Food and Agriculture Organization of the United Nations

The International Treaty ON PLANT GENETIC RESOURCES FOR FOOD AND AGRICULTURE

INTERNATIONAL TREATY ON PLANT GENETIC RESOURCES FOR FOOD AND AGRICULTURE

Rome, Italy, February 2016

REPORT AND ANALYSIS OF THE GLOBAL SURVEY ON DESCRIPTORS REQUIRED FOR PGRFA

EXECUTIVE SUMMARY

This report has been prepared by the Secretariat of the International Treaty on Plant Genetic Resources for Food and Agriculture (PGRFA) as a critical input to facilitate the design and development of the Global Information System (GLIS) in the context of Article 17 of the International Treaty regarding descriptors required to identify PGRFA. It contains the analysis of the two consultations conducted during 2015 on the minimum essential information (descriptors) required and other highly recommended data to be declared and aggregated through GLIS to facilitate access to scientific information about PGRFA.

Both, the survey and the subsequent expert consultation collected useful views and information from Contracting Parties and stakeholders from all over the world. This document outlines the major findings on each of both consultations and also analyses some of the limitations and obstacles faced to devise the initial key set of mandatory descriptors.

The strategic key set of descriptors defined constitutes an essential step to uniquely identify PGRFA samples transferred under the Multilateral System of Access and Benefit Sharing and is critical to require a Digital Object Identifier (DOI) to register distinct types of plant material from different types of holders in the Global Information System.

Finally, it enumerates a series of technical issues and questions for further research and consideration during the early implementation of the first Programme of Work on the Global Information System.

TABLE OF CONTENTS

I. INTRODUCTION

1. The request to work on the further definition of global permanent unique identifiers came from the Consultation on the Global Information System on Plant Genetic Resources for Food and Agriculture meeting Report (IT/COGIS-1/15/Report) which took place in San Diego, USA, on 7-8 January 2015.

2. The need for Permanent Unique Identifiers (PUIDs) emerged as critical to unambiguously and permanently identify plant genetic resources for food and agriculture (genebank accessions or breeding material) being exchanged not only in the context of the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) and its Multilateral System, but also outside. Once a PUID is assigned, the material can be referenced easily and unambiguously forever, even across organizations.

3. With PUIDs, non-confidential information about PGRFA made available by different organizations and scattered across multiple databases may be gathered and harvested more efficiently. In other words, assigning a PUID is a critical first step towards more organized, manageable, effective and available information on PGRFA at the global level to support plant breeding, research and utilization of material, among others.

4. Following on such request, the Secretariat of the International Treaty organized a Task Force on Permanent Global and Unique Identifiers for PGRFA[1] in March 2015 which determined that the use of Digital Object Identifiers was the most promising technical option to increase data connectivity in the area of plant genetic resources.

5. In its final report, the Task Force advised on additional research to be undertaken by the Secretariat, in consultation with appropriate parties and stakeholders, for the elaboration of best practices and standards to be associated with DOIs as an essential element for the implementation of the Global Information System (GLIS) on Plant Genetic Resources for Food and Agriculture.

6. The present report gives insights into the consultation processes conducted during 2015 and the primary outcomes, in particular the *Global Survey on descriptors required to register material in the Global Information System on Plant Genetic Resources* and the subsequent *Focus Group Consultation.* In doing so, the document also illustrates the methodology adopted for the consultation and for the work of the Focus Group.

7. This paper also identifies a few elements for further consideration during the implementation of the first Programme of Work on the Global Information System (2016-2022), adopted by the Governing Body at its Sixth Session in October 2015[2].

---

[1] 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

[2] Resolution 3/2015, The Vision and the Programme of Work on the Global Information System.

II. METHODOLOGY

8. The structure and the concepts of the Consultation process took into consideration:

(i) The FAO/Bioversity List of Multi-Crop Passport Descriptors (MCPD)[3], which are the most advanced and globally adopted standards for passport data *of ex situ* genebank accessions; and

(ii) The conceptual work developed during the design of International Rice Information System (IRIS) and the International Crop Information System (ICIS)[4].

9. The Secretariat collected information from Contracting Parties and stakeholders from all over the world, as follows:

First phase: Global Survey on descriptors required for the assignation of a Digital Object Identifier (DOI) to PGRFA in the Global Information System; and

Second phase: Focus Group Consultation to validate survey results and ensure their wide applicability.

10. The Secretariat issued an online notification on 16 July 2015, which was sent by email to representatives of Contracting Parties- including the national focal points of the Treaty- and interested stakeholders, requesting them to complete and share the survey. The survey was structured in 13 questions divided in two sections listing 'mandatory' and 'highly recommended' descriptors. The survey was online until 5 September. Most of the questions required a positive or negative reply on whether the type of descriptor was relevant for the identification of the material. Respondents were also requested to provide any additional missing information with a substantiated justification for its inclusion. A total of 219 experts (including plant breeders, genebank curators, National Focal Points of the Treaty, researchers, information specialists, geneticists and bioinformaticians, among others) from 98 organizations distributed in 60 countries participated in the survey.

11. The validation phase was conducted through a Focus Group Consultation with selected experts that participated in the survey and that presented different profiles. The collection of inputs was carried out through the email system and videoconference until the end of October 2015. This activity engaged 15 key renowned scientists from different organizations involved in plant breeding, *in situ* and *ex situ* conservation, and PGRFA documentation and data exchange activities, who participated in the Global Survey.

[3] Alercia A, Diulgheroff S, Mackay, M. 2015. FAO (Food and Agriculture Organization of the United Nations), Bioversity International.

[4] International Crop Information System

III. *First phase:* Global Survey on descriptors required for the assignation of a Digital Object Identifier (DOI) to PGRFA in the Global Information System

(a) Main principles

12. In the process of defining the first priority list, survey participants were asked to apply the following criteria to select and prioritize passport descriptors:

- Initial strategic minimum mandatory set that uniquely *identify* PGRFA samples;
- It should include *all PGRFA* (i.e. *ex situ*, *in situ*, on-farm), including material under development, genebank accessions, farmers and research material among others;
- It should start with PGRFA *available* material shared under the Multilateral System (MLS) of Access and Benefit-sharing;
- It should not include PGRFA material that no longer exists, for example, accessions that have been lost; and
- Data on the *minimum set* should be available.

13. The following steps underpinned the development of the descriptors list contained in the survey:

(b) Information collection and reference documents

14. One of the aims of the research was to determine how the plant material was identified in genebanks and in research pools and their common elements. The research started with a comparative analysis of all descriptors listed in the List of Multi-Crop Passport Descriptors v2.1 and those relevant to the ICIS system.

15. As result of this exercise, and to assist in the selection of a "reduced" set of mandatory fields, a comparison table was elaborated using also sources such as *The ECPGR concept for* in situ *(on-farm) conservation in Europe*[5]; the *Core descriptors for* in situ *conservation of crop wild relatives v.1*[6]; the List of requirements of Permanent Unique Identifiers (PUIDs) in the context of the Global Information System findings[7]; fields used in the Focused Identification of Germplasm Strategy tool (FIGS)[8]; the Darwin Core Germplasm[9] standards; the Access to Biological Collections Data (ABCD)[10]; and with data available in regional and global portals like USDA-GRIN, EURISCO and GENESYS.

16. Other sources of information were also consulted, such as scientific papers and case studies linking *ex situ* and *in situ*/on farm data from other internationally recognized organizations. The result was an initial list of essential descriptors to identify PGRFA, along with other descriptors describing the sample and its provenance. By focusing on a small mandatory set of descriptors, the Global System would face the challenges to get data registered and would respect the principles of decentralization, ownership and proper attribution.

(c) Dissemination of the survey

17. A list of stakeholders/potential respondents was prepared taking into account different communities such as members of the Task Force on PUIDs;all National Focal Points of the Treaty; experts belonging to the DivSeek Initiative; the Integrated Breeding Platform (IBP), the ECPGR Programme; reviewers involved in the revision of the FAO/Bioversity MCPD List; the private sector, as well as experts taking part in related descriptors consultations and other potential respondents from partner organizations, particularly from developing countries.

18. The distribution list was composed of some 500 experts belonging to different communities with particular attention given to plant breeding and *in situ*/on farm communities. The survey collected inputs from 219 participants from countries in all regions as graphically displayed in Figure 1.

---

[5] ECPGR Concept

[6] Core descriptors for in situ conservation of crop wild relatives v.1.

[7] Contained in Appendix 4 of the Report of the Task Force on Permanent Unique Identifiers.

[8] 'FIGS' - New tool for mining genebank collections

[9] Darwincore-germplasm

[10] Access to Biological Collections Data - ABCD

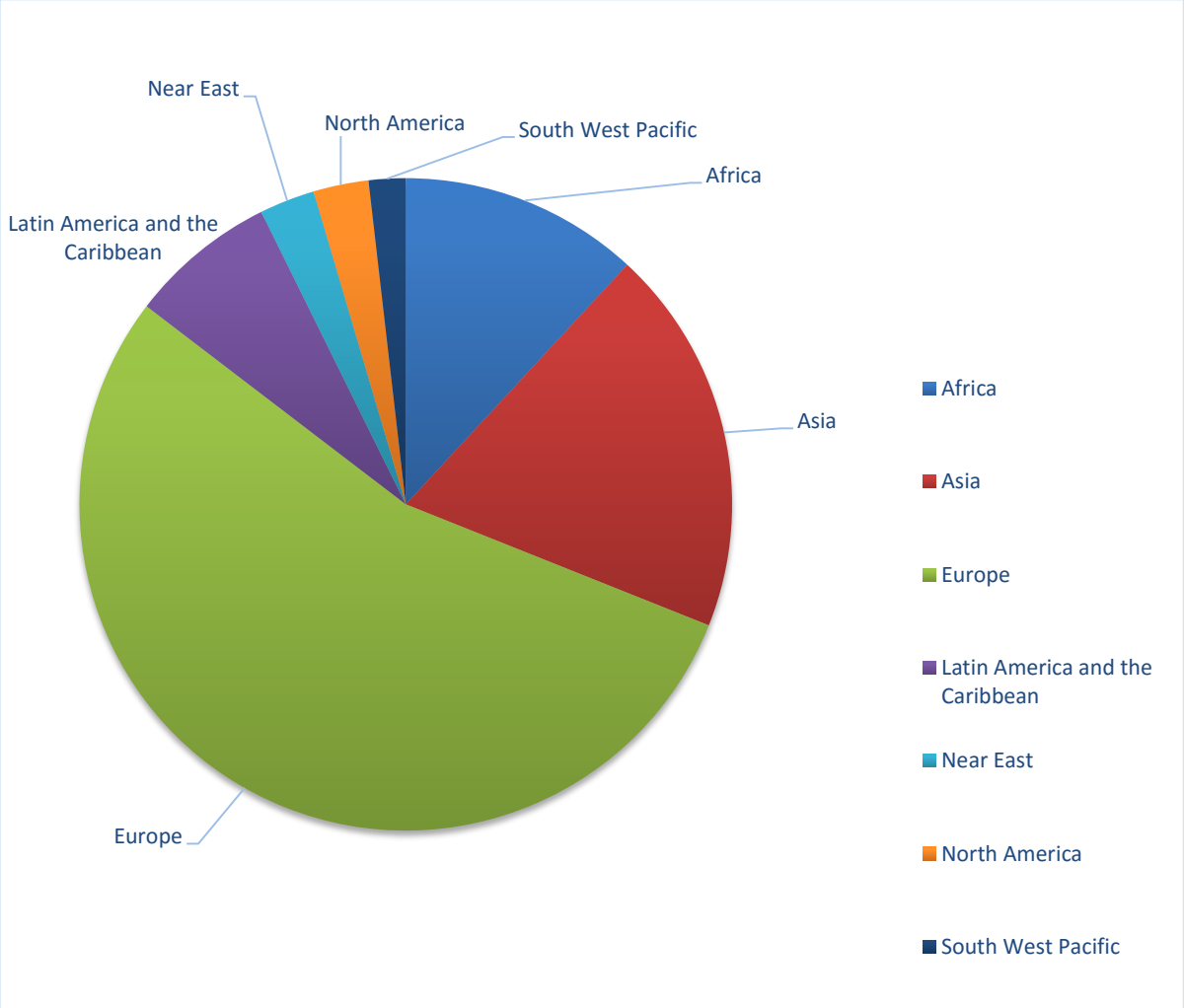

Figure 1. Respondents by region

19. Organizations that participated included, but were not limited to, CGIAR centres (7) ; Universities (17); National programmes and research organizations (INIA, INIAP, INIFAT, NBPGR, NARO, MARDI , CIRAD, INRA, IPK, IBP, ENEA; USDA, among others); several scientists from the private sector (Bayer; GmbH - Nordsaat Saatzucht); FAO National Focal Points; the World Bank, national genebanks, agricultural research institutions, as well as regional and thematic initiatives and networks such as various ECPGR experts.

20. The areas of expertise of survey respondents were also well-balanced and included researchers, breeders, curators, National Focal Points, geneticists and information specialists among others, as shown in Figure 2.

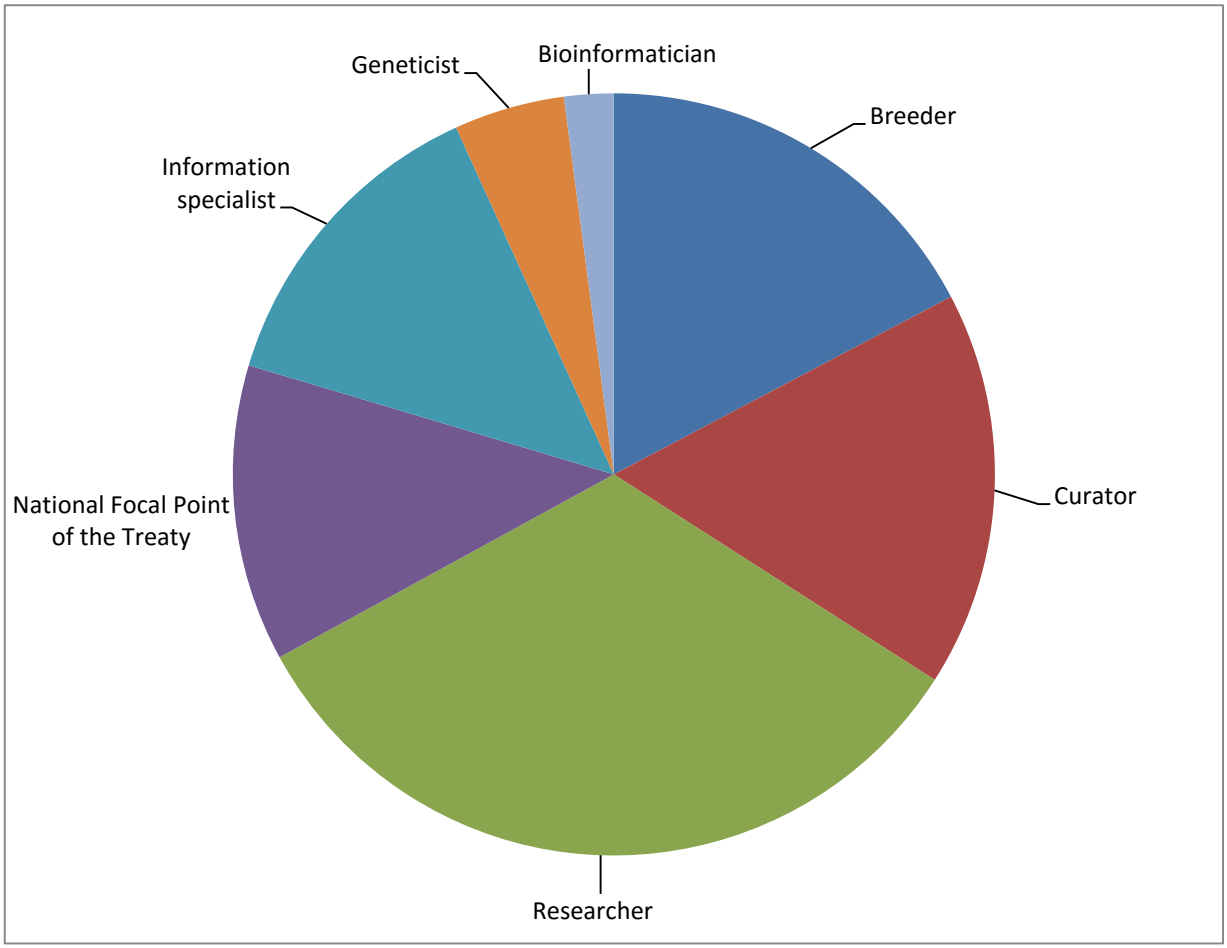

Figure 2. Areas of expertise of respondents

(d) Priority descriptors and major challenges

21. The survey included a set of 'mandatory' descriptors (as given in Table 1) required to uniquely identify all available PGRFA material (not only *ex situ* genebank accessions), transferred under the Multilateral System of Access and Benefit Sharing, with the aim of improving its access to information at the global level and to increase its utilization.

Table 1. Priority descriptors as listed in the survey

1. PGRFA sample identifier: The identifier that you use to identify your PGRFA sample within your collection, organization, laboratory, research institution or activity (e.g. Accession number, selection identifier, etc.)
2. Other PGRFA sample identifiers: Any other identifier assigned to the PGRFA sample in your system (e.g. LSID; system-specific identifiers, etc.)
3. Location: Location where you maintain the sample (e.g. Institute code, Institute (or person) name and address, or other way of identifying your collection, organization, laboratory, research institution or activity)
4. Date: Date on which you created or acquired the sample (e.g. Acquisition date)
5. Genus: Genus name of taxon (e.g. *Oryza*)
6. Species: Species name of taxon (e.g. *sativa*)
7. Crop name: Common name of the crop (e.g. rice)
8. Sample designation: A generic registered or other designation, such as the name of the variety or line or product to which your sample belongs. The same name may be shared by other samples held by you or by others, and thus does not serve to distinguish your sample from other samples of the same variety or line or product (e.g. Emma; Symphony)
9. Method of sample creation: The method by which you created (harvested or acquired) the sample (e.g. Single cross, backcross, Haploids, Collected Sample; Accession into Genebank; Copy In Working Collection; Cultivar Release; etc.)

22. The survey also asked respondents to select other descriptors important to identify PGRFA to which the DOI should be associated. They are presented in Table 2.

Table 2. Other descriptors important to identify PGRFA samples included in the survey*

*Numbers in parentheses on the left-hand side are the corresponding descriptors numbers as published in the FAO/Bioversity MCPD List

10. DATA DESCRIBING YOUR SAMPLE
Species authority (7)
Subtaxon (8)
Subtaxon authority (9)
Biological Status of sample (19)
MLS Status of the sample (27)
Location of safety duplicates (25)
Institute maintaining safety duplicates (25.1)
Type of germplasm storage (26)
11. DATA DESCRIBING THE IMMEDIATE SOURCE OR PARENT OF YOUR SAMPLE
Donor Institute Code/name (22, 22.1)
Donor Accession Number (23)
12. DATA DESCRIBING THE ORIGIN OF YOUR SAMPLE, IF IT WAS ORIGINALLY COLLECTED FROM *IN SITU* OR ON FARM CONDITIONS
Collecting Mission Identifier (4.2)
Collecting Institute Code (4)
Country of Origin* (13)
Collecting date of sample (17)
Location of collecting site (14)
DEC Latitude of collecting site (15.1)
Latitude of collecting site (15.2)
DEC Longitude of collecting site (15.3)
Longitude of collecting site (15.4)
Coordinate uncertainty (15.5)
Coordinate datum (15.6)
Georeferencing method (15.7)
Elevation of collecting site (16)
Collecting/acquisition source (21)
13. DATA DESCRIBING THE ORIGIN OF YOUR SAMPLE, IF IT WAS BRED *EX SITU*
Date of creation of original sample
Pedigree / ancestral Data (20)
Breeding Institute Code/Name (18)
Country of Origin* (13)

(e) Weighted responses

23. Results from Tables 1 and 2 were then analysed and descriptors were ranked by percentage of importance. This percentage was calculated by multiplying the number of experts that considered the descriptor very important by 100, and dividing the result by the number of respondents that took part in the survey.

24. To avoid any possible mistake in preparing the results, also rating averages were performed making possible to obtain detailed statistical information. This rating is a weighted average per column. Each rating scale choice was assigned a value. A sum was made of the weighted values of the number of respondents who picked the positive, negative, or skipped the questions. Then, the weighted value calculation was divided by the sum of respondents. The order of priority assigned by the respondents to the descriptors based on average rating higher than 100 is presented in Figure 3 below.

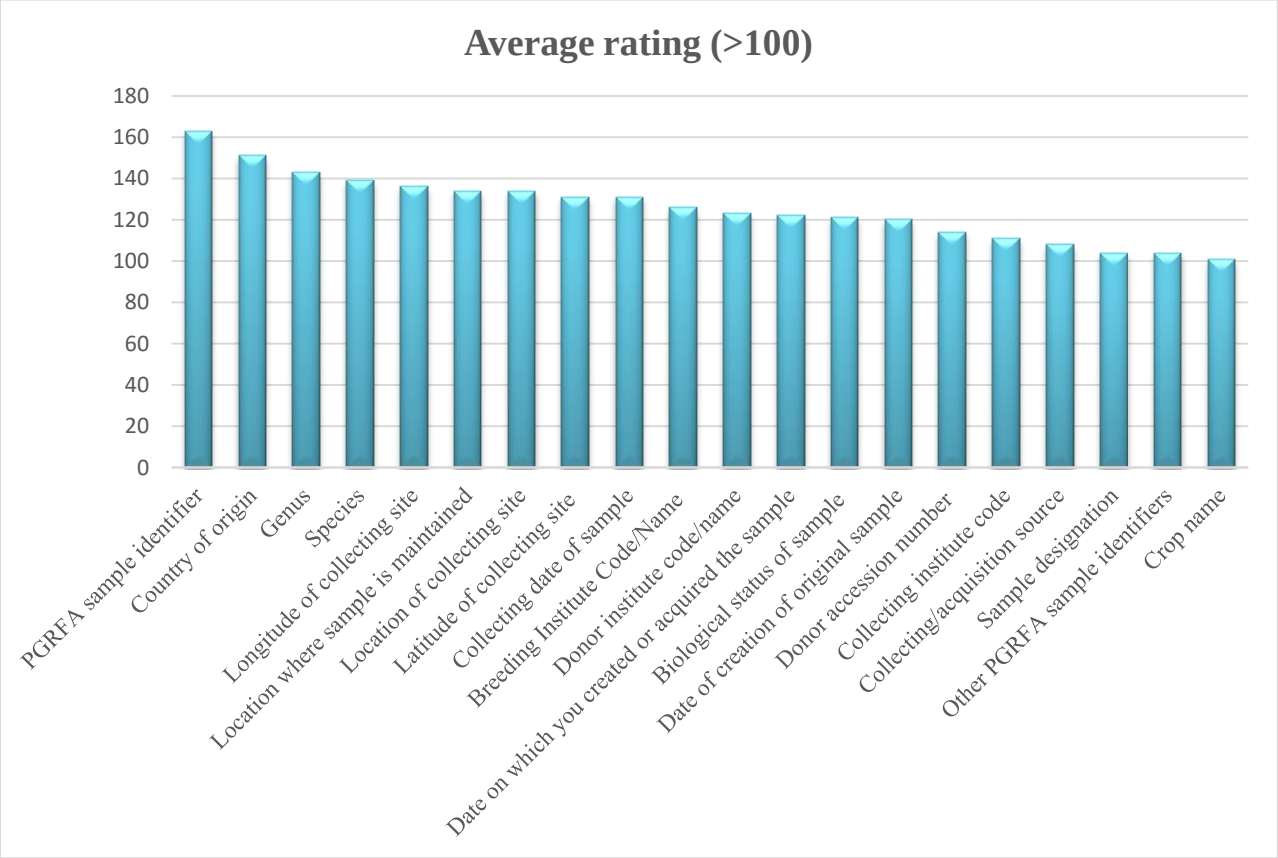

Figure 3. Descriptors resulting from the survey

(f) Major outcomes

25. Preliminary results showed that there was high consensus on some descriptors to be considered mandatory and others requiring further consultation. The analysis also indicated that: (i) there were no major concerns with most of the descriptors proposed in the survey, thereby supporting its validity, (ii) some additional degree of flexibility with existing descriptors was considered necessary; and (iii) a few new descriptors useful to identify non-genebank material (e.g. Population identifier), would be required.

26. In addition to these results, the survey collected also more than 50 comments including views and opinions about the content of the Global System. Many respondents declared in the Remarks section that for uniquely identifying PGRFA samples, only three descriptors were required, namely: FAO Institute Code/Name, Genus and Accession number, or permanent unique identifiers if available, while few of them indicated that only *one* PGRFA sample identifier was needed as the PGRFA sample ID should allow to retrieve all other data and thus it would suffice as many mandatory descriptors are available in the collecting protocols of genebanks and of other collectors.

27. Besides, several respondents suggested that all of the MCPD descriptors, including additional taxonomic information such as species authority and subtaxon, should be highly recommended when reporting PGRFA data as it would not be possible to distinguish, for example, broccoli from cauliflower, or the cultivated taxon from its wild ancestor.

28. Notwithstanding most communities usually assign their own *local* identifiers to PGRFA that they conserve (e.g. genebank community), there are different standards being used by other non-genebank communities. For example, most communities do not assign their own unique identifiers to PGRFA that they acquired from others, and commonly they refer to their samples only by species or variety or crop name. The survey revealed there was a need for further research on this to accommodate all these differences.

29. Other concerns indicated that definitions for ‘Method of creation’ or ‘Sample designation’ lacked clarity, so they were unsure if required or not.

30. Several respondents suggested that the DOI’ system should allow for adding many secondary identifiers and other numbers associated with the material to facilitate the generation of added value. Also the identifier of the donating institution (donor institute code, donor accession number) and the first assigned identifier such as collecting institute code, collecting number or breeding institute code, line number and landrace or variety name should be stored.

31. Respondents also indicated that DOIs should allow additional descriptors related to regeneration activities for a single seed lot that is exchanged (i.e. location, year and seed lot number). Although it does not directly contribute to the identification of the seed lot, it was recommended to keep track of the provider/donor, collector, date of collection and country of provenance.

32. These data are needed to fulfil legal requirements and is also important because genetic resources are often exchanged between collection holders. Therefore seed lots with the same accession name might no longer be identical (due to genetic drift or labelling mistakes), or germplasm coming from different origin but with similar variety names might not be the same, therefore clarity on the original source is essential. The DOI system should be able to model relationships between records.

IV. *Second phase*: FOCUS GROUP CONSULTATION TO VALIDATE SURVEY RESULTS AND ENSURE THEIR WIDE APPLICABILITY (N15)

33. Taking into account comments received, the Secretariat of the Treaty set up a Focus Group Consultation. It consisted of 15 key experts from the user and stakeholder groups belonging to different organizations and areas of expertise to ensure the relevance and wide applicability of the descriptors resulting from the Global Survey. Special attention was paid to engage skilled experts from the breeding community.

34. The core list of descriptors to be associated with a DOI obtained from the survey was harmonized with feedback received from the Focus Group consultation and indicated that there was unanimous consensus on few descriptors and showed that responses gathered were similar on few categories of data which would answer key questions: What, Where, and Which that are required to identify material. Answers to these questions would address each one of the different data types as they constitute a formula for data-gathering to get the complete story on a subject, in this case, the PGRFA material.

(a) Priority descriptors

35. The result of the Focus Group Consultation is presented in Figure 4.

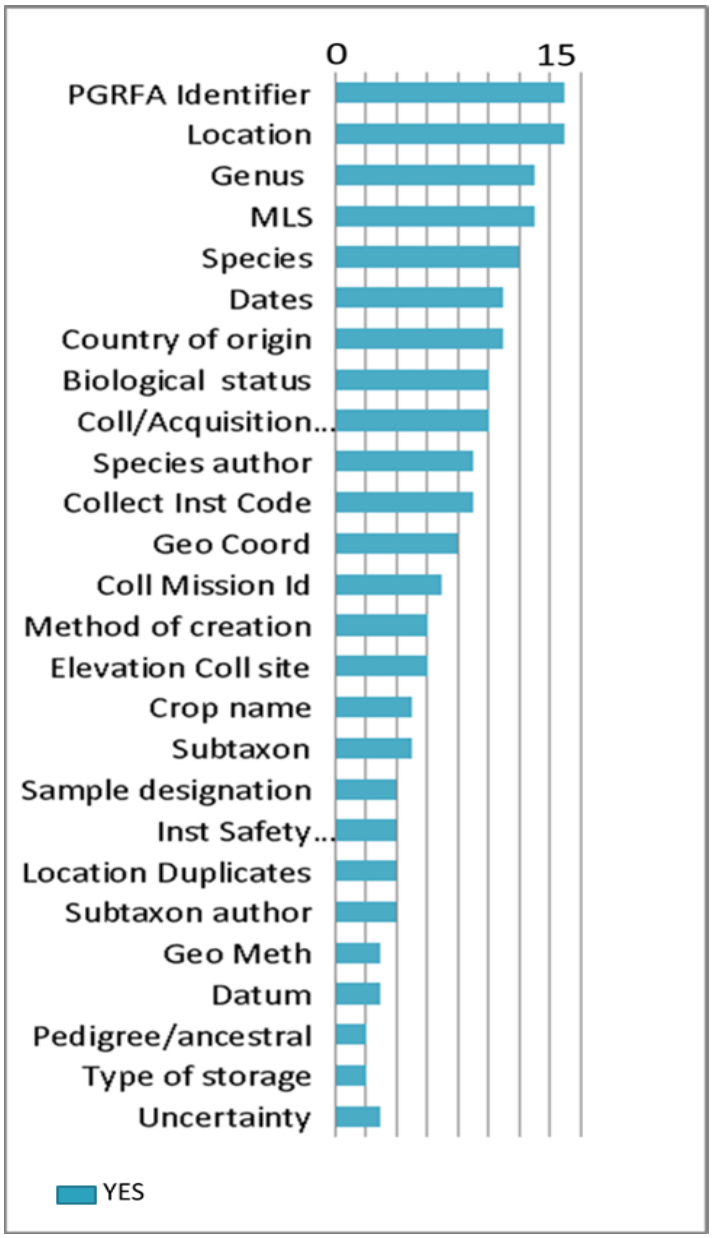

Figure 4. Priority descriptors resulting from the Focus Group Consultation

(b) Challenges for the registration of data in the Global System

36. Some of the most interesting outcomes of the Focus Group are the recommendations listed below which will need to be addressed:

- MLS status got a medium rating average in the Global Survey, but unanimous consensus by the Focus Group experts as a mandatory descriptor.

- Although 'Method of creation' got medium scoring in both consultations, it was felt by some experts as critical to uniquely identify samples.

- A new additional descriptor related to historical *vs* active material was proposed to add, nevertheless considering that GLIS should initially contain data on *available* material, it is felt that it is not required at this stage, but it could be required in the long run.

- All MCPD descriptors that are available for a sample should be 'mandatory'.

- Descriptors that are not applicable for certain types of plant material could be provided as 'NA'(not applicable).

- The DOI registration system should be able to allow users to register non-mandatory descriptors.

- It is best to promote a minimum set of mandatory descriptors as simple as possible to facilitate the adoption of DOIs, rather than trying to get everyone to agree on a wider list.

- There are descriptors corresponding to the 'Location' or PGRFA identifier that may be vague or confusing for users. Besides, geographical coordinates that are required for wild species, though receiving a good rating, are not always available for other types of material.

(c) Looking forward

37. There is a need to develop guidelines for the adoption and optimal use of DOIs to assist users during the registration process. In particular, the development of a set of basic rules for users to determine when to assign them and which descriptor to use according to the type of material they hold.

38. The guidelines should also describe the main features and benefits of using DOIs associated to germplasm samples and provide information on the modalities in which the information can be uploaded.