

April 2021



**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**

**FOURTH MEETING OF THE SCIENTIFIC ADVISORY COMMITTEE ON THE
GLOBAL INFORMATION SYSTEM**

20–21 April 2021

UPDATE ON THE DOI MODULE

I. INTRODUCTION

1. This document provides an update on the module for the management of Digital Object Identifiers (DOIs) and, in particular, on the latest development of the Graph Browser. The browser is a tool for plant breeders and PGRFA users to benefit from the relationships and links established by the holders of the material when creating new materials or receiving materials that have a DOI assigned. The Secretariat will make a presentation to the Committee with detailed illustrations at the meeting.
2. This document further explains the links between the Graph Browser and DOIs, its functioning and its main features. It also invites the Committee to provide advice on the mechanism to obtain feedback from additional stakeholders to improve the prototype.

II. BACKGROUND

3. One of the value-added services offered by the use of the Digital Object Identifiers to PGRFA users is the maintenance of the *relationship graph*, i.e. a representation of the parent-child relationships among those PGRFA to which a DOI has been assigned. Such a graph is populated using the information provided upon DOI registration, namely descriptors M04: Method and R02: DOI(s) of Progenitor(s).
4. Four of the six acquisition methods currently supported by DOIs for food crops actually establish parent-child relationships as explained in Table 1 below.¹
5. As of the end of March 2021, the GLIS relationship graph contained 30 583 nodes representing the PGRFA and 36 459 relationships among them distributed across the four methods as follows: 1 225 under acquisition, 107 under in-house copy, 14 177 under in-house variant and 20 950 under novel distinct PGRFA.
6. In close collaboration with the FAO Information Technology Division, the Secretariat analysed several tools to exploit the properties of the relationship graph fully. A joint decision was made with FAO to store the graph in a dedicated database using Neo4J, a product intended to perform optimised operations on the specific structure of the graph. Neo4J complements the traditional relational database management system (RDBMS - PostgreSQL) that powers the rest of the DOI Management Module. The two databases are managed and kept into synchronisation through the business logic.

¹ For full details on methods, please refer to "Data required for the assignation of Digital Object Identifiers in the Global Information System v.2.1" available at <http://www.fao.org/3/bt113e/bt113e.pdf>

Method	Description	Relationship	Parents	Label
Acquisition	The PGRFA was obtained from someone else or collected in the field	Yes	1	acfr
In-house copy	The PGRFA was created to be as much as possible a genetic copy of the parent	Yes	1	cpfr
In-house variant	The PGRFA was subsampled or harvested from a genetically variable parent	Yes	1	defr
Novel distinct PGRFA	The PGRFA was obtained through a process that makes it genetically distinct from its parent(s)	Yes	1 or more	crfr
Observation – Natural	The PGRFA is held in situ and appeared without holder's intervention	No	N/A	N/A
Inherited	The PGRFA was present when the holder took office	No	N/A	N/A

Table 1. Four of the DOIs acquisition methods establish relationship that can be displayed graphically.

III. PURPOSE

7. The purpose of the design and implementation of the Graph Browser is manifold. Firstly, it is intended to provide plant breeders, PGRFA holders and other users with a powerful yet simple to use tool to **explore the relationship graph** containing valuable information on the ancestry and progeny of a given PGRFA, across organisations and after multiple genetic modifications and/or transfers.

8. Secondly, it benefits from the content declared by users under “Highly Recommended Descriptors” and it may become one strong **incentive** in the future to provide this information. These descriptors are not required to obtain a DOI as they may not always be available. But they represent valuable information about the PGRFA and, if available, allow GLIS to provide more advanced services to its users, such as the Graph Browser.

9. The Graph Browser visually connects the work of genebanks, plant breeders, researchers and other communities, and it also works as a **dynamic discovery tool**. The ability of DOIs to capture these complex relationships was one of the key aspects considered for their adoption as permanent unique identifiers in the PGRFA context. At the same time, the development of such a function was highly demanded by users and stakeholders.

10. Lastly, the current, initial implementation of the Graph Browser aims at stimulating discussions among stakeholders and in the scientific community about concrete **use cases** that could be addressed by future Graph Browser releases. The functions currently provided are based on an assessment of the information available and a preliminary consultation with a reduced group of potential users.

IV. MAIN CHARACTERISTICS

11. The Graph Browser, in its initial implementation, offers two main characteristics. Firstly, the graph can be **interactively explored** by simple queries and visually adding new nodes. The mouse is used to navigate through the graph, zoom in and out, obtain details on a node and expand a node to further investigate its neighbourhood.

12. Secondly, it is very **user-friendly**. Indeed, key queries are performed using a form that hides the complexity of the two databases to which they are addressed. The user does not need to know where a specific information is located or how each database is queried.

V. STAKEHOLDER FEEDBACK

13. One of the purposes of the prototype of Graph Browser is to solicit feedback from potential users. It would be interesting to gather further details about the kind of research stakeholders would like to be able to perform using the Graph Browser, besides the ones already identified.

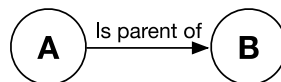
14. The Secretariat will publish a demo online to collect feedback through interviews. In addition, interested stakeholders are invited to send their input to pgrfa-treaty@fao.org using short user stories such as "As <role>, I would be able to <operation> in order to <purpose>". In one of the user stories already supported, <role> would be something such as "genebank manager" or "breeder"; <operation> would be "find all PGRFA that are genetically similar or identical to the starting node and for which genomics information is available" and <purpose> would be "investigate the genetic identity of the material". Additional narrative would be welcomed to better understand what potential users may look for.

15. All the inputs received will be analysed, categorised and assessed against the functions available in the Graph Browser. They may also help the Scientific Advisory Committee to identify and provide advice on additional strategic partnerships to pursue.

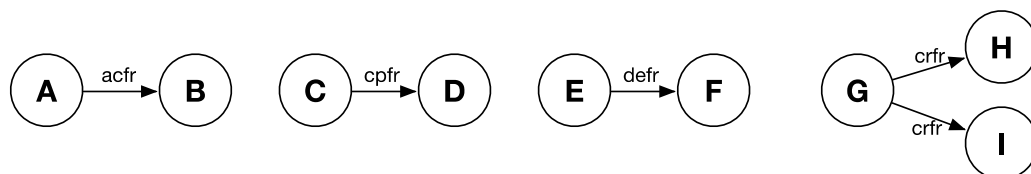
VI. THE GRAPH BROWSER PROTOTYPE

16. The Secretariat will present a demonstration of the prototype at the meeting. What follows is information on how the browser represents the relationships among PGRFA.

17. A graph is a visual representation of the relationships among entities (called nodes). Each node has one or more lines, called arcs, connecting it to other nodes. For example, a relationship "is parent of" is represented in a graph with two nodes, the parent (A) and the child (B) and an arc connecting them. As the relationship is not symmetrical (the parent and the child cannot be exchanged without completely altering the meaning of the graph), the arc has a direction indicated by an arrow at one of the ends and the graph is called "directed". This way, the graph can be read as "A is parent of B" as in the diagram below.



18. The labels assigned to the methods in the table above allow us to read the relationship graph in a similar way. For instance, the examples below, representing methods Acquisition, In-house copy, In-house variant and Novel distinct PGRFA, respectively, can be read as "A is *acquired from* B", "C is *copied from* D", "E is *derived from* F" and "G is *created from* H and I". The labels are shorthand for the text in italics; this helps reduce the clutter in the graph visualisation



19. In graph parlance, moving from one node to another along an arc (e.g. from A to B above) is called a "hop". All examples above contain a single "hop" except the rightmost one where H is 2 "hops" from I: the first "hop" takes you from H to G and the second one from G to I (ignoring the direction of the arcs). Of course, also I is 2 "hops" from H and G is one "hop" from both H and I.

20. The current implementation of the Graph Browser has been designed and implemented more as a proof of concept than as a final product. As explained above, one of its purposes is to

stimulate potential users to participate in its conceptual design offering new features to be considered for future releases until a production-ready version is deployed.

21. To minimise costs, little attention has been devoted to the appearance of the tool as well as to its performance. The focus has been on providing a working implementation to show the potential of navigating across the relationship graph through a graphic, interactive browser.

22. Furthermore, as the number of nodes and arcs in the relationship graph is limited at this time, a demo dataset has been prepared to show Graph Browser's functions in a more meaningful way. In the following, a brief description of the functionalities is provided, along with some commented screenshots.²

23. Access to the Graph Browser will be available, in principle, from the GLIS detail page of any PGRFA included in the relationship graph. For efficiency reasons, only PGRFA that are related to others using any of the four methods described above is actually represented as a node in the graph. Figure 1 shows the "Browse graph" button that appears on the PGRFA detail page. In the future, access to the graph could be made possible from other areas, as appropriate.

PGRFA doi:10.18730/59KM6

Citation: <https://doi.org/10.18730/59KM6>

Main descriptors **Acquisition** **DOI info**

Organization/individual conserving the PGRFA International Rice Research Institute
DAPO BOX 7777
1301 Metro Manila
Philippines
WIEWS code: PHL001 [Details]
Easy-SMTA PID: 00AB40

Local identifier IRIS 4-10190

Date 1986-03-24

Creation method Acquisition **Browse graph**

Taxon Oryza sativa L. var. glaberrima Marsella

Common name Rice

Biological status
Names CO 39
Other identifiers
MLS status
Historical No

Figure 1: Button "Browse graph"

24. Clicking the "Browse graph" button opens a new window as reported in Figure 2 below.

10.18730/59KM6

Max hops: 1 Method: - Any - Keywords: - Any - Query and add Query and replace 1 nodes found

Node label: doi Start from selection Show node details Download as Excel Clear filter Show help

ID	DOI	HPID	HWIEWS	Local ID	Genus	Species	Subtaxa	Method	BioStatus	MLS Status	Provenance
173684	10.18730/59KM6	00AB40	PHL001	IRIS 4-10190	Oryza	sativa	subsp. indica	Acquisition			IND

Figure 2: Initial Graph Browser window

² Please keep in mind that the PGRFA showed belong to a demo dataset in a test environment.

25. The following options are now available:

- the graph can be zoomed back and forth by using the mouse wheel;
- a node can be selected by clicking on it (its colour changes to red);
- a node can be moved by dragging it with the mouse;
- moving the mouse over a node displays a panel with the information about the PGRFA;
- clicking on a node while pressing the SHIFT key opens a new window with the GLIS detail page of the corresponding PGRFA;
- the table at the bottom of the page displays information about the nodes shown in the graph. Clicking on a row in the table selects (paints in red) the corresponding node in the graph. Likewise, clicking on a node selects the corresponding row in the table;
- double-clicking on a node "expands" it showing its immediate neighbours, regardless of the filtering and search criteria.

26. The controls in the green strip are organised into two rows. The top one deals with querying the graph; the bottom one contains general controls. The query controls in the top row are:

- "Max hops" indicates the maximum number of arcs connecting the resulting nodes. If "Max hops" is set to 2, the resulting graph will contain only nodes that are at most 2 "hops" away from the selected node;
- "Method" defines the method that all nodes along the path must match. For example, you can look for PGRFA that have been "acquired from" the current one;
- "Keywords" lets you specify which keyword should be present in the links to external websites associated with the PGRFA. For example, you can look for PGRFA that are associated with "Phenomics" links;
- "Query and add" performs the query indicated by "Max hops", "Method" and "Keywords" and adds the nodes matching the three conditions to the graph;
- "Query and replace" performs the query indicated by "Max hops", "Method" and "Keywords" and replaces the nodes currently in the graph with the ones matching the three conditions;
- at the far right of the first row, the number of nodes added by the last query is reported.

27. The three conditions "Max hops", "Method" and "Keywords" must be satisfied simultaneously (they are combined in logical AND).

28. The bottom line of the green strip contains the following controls:

- "Node label" allows you to change the node label. For instance, you can have nodes labelled with the holder's WIEWS or their provenance;
- "Start from selection" opens a new window with only the selected node allowing you to continue from there in a new browser window. The previous window is still available in your web browser and you can go back to it or just close it;
- "Show node details" opens the GLIS detail page on the selected node. You can obtain the same effect by clicking a node while holding down the SHIFT key;
- "Download as Excel" exports the table content to an Excel document that is downloaded to your computer;
- "Clear filter" resets all filters in the table header;
- "Show help" shows a separate window with the Graph Browser help. The window can be resized and moved as required.

29. When performing a query, the selection criteria set in the green strip can be combined with the table filter. For instance, one could ask to show the graph with all nodes within two hops from the start node, that have been created using the "Acquisition" method, that have

"Phenomics" keywords associated to their links and have Philippines ("PHL") as provenance. The result is show in Figure 3 below.

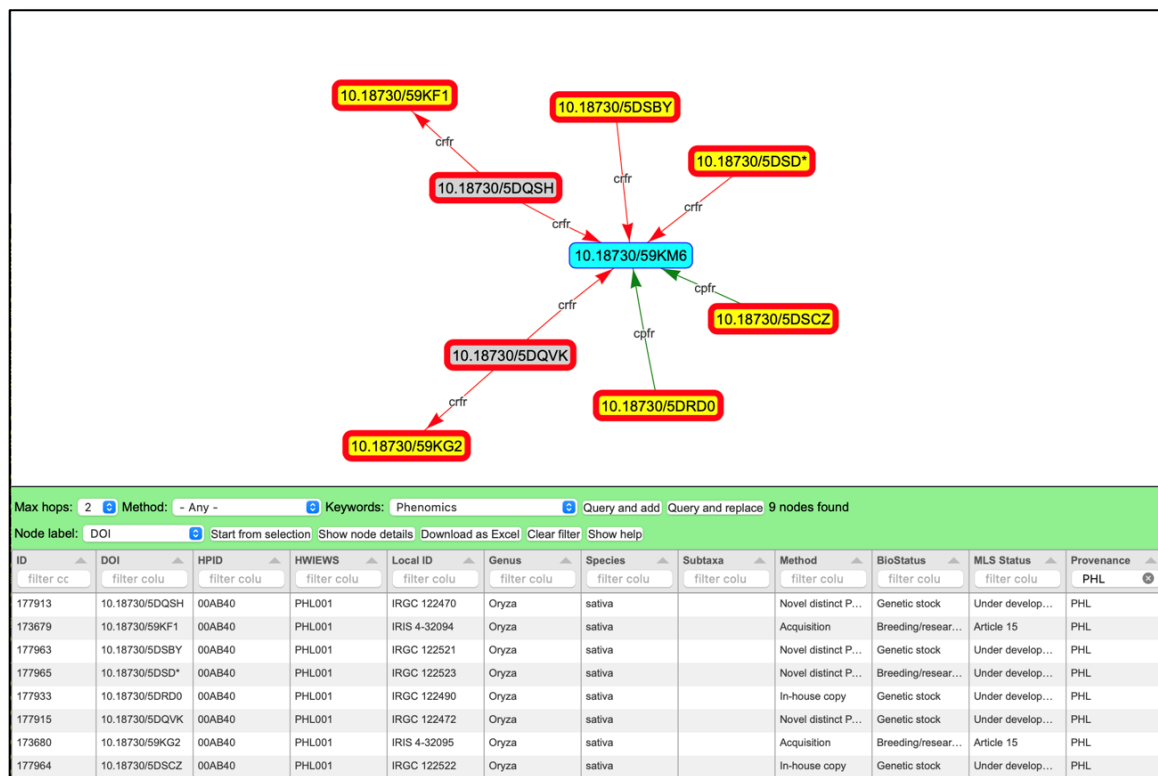


Figure 3: Combined query result

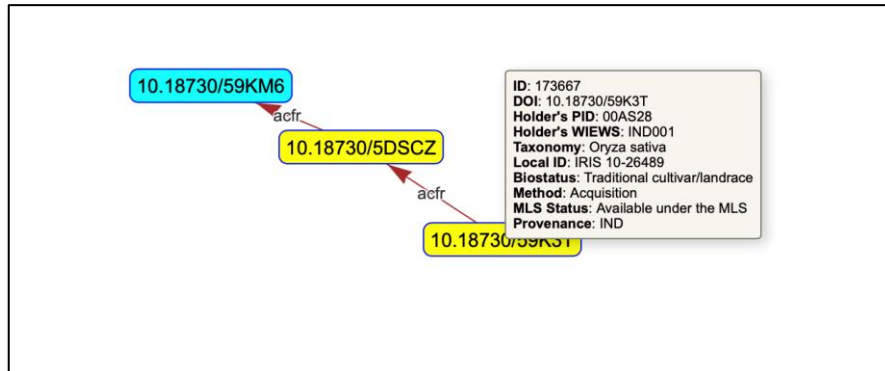
30. As displayed in Figure 3, the nodes and arcs are coloured according to the following tables:

Arc label	Colour
acqu	brown
cpfr	green
crfr	red
defr	blue

Node colour	Description
Yellow	Newly added nodes (by last query or node expansion)
Orange	Node has been added by last double-click
Blue	Node that has been expanded by double-click
Light blue	Node that was present before last operation
Red	Node selected by click
Grey	Newly added intermediate nodes (does not satisfy query, but is in the path to a node which does)
Red border	Node matching table filter

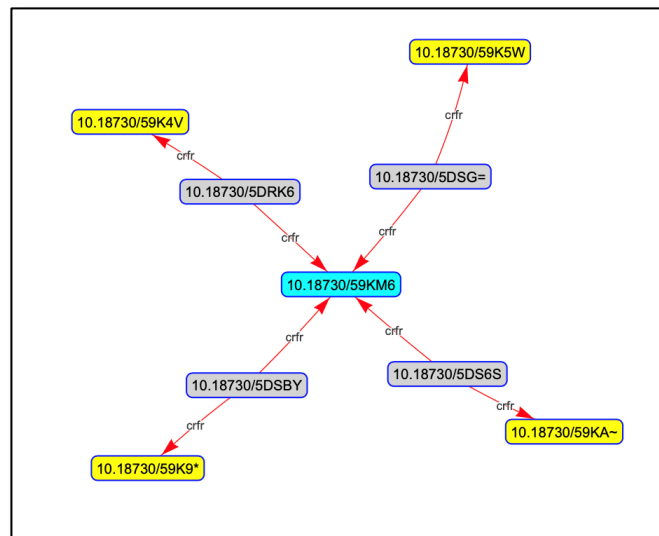
VII. EXAMPLE QUERIES

31. In the following, a few example queries that can be performed by the current version of the Graph Browser are presented. For the sake of simplicity, only the graph is presented. The first query looks for nodes within two hops from the starting node (DOI 10.18730/59KM6) that have been acquired. The screenshot also shows the detail panel that appears when the mouse is moved onto a node.



Example 1: PGRFA acquired within 2 hops

32. The second query looks for PGRFA within two hops from DOI 10.18730/59KM6 that are not under the Multilateral System. Please remark the node shaded in grey. They are intermediate nodes that may not match the query criteria but that are necessary to keep the graph connected.



Example 2: Related PGRFA within two hops that are not under the Multilateral System.

33. The next example involves looking for PGRFA with "Genomics" among keywords associated with their links and within three hops from the starting node. The example shows the Local ID (Accession Number and other identifiers) as node label. Node "IRGC 127093", having been expanded through double-click, is coloured in orange, and its neighbour "IRIS 4-27076" is coloured in yellow, having just been added to the graph.

