collaborative
Ms Ruth Bastow	ruth@globalplantcouncil.org	Global Plant Council
Mr David Marshall	david.marshall@hutton.ac.uk	James Hutton Institute
Ms Sarah Hearne	s.hearne@cgiar.org	CIMMYT, SeeD project
Mr. Eugene Timmermans	eugene.timmermans@agr.gc.ca	Plant Gene Resources of Canada
Mr Axel Diederichsen	axel.diederichsen@agr.gc.ca	Biodiversity and Collections, Saskatoon, CANADA Agriculture and Agri-Food Canada
Mr. Barnabas Kapange	bkapange@spgrc.org.zm bkapange@gmail.com	SADC Genebank
Mr Éamonn Ó Tuama	eotuama@gbif.org	GBIF Secretariat
Ms Adriana Alercia	a.alercia@cgiar.org	Bioversity International
Ms Elisabeth Arnaud	a.alercia@cgiar.org	Bioversity International
Mr Stefano Diulgheroff	Stefano.Diulgheroff@fao.org	Plant Production and Protection Division of FAO
Mr. Matija Obreza	matija.obreza@croptrust.org	CROPTRUST
Marco Marsella	m.marsella@itworks.it	IT Consultant
Gerardo Francione	Gerardo.Francione@fao.org	ITPGRFA Secretariat
Francisco Lopez	Francisco.Lopez@fao.org	ITPGRFA Secretariat

**List of requirements of Permanent Unique Identifiers (PUIDs)
in the context of the Global Information System (GLIS)**

#	Requirement	Description
1	Uniqueness	Each PUID shall be associated to one and only one PGRFA entity
2	Permanence	The association between the PUID and the entity shall be maintained without any time limit
3	Opacity	No information on the entity should be inferable from the PUID alone
4	Resolvability	Suitable web interfaces are needed where a given PUID can be used by humans and client applications to obtain a description of the metadata associated with the PGRFA entity. Humans will receive a HTML page while a machine-readable format will be returned to client applications (see also #10 below)
5	Discoverability	Given one or more attributes of the entity, a web interface exists where those attributes can be entered to obtain the list of entities matching those attributes along with the associated PUIDs. Again, this applies to both humans and client applications (see also #10 below)
6	Security	Access to PUID management functions (e.g. editing of associated metadata) shall be allowed only to authorized users
7	Scalability	The adopted PUID type shall be designed to handle very large number of identifiers (hundreds of millions)
8	Interoperability	Interoperability with other PUID types shall be ensured
9	Compatibility	Local identifiers already assigned to PGRFA entities (e.g. Accession Numbers, Collecting Number) shall be preserved and provided as an attribute of the new PUID in the resolution service
10	Content negotiation	The default response format is HTML, but client applications shall be able to specify a preferred format (e.g. XML, RDF, JSON)
11	Accepted standard	Selecting a PUID type that follows an internationally accepted standard is considered a plus because it will guarantee a coordinated development of the framework
12	Acquisition and maintenance costs	The cost of acquiring licenses or registration with a central authority, as well as any other cost associated in acquiring the PUID technology (e.g. software tools) should be assessed
13	Acceptance by publishers	It is expected that GLIS entities will be cited in journals, books and papers. Selecting a PUID type that is widely accepted by publishers is considered a plus
14	Popularity	Adopting a PUID type that is widely used facilitates acceptance by users
15	Availability of tools	Software tools available in the most common programming languages (e.g. PHP or Java) will facilitate integration in GLIS
16	Resolution service and multiple resolution	Availability of a reliable, global resolution service is considered a plus because it will allow users outside the GLIS community to resolve GLIS PUIDs from third party websites. As an optional and powerful feature, a list of links to multiple destinations (with specific, multidisciplinary information on the same entity) can be returned in the resolution response. In this case, filtering of destinations according to some metadata description can be supported
17	Framework design	The detail and completeness of the logical and technical design framework should be evaluated because functions and services that are designed and specified for the PUID types in such frameworks are likely to be more reliable, powerful and interoperable with existing and future management systems compared to areas that are not specified
18	Metadata	The ability to incorporate different entity types and metadata descriptions is critical
19	Relations	The PUID shall support modeling of complex relations among entities such as instantiation, hierarchy, derivation, inclusion and so on
20	Identification of fragments	The possibility of identifying individual attributes or fragments of an entity

Comparative analysis of ARK, DOI and LSID with assigned scores

#	Requirement	ARK	DOI	LSID
1	Uniqueness	2	2.7	3.0
2	Permanence	2.5	2.7	1.7
3	Opacity	2.5	3.0	2.0
4	Resolvability	3	3.0	1.0
5	Discoverability	2.5	2.7	2.0
6	Security	2.5	2.0	2.0
7	Scalability	2.5	3.0	1.7
8	Interoperability	2	2.7	2.3
9	Compatibility	3	2.7	2.0
10	Content negotiation	2	2.7	2.3
11	Accepted standard	1.5	3.0	1.0
12	Acquisition and maintenance costs	2	2.0	2.7
13	Acceptance by publishers	2	3.0	1.3
14	Popularity	2	3.0	1.0
15	Availability of tools	3	2.7	1.0
16	Resolution service and multiple resolution	2	2.7	1.0
17	Framework design	2.5	3.0	2.5
18	Metadata	2.5	2.7	1.7
19	Relations	2.5	2.7	2.0
20	Identification of fragments	3	1.7	1.0
	Total	47.5	53.3	35.2

**List of requirements of Permanent Unique Identifiers (PUIDs)
in the context of the Global Information System (GLIS)**

M for Mandatory; O for Optional; ? for further research

Descriptor	Findings
INSTCODE	M
ACCENUMB	M
COLLNUMBER	?
GENUS	M
SPECIES	M
SPAUTHOR	M
SUBTAXA	M
SUBAUTHOR	M
CROPNAME	O
ACQDATE	?
ORIGCTY	?
COLLDATE	?
BREDCODE	?
SAMPSTAT	?
ANCEST	?
DONORCODE	M
DONORNUMBER	M
OTHERNUMB	M
MLSSTAT	M

Other fields to be considered, like Recordstatus (Active/historical)