November 2014

The International Treaty
ON PLANT GENETIC RESOURCES FOR FOOD AND AGRICULTURE

INTERNATIONAL TREATY ON PLANT GENETIC RESOURCES FOR FOOD AND AGRICULTURE

FIRST MEETING OF THE EXPERT CONSULTATION ON THE GLOBAL INFORMATION SYSTEM ON PLANT GENETIC RESOURCES FOR FOOD AND AGRICULTURE

San Diego, USA, 7-8 January 2015

TECHNICAL OPTIONS TO FACILITATE THE ESTABLISHMENT OF DATA LINKS IN THE FIELD OF PLANT GENETIC RESOURCES FOR FOOD AND AGRICULTURE: PERMANENT UNIQUE IDENTIFIERS

EXECUTIVE SUMMARY

1. *This document has been prepared at the request of the Secretariat of the International Treaty on Plant Genetic Resources for Food and Agriculture to facilitate the discussions in the context of the development of the Global Information System on Plant Genetic Resources for Food and Agriculture referred to in Article 17 of the International Treaty. The document presents some of the technical options that could be considered in setting up a global common method to apply permanent unique identifiers to plant genetic resources for food and agriculture, particularly those available in the Multilateral System of the International Treaty.*
2. *Furthermore, it presents a set of related standards and good practices that, if applied in connection with the adoption of permanent unique identifiers, would help to establish data links in a more automated way and hence increase the value of current genetics stocks for plant breeders. The study also analyses some of the limitations and obstacles faced by data curators and offers some preliminary considerations, to be further explored, on how to overcome them.*
3. *Finally, it enumerates a series of technical measures and recommendations to improve data interoperability for the benefit of plant breeders and data curators and that could become standards and common good practices if consensus is reached in the short and medium term in the context of the Global Information System.*

Table of Contents

Paras.

I. INTRODUCTION

1. The development of the Global Information System offers an opportunity to the Treaty Community to provide a wider set of non-monetary benefits in the form of services and to enhance the delivery of services to genebanks and data curators.

2. Some of the immediate benefits that the development of the Global Information System can bring to plant breeders, farmers, genebanks and data curators are related to the enhancements of ongoing operations within the Multilateral System. This is particularly relevant to the availability and accumulation of public information on the Plant Genetic Resources being transferred.

3. To achieve these and other advantages within the Global Information System, the Treaty Community needs to be able to connect data sets maintained by a wide variety of organizations. This document provides background information on some of the technical options most frequently adopted to facilitate the establishment of data linkages at Accession level, particularly involving some kind of unique permanent identifiers, and puts forward recommendations for the consideration of the experts of the Global Consultation on the Global Information System.

II. BACKGROUND

4. Through Article 12.1 of the Treaty, Contracting Parties agreed to facilitate access to plant genetic resources for food and agriculture under the Multilateral System and in accordance with the provisions of the Treaty.

5. Among the conditions of the transfer, Article 12.3.c. of the Treaty states that "*All available passport data and, subject to applicable law, any other associated available non-confidential descriptive information, shall be made available with the plant genetic resources for food and agriculture provided*".

6. Article 12.4 of the Treaty provides that facilitated access under the Multilateral System shall be provided pursuant to a Standard Material Transfer Agreement (SMTA), which was adopted the Governing Body of the Treaty, in its Resolution 1/2006 of 16 June 2006.

7. Article 3 of the SMTA states:

> *"The Plant Genetic Resources for Food and Agriculture specified in Annex 1 to this Agreement (hereinafter referred to as the "Material") and the available related information referred to in Article 5b and in Annex 1 are hereby transferred from the Provider to the Recipient subject to the terms and conditions set out in this Agreement."*

8. The Multilateral System of Access and Benefit-sharing started to operate in January 2007 and in order to facilitate the use of the SMTA, the Secretariat developed an offline application, distributed on CD-ROM, to assist the provider in the compilation of the agreement using ordinary personal computers. The application was presented at the Second Session of the Governing Body in 2009.

9. Since then, the Secretariat has continued assisting users in the operations of the SMTA tools. This effort culminated in 2011 with the deployment of Easy-SMTA, a system to allow the online reporting of SMTAs to the Governing Body.

10. Easy-SMTA was further improved by 2013: on one hand to support the online generation of the SMTAs, a feature particularly requested by small and medium size providers, and on the other hand to provide the system-to-system integration requested by large providers.

11. In the development of these tools and systems, the Secretariat was not alone and number of experts and stakeholders had been involved in the different stages of development to make the system match the needs of the users.

12. With the view to facilitate the operations of the Multilateral System at the beginning of the legal and technical consultation process set up but the Secretary, a number of needs were identified, including:

    i. Discovery of material included in the Multilateral System

    ii. Compiling agreements

    iii. Reporting

13. It was agreed that the discovery of material had implications going beyond the SMTA and would need to be considered in the larger picture of the other provisions of the International Treaty, particularly in relation with the development Article 17, on the Global Information System.

14. It was also noted that a system was not in place to easily refer to the associated available non-confidential descriptive information that shall be made available with the plant genetic resources for food and agriculture transferred through a SMTA.

15. Easy-SMTA allows the user to attach the associated information on the PGRFA being transferred to the SMTA agreement, resulting, in some cases, in very large documents. Alternatively, the Provider can indicate an URL where on the Internet such information can be found. While this second option is quick and easy to use, it suffers from the inherent unreliability of URLs whereby links are broken when the domain or the location of the Provider's changes over time. Because of this, URLs are not *permanent*.

16. At the same time, Easy-SMTA treats this information as confidential as it is associated to the SMTA. This means that it can be used only by the intended recipient and even for the recipient it may be difficult to properly record all that information in an structured way.

17. The information associated to PGRFA has an important role in helping users of the Multilateral System to find what they are looking for. Maybe even more importantly, it can be improved or enriched during transfers because each subsequent recipient may increase the value of the PGRFA with further analyses and evaluations.

18. Unfortunately no mechanism is currently available to facilitate the easy aggregation of all the non-confidential information related to PGRFA. This would be of benefit not only for the subsequent recipients, but also for the all the providers.

19. To date, many initiatives, in several domains, have attempted to set up a global information system following a centralized approach. These methods, while producing some valuable results, have failed over time due to the enormous transaction costs and, among other limiting factors, because data curation was handled away from the originator.

20. In the field of life sciences, some innovative approaches have been developed and already tested with preliminary encouraging results helping researchers find raw data and share

results of their work. This was possible thanks to the development and adoption of standards, technologies and tools facilitating documentation and data sharing in a more automated way.

21. This document analyses some of the standards and technical options for documenting and sharing PGRFA-related information that, if adopted by the Treaty Community and the PGRFA community in general, could boost the way in which plant breeding and plant conservation is done.

III. LINKING DATA THROUGH PERMANENT UNIQUE IDENTIFIERS (PUI)

22. The Vision for the Global Information System does not present a single, monolithic system storing all such diverse and dynamical information. Rather, it is more likely to be a constellation of systems, many already existing, where each one contributes the corresponding institution's perspective and expertise and a set of related services and tools.

The Global Information System is likely to be a constellation of systems

23. A critical need therefore arises to make easily available the information about the same PGRFA[1] that is distributed across all the systems. The aim is to facilitate the integration of existing phenotypic, genotypic and geographic data with passport information and other sets of relevant data to create added value.

We need to link information stored in the different systems and make it easily accessible

24. It is possible to generate added value within the existing information technology infrastructure developed to assist the users of the Multilateral System as an initial step. The workflow is illustrated in Figure 1. When the Provider transfers PGRFA to the Recipient using Easy-SMTA, together with the Accession number, he could also report a Permanent Unique Identifier (PUI) for each sample being transferred, rather than the actual URL address. The Recipient can subsequently use the PUI on the Global Information System and access all information related to the material received. Along the way, more information may be obtained and contributed by the users of the Multilateral System through the Global Information System.

Permanent Unique Identifiers (PUIs) are a necessary step towards this integration goal

25. This function would preserve the confidentiality of the information associated with the transfer while making available only the associated information with the PGRFA being transferred.

[1] The term PGRFA in this document is to be read as defined in the SMTA: "any genetic material of plant origin of actual or potential value for food and agriculture".

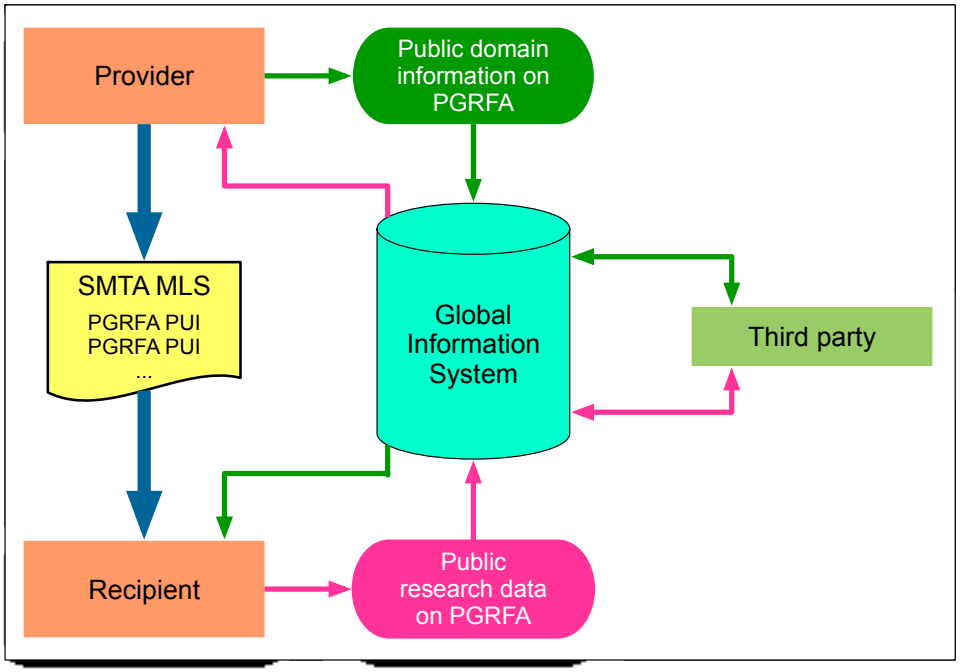

Figure 1. Workflow of PGRFA and associated information.

What is required to make all this work is:

i. a Permanent Unique Identifier (PUI), i.e. a text string that, as a minimum requirement, unambiguously and permanently identifies a single object of interest. This concept can be applied to each sample being transferred, of course, but also to any individual object or piece of information that we need to be able to unambiguously and permanently identify;

ii. a set of rules that the PGRFA Community, i.e. the constellation of systems in the GIS and any other system involved, agrees to adopt in the assignment of a PUI to an individual object of interest and their management;

iii. a set of technical solutions, e.g. communication protocols, to interconnect the systems and make them work together to the user's benefit.

26. Adopting a common PUI strategy is an important step towards structured data, i.e. information that is much easier to deal with using computers and automated processes reducing human intervention. This, at the same time, increases that availability of the information, improves data quality reducing error and would significantly improve the way users of the Multilateral System find and access PGRFA. In fact, its potential benefits go beyond the community and expand to the plant breeding and conservation systems.

Making information directly accessible by computers lets us tap much larger datasets and establish relations more effectively

IV. MAIN FEATURES AND TYPES OF IDENTIFIERS

27. According to the above, the first step to create a decentralized global information system consists in providing guidelines on how to properly label the information objects and how to link them. These questions are central to the architecture of the global system, which, on one hand, cannot be just a repository of individual data entries and, on the other hand, needs to preserve the investments that have been already made by the participating institutions.

28. At the first preparatory meeting for the Global Consultation organized by the Treaty Secretariat in Bonn, Germany, in April 2014, one of the suggestions of the participants was for the Treaty Secretariat to articulate a discussion on the kind of unique identifier that are needed to build the global system architecture. In fact, the discussion on the adoption of unique identifiers in the PGRFA domain is not new and it has been addressed several times in diverse technical meetings over the last decade with modest results to date. Some of the reasons of this stagnation being the lack of a wide consensus, the need to bring the issue to the attention of a global normative body and the necessity of establishing support mechanisms and incentive for most data curators to embrace the initiative.

Discussions on PUIs have already been ongoing but the lack of a coordinated effort had led to stagnation

29. Based on analysing the available literature and consulting with technical experts involved in the biodiversity informatics area and in order to stimulate further consultation in the context of the Global Information System, the Treaty Secretariat has categorized the main features of the PUIs in our context.

30. Any identifier should:

    i. be *unique*, i.e. unambiguously identify a specific object or intellectual asset;
    ii. be *permanent, i.e.* always valid: the same object will be forever associated to the same identifier;
    iii. be *opaque*, i.e. nothing about the associated object should be inferrable by the structure of the identifier;
    iv. be *actionable*, i.e. a defined procedure for name resolution exists to access the information associated to the object once the identifier is known. In other words, the resolution is the process in which an identifier is input to a network service that returns as output one or more information pieces related to the object[2]. Resolution should be available in both human and computer-readable forms;
    v. be *discoverable*, i.e. given details of an object, it is possible to retrieve its identifier.

Requirements of a PUI: unambiguous, permanent, opaque, actionable, discoverable

31. Simultaneously covering all these needs is possible through the use of Permanent Unique Identifiers (PUIs). Indeed, PUIs become a very effective way to associate information *on the*

[2] http://www.doi.org/factsheets/DOIIdentifiers.html

same object stored in two or more different databases. It is enough for all databases to associate the same PUI to their own data for it to become unambiguously and permanently linked, which is exactly what is required. All the databases are now effectively connected as a larger distributed system.

Opacity: The opacity requirement may seem surprising at first but it should be remembered that PUIs are intended to be used by computers to communicate or interact with other computers. In fact, opacity is a desirable quality in a PUI because it prevents humans from making assumptions on the associated object or on its relations to other objects that may be totally incorrect.

32. Given the hierarchical nature of some PUIs, it will also be possible to attach hierarchically related PUIs to specific attributes of the object and share them separately or in the context of the parent object. Likewise, some PUIs can be linked so that relations among objects can be modelled.

For PUIs to be effective, it is necessary to agree on a common, controlled vocabulary to describe the objects of interest

33. But how do we agree on the object to which a PUI is attached? How do we know that two objects are different enough to deserve two separate PUIs? There is no single, easy and quick solution, but a high level of success can be achieved by adopting a combination of good practices and standards such as controlled vocabularies that come to the rescue when associated with the PUI. An example of such vocabularies, that has been tailored to the needs of describing biological specimens and based on the standard Dublin Core Metadata Initiative[3], is Darwin Core[4].

Darwin Core is a good candidate as controlled vocabulary, although more work is required

34. A vocabulary like Darwin Core helps us declare that two objects are different and, therefore, that they need two different PUIs, when an agreed set of descriptors contains different values. For instance, two samples differing only in some secondary Darwin Core descriptors are likely the same whereas two samples where many more fields differ most likely require different PUIs.

35. It is to be noted that, in order to avoid any ambiguity on the Darwin Core terms and to keep track of all the changes that a term has gone through and of all intervening decisions, each Darwin Core term is associated to a PUI[5].

[3] http://dublincore.org
[4] http://rs.tdwg.org/dwc/
[5] *In the form of an Uniform Resource Identifier or URI.*

36. It is expected that PUIDs will be initially assigned to individual genebank Accessions. However, the PUID strategy is more general and they can be assigned to any information object and entity related to PGRFA and related activities.

37. Although the advantages of adopting PUIs are clear, it is to be noted that there are several types of them and that it would be advisable that the global community discusses and, if possible, agrees on one of them in order to maximize interoperability. The issue is which one of the several PUI types to adopt. The following section compares some of the most widely used PUI types[6].

IV.1. Uniform Resource Identifier (URI) and Persistent Uniform Resource Locator (PURL)

38. These two are based on the familiar URL schema: 'http://name/path' where 'name' is the portion that identifies the authority (i.e. the institution assigning the PUI for URIs or purl.org) and 'path' is the hierarchical sequence of context(s) and object identifier. The main advantage is that resolution can be easily carried out through a web browser. The major disadvantage of URIs is that systems move or evolve over time and, because of this, these identifiers may be broken (the familiar "404 not found" message that we see in our browsers when we access a broken link). The PURL is an attempt to mitigate this *dead link* problem by using an intermediate resolution service but they offer no support for integrating or discovering other identifier types. Another problem common to both URIs and PURLs is their lack of opacity.

IV.2. Digital Object Identifier (DOI)

39. DOIs identifies specific abstract or real objects, whether or not they are available on the Internet. Management of DOIs is controlled by the International DOI Foundation that provides a backbone resolution service. This identifier is highly opaque and usually more robust than other PUIs. DOIs support relations between objects which is of obvious interest for the modelling of objects in the plant genetic resources domain. On the downside, DOIs present the disadvantage for the data curator that a cost may be associated to obtaining or managing DOI names while the DOI name resolution is free of charge.

IV.3. Life Science Identifier (LSID)

40. This type of identifier was developed by IBM with the needs of the life-science community in mind. It cannot be resolved by a normal browser but requires a rather articulated resolution procedure. It is not as opaque as DOIs. Despite its more focused origin, LSIDs have failed so far to gain significant popularity.

IV.4. Universally Unique Identifier (UUID)

41. The UUID[7] is registered as a URN scheme and provides a robust identifier technology independent from any centralized authority. UUIDs can be generated on demand using one of the specified UUID generation algorithms. UUIDs will not need to be minted by a central authority and will scale well to networks with a very large number of objects such as the biological material sample collections. As for other URN schema a resolver service can be established to provide full resolution of UUIDs registered to this resolver.

[6] A Beginner's Guide to Persistent Identifiers Version 1.0, February 2011, GBIF.
[7] http://tools.ietf.org/html/rfc4122

IV.5. Archival Resource Keys (ARK)

42. The California Digital Library developed the ARK[8] identifier scheme largely based on experiences with the Handle system, DOIs and PURLs. The ARK scheme proposed a technological solution to improve the long-term persistence of the service through a minimal complexity of the resolver infrastructure. The ARK technology separates the original name assigning authority from the diverse and multiple name mapping authorities (resolution services). EZID[9] is a related identifier technology provided by the California Digital Library offering an easy way to create and mange long-term identifiers, including ARKs and DataCite DOIs.

V. THE DIGITAL OBJECT IDENTIFIERS (DOIs) AS A VIABLE OPTION

V.1. The Distinctive Features of DOIs

43. The key question regarding the consideration of the Digital Object Identifiers as a viable, and possibly, the best current option is which aspects would make them attractive to the Plant Genetic Resources Community. While extensive documentation with all the details is available in the DOI Handbook[10], this section presents just a summary, which is also illustrated in Figure 2:

    i. The DOI system is based on a set of **ISO and ANSI standards** ensuring a solid foundation and a controlled environment;
    ii. An object associated with a DOI name is described **unambiguously** by DOI metadata, based on a structured, extensible data model that enables the object to be associated with metadata of any desired degree of precision and granularity to support community-specific description and services;
    iii. DOI names support **multiple resolutions**, i.e. resolving a DOI name can return different levels of information: from the simple HTTP redirect to a single description page to a list of destinations or complex metadata to be used by the client system according to its needs. Multiple resolution is very useful to link information on the object spread across multiple systems as it happens with the information associated with the PGRFA transferred with the SMTA;
    iv. The DOI system supports **Linked Data**, a set of best practices aimed at facilitating access to multiple data sources, which is very good to promote participation of existing systems in the Global Information System on PGRFA with minimum effort;
    v. The DOI system supports **Content Negotiation**, a way for the client to specify the preferred format of the response while supporting widely accepted default standards such as RTF or JSON;
    vi. DOI names can model **relationships between objects**. A minimum set of relation operators is included in the DOI system (such as "is derived from" that could be used to record the transformations of PGRFA during their transfer from Providers to Recipients), but custom extensions are supported according to the needs of the community of users;

[8] https://tools.ietf.org/html/draft-kunze-ark-18
[9] http://ezid.cdlib.org/
[10] http://www.doi.org/hb.html

vii. DOI names support the **identification of fragments of objects**, i.e. cases where individual attributes or groups of attributes of the object need to be identified separately. This can be obtained without registering new DOI names but rather deriving the desired fragment identifiers through a generating formula from the DOI name associated to the object as a whole

viii. DOI names are **designed to accommodate other PUI types** supporting full interoperability in cases where an investment has been already made in a different PUI type. This feature can be extremely useful to link information from different existing systems in the domain of PGRFA, i.e. to link records from databases where different identifiers have been adopted over the years[11]. This allows for existing legacy schemes, if available, to be used. The Global Information System will need to connect a wide variety of resources and adopt a PUI able to preserve and re-use other identifiers[12].

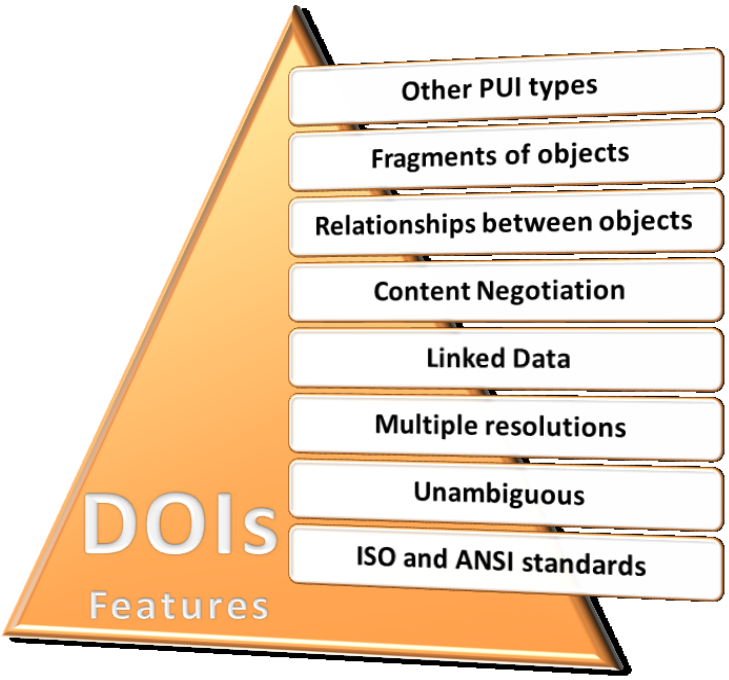

Figure 2. Main Features of DOIs

V.2. What does a DOI look like?

44. Now that we have introduced DOIs in general, let's have a look at few examples. The following are real DOI names:

```
10.1371/journal.pgen.1003477
10.1111/1467-9388.00298
10.1038/nrg1729
```

A DOI name takes the form of a character string divided into two parts, a prefix and a suffix, separated by a slash. The prefix identifies the registrant of the name, and the suffix is chosen by the registrant and identifies the specific object associated with that DOI. For example, in the DOI name 10.1111/nrg1729, the prefix is 10.1111and the suffix is nrg1729. The "10." part of

[11] *The existence of multiple (third, fourth, etc.) identifiers should be recognized in the DOI metadata field "referentIdentifier(s)" (other identifier(s) commonly referencing the same referent by multiple values), rather than by incorporation in the DOI name.*

[12] *Identifier Interoperability:* http://www.doi.org/factsheets/Identifier_Interoper.html

the prefix identifies the DOI registry in the Handle System, and the characters "1038" in the prefix identify the registrant. "nrg1729" is the suffix, or item ID, identifying a single object. DOIs should display on screens and in print in the format "doi:10.1038/nrg1729". The DOI can be resolved online and the information associated with it obtained by both a human and a machine. While the human will be redirected to a web page with the information, a computer may receive the associated information in several formats based on the content negotiation feature of DOIs.

45. To resolve a DOI, we can just take any browser and type the following URL:

```
http://dx.doi.org/{DOI name}
```

or, using the above example:

```
http://dx.doi.org/10.1038/nrg1729
```

this result is displayed as Figure 3.

Through this simple action, we have resolved the DOI name accessing the object it refers to, i.e. the Nature article). This is the minimum level of resolution services, one easily accessible to humans. Computer application would be able to access different representations of the object associated to the DOI and, through Content Negotiation, even specify in what format they want to receive such information. For example, an application could query the DOI register for objects with specific attributes, say articles on biodiversity, then query each returned source to compile a digest and present it to the user.

Figure 3. Example of DOI resolution

VI. RESPONDING TO THE NEEDS OF THE PGRFA COMMUNITY

VI.1. Example from other domains

46. There are many instances where the use of the Permanent Unique Identifiers has helped document and share research data. For instance, the University of Bristol manages a dedicated research data repository as part of their Research Data Service[13]. They are currently using the DataCite service to assign digital object identifiers (DOIs) to research datasets in order to provide unique and perpetual identifiers for data, hence facilitating citation and discoverability. The repository at the British Library hosts data underpinning research outputs and provides a home for data with immediate and straightforward access. The Bristol

[13] http://data.bris.ac.uk/

Research Data Service provides guidance on how to use the identifiers to cite data and is developing appropriate policies to monitor usage[14].

DOIs are already widely used in many and diverse research fields

47. The use of permanent identifiers has also been integrated and improved the architecture of Wikipedia, the famous knowledge repository, through the support of CrossRef[15][16]. Wikipedia uses DOIs to avoid a number of common issues with citations like broken links, copy and paste errors made by humans in citations text, copyright violations whether accidental or deliberate. It also provides verifiability enhancements, as the DOIs always leads to the correct sources, so modifications of abstracts or even content is avoided. When more than one source exist, DOIs links, like ISBN book numbers, let users choose the source of preference.

VI.2. Benefits for the users of the Multilateral System

48. The advantages of adopting PUIs in the context of the Treaty are large and numerous, for instance:

    i. Establishing a connection between the information generated by the Provider of the PGRFA and the corresponding associated information likely to be generated by a Recipient. This may become a very powerful incentive for actual and potential providers of PGRFA within the Multilateral System of Access and Benefit-sharing.
    ii. Establishing an effective way of accessing the information accrued about each individual PGRFA, regardless of where this information is maintained. This would greatly simplify the establishment of a global distributed information system linking associated information for each PGRFA occurrence during multiple transfers.
    iii. Facilitate the adoption of a standard vocabulary by the PGRFA Community, such as Darwin Core, to provide the necessary metadata structure to be associated to DOI names.

Adopting PUIs would greatly improve the effectiveness of the Multilateral System of Access and Benefit Sharing

49. These advantages are not specific to Digital Object Identifiers (DOIs), but they are a very effective way of attaining these advantages.

VI.3. The contribution of identifiers to genomics

50. The facilitation of processes related to the collection of relevant metadata in the field of genomics would clearly benefit the community by reducing ongoing replication of efforts and by maximizing the capacity to share and connect data within the genomics community and the plant genomics community in particular. In this sense, some initiatives have already

[14] *Duke, M. and Gray, S. (2014). 'Assigning Digital Object Identifiers to Research Data at the University of Bristol'. DCC RDM Services case studies. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/persistent-identifiers*

[15] *At the beginning of 2000, the world's leading scholarly publishers joined to form the non-profit, independent organization, Publishers International Linking Association, Inc. (PILA), which operates CrossRef, http://www.crossref.org/02publishers/doi_display_guidelines.html*

[16] http://crosstech.crossref.org/2014/08/citation-needed.html

discussed the need for standard processes related to the way we document and share information.

DOIs are already used in leading genomics initiatives

51. One of these initiatives is the Genomic Standard Consortium (GSC), an international body established in 2005 to promote mechanisms to standardize the description of genomics and the exchange and integration of genomic data. In 2008 GSC proposed to define a set of core descriptors for genomes and metagenomes in the form of "Minimum Information about Genome Sequence (MIGS) to be applied to genome publications"[17]. Among the list of recommended core fields we find "source material identifier", which at that time was not a must as many repositories did not yet apply them. This list of minimal information was recognized by the GSC as just a starting point for the description of genomes and metagenomes and they already listed among them, the Digital Object Identifier (DOIs).

VI.4. Possible implementation pathways

52. Adopting DOIs requires a sequence of steps, according to the experience already accumulated and the procedure described in the existing literature, including the DOI Handbook. As a background, further refinements on Darwin Core would be advisable to make it a solid foundation for the community DOIs. This has been recognized by the Global Conference on Bioinformatics held in 2012 in Copenhagen and by the Global Biodiversity Informatics Outlook within the Focus Area A3 "Persistent storage and archives"[18]. The participants also indicated that limited benefits were to be created unless the associated data to the persistent universal identifiers for biodiversity data is made available on the web.

53. In the context of the work already undertaken for the setting up of the information technology tools in support of the Multilateral System of Access and Benefit-sharing, the Secretariat of the International Treaty could explore how to facilitate the adoption of DOIs reducing transactions costs. Currently, acquiring a DOI involves the payment of a fee that is set by the Registration Agency (RA), i.e. the body managing DOIs specific to a community. In turn, RAs pay an annual fee to the International DOI Foundation (IDF) that charges them on a cost-recovery basis. The IDF, among other activities, runs the worldwide backbone resolution service, an extensive network of servers worldwide, that guarantees the resolvability of DOIs.

The major obstacle to widespread adoption of DOIs is the cost associated to obtaining one but this could be alleviated by establishing a Registration Authority

54. As a first step there is the need to set up a DOI system and identify a suitable Registration Agency (RA), i.e. the organization managing the subset of DOI names specific to a community. There are several RAs offering DOI registration services but they are mostly run

[17] *The minimum information about a genome sequence (MIGS) specification, at Biotechnol. May 2008; 26(5): 541–547. DOI: 10.1038/nbt1360*

[18] *Global Biodiversity Informatics Outlook, delivering Biodiversity Knowledge in the Information Age. Global Biodiversity Information Facility. 2012. http://www.biodiversityinformatics.org/*

as business operations and seek a financial return or serve very specific needs. The International Treaty and its Secretariat, as trusted third party already managing a system critical for the Treaty such as Easy-SMTA, could alleviate the aforementioned cost issue becoming a RA for the Treaty community.

55. Some of the advantages of such option are:

    i. the annual RA fee paid to the IDF could be borne by the Treaty budget so that SMTA parties will not have to pay anything for the use of DOI names, removing the major obstacle to DOI adoption;
    ii. community-specific, value-added services could be offered by the Secretariat in collaboration with major Providers and initiatives, whereby, by resolving a PGRFA DOI, any user will be able to access the wealth of information distributed across all such systems.

56. The Secretariat already provides technical coordination and management functions on major Treaty Systems, particularly the Multilateral System of Access and Benefit-sharing, and it would be a natural extension of its duties to offer RA services as well as a non-monetary benefit in the context of the Global Information System[19].

Managing DOIs for the Treaty community would be a natural addition to the Secretariat's duties

57. Once the RA status is obtained, the technical infrastructure needs to be put in place to provide the minimum resolution services required. At first, this could be outsourced to the IDF itself that offers such service on request. Once the adoption of DOIs widens and services required by the community grow, the Governing Body of the Treaty could consider other options, like the managing its own resolution service, either through a group of collaborating institutions hosting mirror servers or through the management a number of virtual instances on the cloud.

58. As a background activity, the DOI metadata descriptor will have to be customized according to the needs of the community. In particular, mapping of metadata to the DOI metadata kernel, the minimum set of metadata that needs to be shared across the entire DOI system to make DOIs discoverable, will have to be achieved. This leverages the support for community-specific metadada descriptions offered by DOIs.

59. In the meantime, the strengthening of partnerships with existing capacity building and training providers to organise awareness seminars and training courses would be required to educate the community on PUIs in general (and DOIs in particular) and on controlled vocabularies such as Darwin Core and ontologies so that an effective use is made of this new opportunity. The discussion on controlled vocabularies and ontologies is beyond the scope of this document, but several initiatives have been advancing on the subject and should be taken into account.

60. Also network of providers of technical support will have to be set up without the need for the Treaty Secretariat to deliver these services directly. This way, participating institutions could train their staff and upgrade their systems to accommodate the DOIs and associated services

[19] *A more detailed cost and feasibility study would be required for the consideration of the Governing Body. Alternative options, like the development of partnerships with an existing RA, could also be further explored.*

with adequate assistance, with special attention to those in developing countries where more capacity to adopt new technology and standards is needed.

Adopting PUIs would require enhancements to existing systems, but this can be mitigated by establishing a support network, especially aimed at developing countries

VII. CONCLUSIONS AND RECOMMENDATIONS

61. The adoption of Permanent Unique Identifiers (PUI) is a necessary step to achieve data integration in the field of plant genetic resources for food and agriculture and for the future development of a coherent Global Information System. Its adoption and use in connection with the transfer of PGRFA material would facilitate also the SMTA reporting. While there are many good-enough options identified for the adoption of PUI, the international community would benefit from an agreed common standard, particularly to facilitate data integration.

62. The use of Digital Object Identifiers (DOIs) offers some clear technical advantages compared to other PUI options. Among these advantages it is worth noting its interoperability, the possibility to enable users to re-use them with other existing identifiers. One of the major limitations for their initial adoption is related to the financial cost of acquiring packages of DOIs. Although these costs are not prohibitive, it could become an obstacle for some users in the context of the Global Information System on PGRFA.

63. If the DOIs are to be the preferred technical option, the issue of eliminating or further discounting the costs associated with their issuance could be further explored at community level through a cost study to be prepared by the Treaty Secretariat for the consideration of the Governing Body, including the option to establish a dedicated Registration Agency and other alternatives, in association with the support services being offered for the operations of the Multilateral system of Access and Benefit-sharing.

64. Training and capacity building, particularly for PGRFA data managers in developing countries, would also be needed no matter what kind of PUI is chosen. These services could be provided through the establishment of a network of training and capacity building providers and the design of a collaborative programme oriented to upgrade the skills of bioinformaticians and other staff working in genebanks and other institutions curating relevant PGRFA data repositories. The International Treaty, through the third round of its benefit-sharing Fund, has also started to consider proposals in support of the implementation of Permanent Unique Identifiers.