The International Treaty
ON PLANT GENETIC RESOURCES FOR FOOD AND AGRICULTURE

E

INTERNATIONAL TREATY ON PLANT GENETIC RESOURCES FOR FOOD AND AGRICULTURE (IT-PGRFA)
SECOND TECHNICAL CONSULTATION ON INFORMATION TECHNOLOGY SUPPORT FOR THE IMPLEMENTATION OF THE MULTILATERAL SYSTEM OF ACCESS AND BENEFIT SHARING OF THE INTERNATIONAL TREATY
Rome, 2-3 December 2008
Information management in support of the global system for the conservation and sustainable use of plant genetic resources for food and agriculture - Global Information on Germplasm (GIG)

1.   The Governing Body of the International Treaty at its Second Session "*stressed the need for closer and effective cooperation with the Global Crop Diversity Trust*"[1] and "*recognized the importance of collaboration between the Secretary of the Treaty and the International Agricultural Research Centers in the implementation of the Treaty. It expressed appreciation for the joint programme set up by FAO and Bioversity International to provide technical assistance to facilitate developing country implementation of the Treaty, and in particular, to initiate the Multilateral System*".[2]

2.   In furtherance of this guidance from the Governing Body of the Treaty and in extension of existing collaboration between the Interim Secretariat of the Treaty and Bioversity International, a collaborative project on information systems has been developed for information management in support of a global system for the conservation and sustainable use of plant genetic resources for food and agriculture. The project's implementation is through a partnership between the Global Crop Diversity Trust (GCDT), Bioversity International and the IT-PGRFA Secretariat.

3.   As part of its mission to ensure the long-term conservation and availability of crop diversity worldwide, the GCDT has supported the development of crop and regional conservation/use strategies. These strategies and other consultations over the past years have highlighted the need for genebanks to better manage and share accession-level information in order to facilitate and stimulate use of collections. The Trust therefore invited Bioversity International, on behalf of the CGIAR System-wide Genetic Resources Programme (SGRP), to

---

[1]   IT/GB-2/07/REPORT, Paragraph 59

[2]   Ibid., pargraph 84

address this important impediment to building an efficient and effective global system for the conservation and use of plant genetic resources for food and agriculture.

4.		Since 2006, the Centres of the Consultative Group on International Agricultural Research (CGIAR) have been leading a series of major informatics projects, including for the design, implementation and promotion of a one-stop entry point for information on and access to the collections held in trust for the world community by the CGIAR Centres. Through the CGIAR System-wide Information Network for Genetic Resources (SINGER), users can already search across individual Centre genebank holdings to identify germplasm samples of interest from a single entry point.

5.		Based on work undertaken by the Interim Secretariat of the Treaty, in 2007 Bioversity began collaboration with the Treaty Secretariat to develop prototype IT modules for ordering germplasm under the Treaty's Multilateral System of Access and Benefit-sharing (MLS). The above-mentioned current project, entitled "Global Information on Germplasm (GIG): Information management in support of the global system for the conservation and sustainable use of PGRFA", will build upon this collaboration with the IT-PGRFA Secretariat and its outputs will be fully compatible with the 'IT-PGRFA module' for ordering and exchanging germplasm. This will ensure that all transfers of MLS material facilitated by the accession level information system resulting from the GIG project will be fully compliant with the operation of the Standard Material Transfer Agreement of the Treaty.

6.		Annex 1 of this document contains the short version of this project.

Annex 1

A. PROJECT SUMMARY

Global Information on Germplasm (GIG): Information management in support of the global system for the conservation and sustainable use of PGRFA
Abstract: The overall goal of the Global Information on Germplasm (GIG) project is to improve access by breeders and other users to the germplasm they need in genebanks around the world. The project specifically addresses the challenge of making information about germplasm collections readily available, which has been repeatedly identified as a key to increasing use. The project has three components: developing information standards to describe the characteristics of genetic resources of most interest to users; deploying a genebank data-management system appropriate to users in developing countries; and developing a global accession-level information system. The first component will mobilize germplasm users to agree on core subsets of characterization and evaluation standards. These will be incorporated into the state-of-the-art GRIN-Global genebank data-management system being developed by USDA through a separate project supported by the Trust. GRIN-Global will be made available to national programmes and other genebanks through the second component. Finally, the third component of the project involves the development of a global system for accessing and managing accession-level data in support of conservation and use of PGRFA by linking up national, regional and international genebank databases.
Crops/species: All major food crops, including CGIAR mandate crops, with a focus on the following 22 crops: banana, barley, beans, breadfruit, cassava, chickpea, coconut, cowpea, fababean, finger millet, grass pea, maize, major aroids, lentil, pearl millet, pigeon pea, potato, rice, sorghum, sweet potato, wheat and yam, and a limited number of other crops of interest to the project collaborators.
Countries: The project is worldwide in scope.

Activities (check relevant boxes):

☐ Seed storage   ☐ Safety duplication

☐ Storage *in vitro*   ☐ Safety backup at Svalbard Global Seed Vault

☐ Field maintenance   ☐ Conservation research/protocols

☐ Regeneration   ☐ Plant health research (safe movement)

☐ Characterisation   ☒ Collection management standards/procedures

☐ Evaluation   ☒ Training

☒ Documentation/database development   ☐ Other (specify):

☐ Germplasm distribution

☒ Information systems

I. Applicant Institution:	Bioversity International, as the Convening Centre of the of the CGIAR System-wide Genetic Resources Programme (SGRP)
Principal investigator	Michael Mackay
	Senior Scientist, Biodiversity Informatics Project Coordinator
Co-Principal investigator:	Elizabeth Arnaud
Title:	Scientist, Understanding and Managing Biodiversity Programme
Address:	Via dei Tre Denari 472/A 00057 Maccarese, Rome, Italy

Key collaborating institutions

II.	1. Collaborating Institution:	Secretariat of the ITPGRFA
	Principal investigator:	Shakeel Bhatti
	Title:	ITPGRFA Secretary
	Address:	FAO Headquarters Viale delle Terme di Caracalla, 1 00153 Rome, Italy
	Telephone:	+39 06 5705 3057
	Fax:	+39 06 5705 6347
	Email:	pgrfa-treaty@fao.org

III.	2. Collaborating Institution:	SGRP
	Principal investigator:	David Williams
	Title:	Coordinator
	Address:	c/o Bioversity International Via dei Tre Denari 472/A 00057 Maccarese, Rome, Italy
	Telephone:	+39 06 61181
	Fax:	+39 06 6197 9661
	Email:	d.williams@cgiar.org
	Principal investigator:	Brigitte Laliberté
	Title:	Coordinator of GPG2 Project
	Address:	Bioversity, Via dei Tre Denari 472/A 00057 Maccarese, Rome, Italy
	Telephone:	+39 06 6118 272
	Fax:	+39 06 6197 9661
	Email:	b.laliberté@cgiar.org

IV.	3. Collaborating Institution:	ICWG-GR (SGRP steering committee)
	Principal investigator:	Jean Hanson
	Title:	Chair
	Address:	International Livestock Research Institute P.O.Box 30709 Nairobi 00100, Kenya
	Telephone:	+254 20 422 3000
	Fax:	+254 20 422 3001
	Email:	j.hanson@cgiar.org

V.	4. Collaborating Institution:	United States Department of Agriculture
	Principal investigator:	Peter K. Bretting
	Title:	National Program Leader, Plant Germplasm and Genomics
	Address:	USDA/ARS National Program Staff 5601 Sunnyside Avenue Beltsville, MD 20705-5139
	Telephone:	+1 301 504 5541
	Fax:	+1 301 504 6191
	Email:	peter.bretting@ars.usda.gov

Total budget:	US$3,538,780 (including in-kind contributions)
Amount requested from Trust:	**US$1,675,000**
Duration of the project:	June 2008 through March 2011

B. PROJECT NARRATIVE DETAIL

Project background and justification

As part of its mission to ensure the long-term conservation and availability of crop diversity worldwide, the Trust is supporting the development of crop and regional conservation strategies. These strategies – and other consultations over the past years – have repeatedly highlighted the need for genebanks to better manage and share accession-level information. Users currently find it difficult to access the information they need to make informed decisions about what germplasm to order. The Trust has therefore invited Bioversity International, on behalf of the CGIAR System-wide Genetic Resources Programme (SGRP), to address this important impediment to building an efficient and effective global system for the conservation and use of plant genetic resources for food and agriculture. An invitation to join the partnership will also be extended to member countries of the European Cooperative Programme for Plant Genetic Resources (ECPGR).

The Secretariat of the International Treaty for Plant Genetic Resources for Food and Agriculture (ITPGRFA) has been closely involved in the process of developing this proposal and is a partner in the work. In 2007, Bioversity began collaboration with the ITPGRFA Secretariat to develop a prototype Permanent Identifier (PID) Server for registering users of germplasm through the Treaty's Multilateral System of Access and Benefit-sharing (MLS). This project will build upon this collaboration with the ITPGRFA Secretariat and its outputs will be fully compatible with the 'ITPGRFA module' for ordering and exchanging germplasm.

This partnership is designed to build on a multi-million dollar investment over the past ten years by Bioversity/SGRP to build capacity for sharing germplasm information, particularly through SINGER and EURISCO. In a recent consultation, the Executive Committee of the SGRP Inter-Centre Working Group on Genetic Resources (ICWG-GR) discussed how the expertise of the SGRP can be employed to build such a system, as well as its benefits to SGRP members and beyond.

The Global Information on Germplasm (GIG) project will be built upon the ongoing investment of US$10 million by the World Bank to support SGRP and Bioversity in information management through the project "Collective action for the rehabilitation of global public goods in the CGIAR genetic resources system, Phase 2" (GPG2). The GPG2 project includes activities on refining characterization strategies, developing germplasm inventory systems, improving the quality of location data, creating crop registries and implementing a one-stop entry point for information on and access to the CGIAR collections held in trust for the world through SINGER.

Through this partnership, three inter-related components will be implemented in an integrated fashion:

Component 1: Development of characterization and evaluation data standards for 22 target crops; **Component 2:** Deployment of GRIN-Global, a new genebank information management system developed through a complementary USDA project; and

Component 3: Development of a global accession-level information system ("ALIS").

The first component will enhance the usefulness of existing descriptor lists by mobilizing germplasm users to agree on core subsets of characterization and evaluation data standards. These will be incorporated into a globally deployable version of the Genetic Resources Information Network (GRIN) developed by the United States Department of Agriculture Agricultural Research Service (USDA-ARS). This state-of-the-art genebank information

management system – GRIN-Global – will then be deployed to national germplasm programmes and other genebanks through regional networks. Finally, "ALIS" will provide the foundation for a global system for data sharing in support of the conservation and use of plant genetic resources for food and agriculture (PGRFA) by linking national, regional and international genebank databases. In addition to receiving data from GRIN-Global, "ALIS" will be linked to genebanks using other systems as well as to SINGER and EURISCO, allowing searches of CGIAR and national genebank holdings around the world.

Through "ALIS", users will be able to access an online ordering tool kit (OTK) that will be compliant with the ITPGRFA requirements regarding reporting on Standard Material Transfer Agreements (SMTA). Integration of "ALIS" and the ITPGRFA module (Component 3) for the facilitation and reporting of SMTAs will be critical for: (i) the credibility of the MLS; (ii) the planning and operation of the MLS; (iii) the ordering of material by potential users; (iv) the monitoring and review of the MLS by the Governing Body of the ITPGRFA; and (v) creating a sound basis for policy decisions by the Governing Body on the future evolution of the MLS. The project will yield data standards and protocols for the 22 target crops that can also be applied to other plant species or adopted by partners for other purposes. The three components will be promoted together and a community of practice established through training workshops, training of trainers and outreach.

2. Project goal

The goal of GIG is to enable breeders and other users to make more effective use of *ex situ* collections through better data management and increased accessibility.

3. Project activities and approach

The three components of the project (data standards, the deployment of the GRIN-Global genebank information management system and "ALIS") will be carried out in parallel and with consistent interaction between them to reach the project goal. The teams implementing these components will also interact closely with the project's strategic partners to integrate existing systems, as depicted in Figure 1 below.

Figure 1
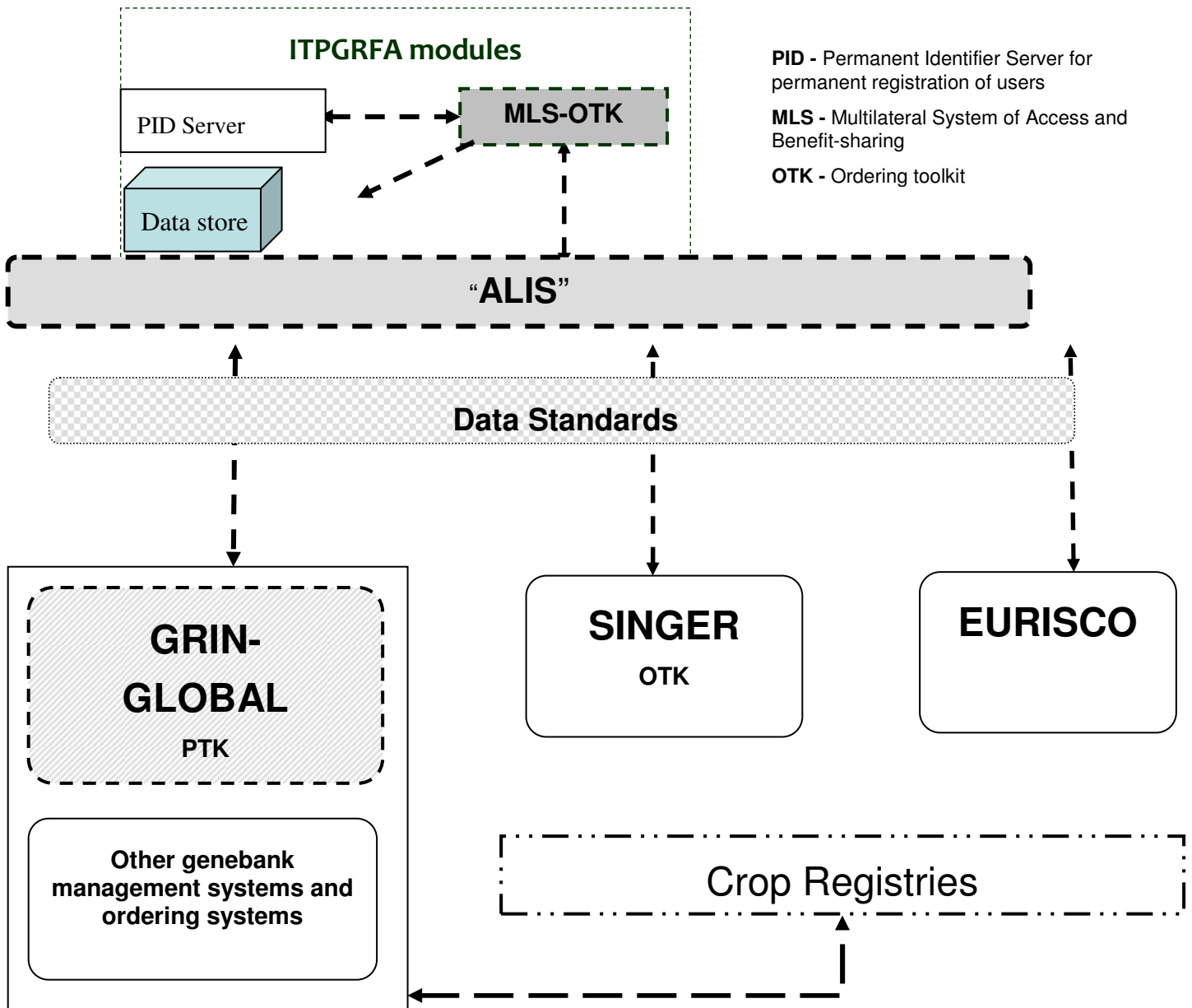
Component 1: Development of characterization and evaluation data standards for 22 target crops

Documentation standards for PGRFA are needed so that information about plant traits can be entered into databases in a uniform manner that will allow cross-searching. To address this need, a set of agreed standards for characterization and evaluation of 22 target crops (banana, barley, beans, breadfruit, cassava, chickpea, coconut, cowpea, faba bean, finger millet, grass pea, maize, major aroids, lentil, pearl millet, pigeon pea, potato, rice, sorghum, sweet potato, wheat and yam) will be developed. This list can be revised if necessary to add other priority crops identified by ECPGR that are not included in Annex 1 of the ITPGRFA, which lists crops covered under the MLS. A summary of past progress in developing data standards for the target crops, and planned meetings of crop experts, can be found in Table 1 below.

Table 1: **Tentative grouping of target crops by previous progress in data-standards**

Crop	Species	CG	GPG2	GCP	Harvest+	DL	Lang	Year	Crop strategy	Planned meetings	Date of meeting
Group 1 Short-term output											
Rice*	*Oryza spp.*	X	X	X	X	X	E	2007	√	No	
Banana*	*Musa spp.*	X	X	X	No	X	E,S,F	1996, 2001	√	Musa Taxonomy Advisory Group Meeting, India (September 2008)	September 2008
Coconut	*Cocos nucifera*	X	No	X	No	X	E	1992, 2007	√	August 2008	August 2008
Grass pea	*Lathyrus spp.*	No	No	No	No	X	E	2000	√	No	
Cultivated potato	*Solanum tuberosum*	X	X	X	No	X	E	1977	√	No	
Taro	*Colocasia esculenta*	No	No	No	No	X	E,F,S	1999	√	No	
Group 2: Medium-term output											
Barley	*Hordeum vulgare*	X	No	X	No	X	E	1994	√	No	
Wheat (Rev.)	*Triticum spp.*	X	No	X	X	X	E	1985	√	International Wheat Genetics Symposium Australia (24-29 August 2008)	24-29 August 2008
Yam	*Dioscorea alata*	X	No	X		X	E,S,F	1997	No	Yam Strategy Meeting, venue to be confirmed (late 2008)	late 2008
Chickpea	*Cicer arietinum*	X	X	X	No	X	E	1993	√	No	
Sweet potato	*Ipomoea batatas*	X	No	X	X	X	E,S,F	1991	√	No	
Maize	*Zea mays*	X	X	X	X	X	E,S,F, P	1991, 2000	√	No	
Sorghum	*Sorghum bicolor*	X	X	X		X	E,F	1993	√	No	
Group 3 : Long-term output											

Lentil	Lens culinaris	X	No	X	No	X	E	1985	√	No	
Pigeonpea	Cajanus cajan	X	X	X	No	X	E	1993	√	No	
Group 3 : Long-term output (cont.)											
Bean	Phaseolus vulgaris	X	No	X	X	X	E,P	1982, 2001	No	Bean Strategy Meeting, venue to be confirmed (mid 2008)	mid 2008
Cowpea	Vigna unguiculata	X	No	X	No	X	E	1983	No	Cowpea Strategy Meeting, venue to be confirmed (mid 2008)	mid 2008
Finger millet	Eleusine coracana	X	No	X	No	X	E	1985	No	No	
Faba bean	Vicia faba		No		No	X	E	1985	No	No	
Pearl millet	Pennisetum glaucum	X	No	X	No	X	E,F	1993	No	Pearl Millet Strategy Meeting, venue to be confirmed (late 2008)	late 2008
Cassava	Manihot esculenta	X	No	X	X	draft	No		No	Crop meeting Cali Colombia (April 2008)	
Breadfruit	Artocarpus altilis	No	No	No	No	No			√	No	

* Rice and banana = Multi-site evaluation; source of evaluation data sets

CG = priority crop of CGIAR

GPG2 = priority crop of SGRP GPG2 project

GCP = target crop of Generation Challenge Programme

Harvest+ = target crop of Harvest+ Challenge Programme

DL = descriptor list already developed for this crop

Lang = language: E – English; S – Spanish; F – French; P – Portuguese

Crop strat. = crop strategy developed for this crop

Existing information on characterization and evaluation standards (including data derived from SGRP/GPG2, Generation Challenge Programme [GCP] and Harvest+ research, crop strategies and published papers) will be compiled for each crop and across crops when appropriate. Groups of crop experts, mainly drawn from the stakeholders involved in the development of crop strategies, will be convened to agree on the most strategic set of characterization and evaluation traits.

The tentative list of crops above will be further refined when the list of crop meetings are completed and all necessary information is compiled. The crops have been grouped as follows:

A. Short-term output: This group will be composed of crops for which standards have already been developed or crop experts have already been identified and can be easily consulted. These outputs can be reached in the first quarter of the implementation period and can be used for testing GRIN-Global and "ALIS". For example, rice should be an easy crop to complete as much work has already been done on traits of interest and a minimum list has already been developed.

B. Medium-term output: The development of data standards for these crops will require more time than the first group, but the basis for their development already exists. For example, descriptors may have been developed but have not been revised, or the crop is included in a GPG2 project activity.

C. Long-term output: The crops in this group lack developed standards or these are in an early draft. The development of data standards for these crops should be started as soon as possible and will last throughout the entire implementation period of Component 1.

The crops will be prioritized according to the current status of data-standard development and the procedures defined for each group; the three groups will be processed in parallel and the composition of each group may change according to progress made and identification of traits. The crops for which standards are most advanced will form the priority set for performing tests of GRIN-Global and "ALIS" while development efforts will be ongoing for crops for which the descriptor lists are less-advanced or lacking. The process of development of characterization and evaluation standards will be documented and the methodology published at the end of the project.

Groups of experts will be consulted on the need for and feasibility of developing data standards on specific multi-crop concepts including post-harvest traits, resistance to abiotic stress and nutritional value. Concepts that are common to all crops ('simplified ontology') will be discussed in the framework of developing the schema for a multi-crop information system. Although most of the consultations with crop experts on characterization and evaluation standards will take place through electronic means, where possible, face-to-face discussions will piggyback on meetings already planned and funded by others. The capacity of genebank staff to use the new data standards will be enhanced through the training activities carried out in the context of GRIN-Global deployment (Component 2).

The characterization and evaluation standards developed through this component will be validated by a larger group of stakeholders, including crop experts and germplasm users from the CGIAR, national agricultural research programmes, universities and the private sector, in order to assure the relevance and wide applicability of the standards. In developing these standards, the team will collaborate with GPG2 Activity 4.2.1: Review of characterization standards and strategies, in which a group is reviewing the crop-specific characterization standards previously developed by Bioversity for seven crops. The output of this GPG2 activity will be a strategy and methodologies for goal-oriented characterization to meet research needs; inputs will be received from this activity to assist in the development of the data standards. GPG2 Activity 3.3 on Crop registries is also designating minimum descriptors for identifying duplicates among holdings and could be an excellent source of information for Component 1.

Considering that descriptors cannot be translated until they are validated, translation will have to be completed beyond the lifetime of Component 1. However, some data standards are already available in several languages (see Table 1); these will be the priority selections for the testing of GRIN-Global. Once validated, the data standards will be made available in a suitable format for both "ALIS" and GRIN-Global, as well as disseminated via the Internet and through crop and regional networks.

Component 2: GRIN-Global deployment

The Trust is partnering with USDA in developing GRIN-Global, a state-of-the-art genebank information management system. This project will provide additional resources for the deployment of GRIN-Global.

Management

Bioversity will be closely involved in the USDA project through its representative on the Technical Steering Group (TSG) convened by USDA. According to the Terms of Reference, the functions of the TSG include: acting as a stakeholder representative on technical issues relative to the project's mission and recommending strategies to encourage its use; providing technical input into the GRIN-Global project; and providing technical guidance to the GRIN-Global development team.

Development

Bioversity staff, in collaboration with national partners, will actively participate in the testing of prototypes before release of the final version of GRIN-Global. Bioversity will contribute a comprehensive GRIN-Global data dictionary and user documentation using the data standards developed in Component 1 and in line with the requirements for linking to "ALIS". A global registry of international data standards will be developed by Bioversity and the necessary technology will be provided to link GRIN-Global to the system. A test group composed of curators from various countries and participants in crop expert groups (Component 1) will validate GRIN-Global functionalities and compliance with the data standards.

Outreach, help desk and capacity building

The project team will participate in the development of a help-desk web site in multiple languages, guidelines for data migration and other supporting documentation, in collaboration with USDA. Based on a list of criteria developed with the Trust and USDA, Bioversity will select 25 genebanks to be targeted for the adoption of GRIN-Global and capacity-building activities. Bioversity will also facilitate the adoption of the system through the efforts of its regional staff, who will build on regional networks as mechanisms to facilitate deployment, and will monitor progress with the help of the network coordinators. A baseline assessment of the 25 genebanks will be conducted as part of the genebank-selection process. Regional workshops will be held to train genebank staff in the system's operation and direct help-desk support will be provided to support use of the system. Training materials will be developed by USDA and translated in collaboration with Bioversity regional offices. The genebanks' information systems will be upgraded when appropriate in order to facilitate migration to GRIN-Global.

Quality data acquisition and checking

The data standards developed in Component 1 will be embedded into the GRIN-Global structure for data entry and will be promoted during training workshops as best practices in documenting holdings. The GPG2 knowledge base on genebank best practices (Activity 2.1.3) will promote the use of the data standards along with the methodology.

The reference documents promoted by GBIF on data quality will be recommended when appropriate: http://www.gbif.org/prog/digit/data_quality

Chapman, A. D. 2005. Principles and Methods of Data Cleaning – Primary Species and Species-Occurrence Data, version 1.0. Report for the Global Biodiversity Information Facility,Copenhagen

Chapman, A.D. and J. Wieczorek (eds). 2006. Guide to Best Practices for Georeferencing. Copenhagen: Global Biodiversity Information Facility, Copenhagen.

Chapman, A. D. (Ed.) 2005. Principles of Data Quality, version 1.0. Report for the Global Biodiversity   Information Facility, Copenhagen.

These data-quality checking procedures will also be promoted to the USDA team.

Component 3: "ALIS"[3]

Management

The Accession-level Information System ("ALIS") will be implemented under the guidance of an international steering committee (ISC) representing the Global Crop Diversity Trust, Bioversity, the Secretariat of the ITPGRFA, the CGIAR Centres and SGRP, ECPGR, USDA, the Global Biodiversity Information Facility (GBIF), the International Seed Federation, breeders and the International Union for the Protection of New Varieties of Plants (UPOV). An invitation will also be extended to member countries of ECPGR to nominate a joint representative to the ISC. The functionalities of "ALIS" will be based on advice of experts at an technical consultation held early on in the project, and will be subject to an external review by breeders and curators before the end of the project. A final evaluation of "ALIS" will also be undertaken, with support from the Trust, by an external group of technical experts nominated by the Trust, the Secretariat of the ITPGRFA and Bioversity. The evaluation report will be submitted to Trust and the "ALIS" ISC.

Development of the "ALIS" structure, data-exchange process and modules

"ALIS" will be designed to link the dispersed databases of national, regional and international genebanks, allowing access to – and searching of – standardized information about germplasm (passport, characterization and evaluation data). Priority will be given to linking the information systems (passport and characterization data) of the CGIAR centres' (via SINGER), the European network (via EURISCO), USDA and the 25 genebanks adopting GRIN-Global. "ALIS" will need to cope with a wide range of data providers at different technological levels. At project startup, a technical consultation will identify the most appropriate technologies for both data providers and end-users. Providers will be enabled to publish data on "ALIS" using a range of technical solutions that meet their individual technical requirements and capacities, from an easy upload function to more stable web services. Tools and standards used by GBIF and GCP Consortium members will be assessed for their stability and potential use by "ALIS" data providers. Through deployment of GRIN-Global, the technical capacities of many genebanks will be standardized and the technology provided for their inclusion in "ALIS". A mechanism will be developed to ensure that each dataset is properly attributed to its specific data provider and data source.

The Secretariat of the ITPGRFA will conduct a wide consultation with genebank managers regarding the functionalities and interoperability of the ordering toolkit (OTK) so that it can be customized and coupled with existing systems. The OTK will make it possible for users to order selected germplasm and receive the appropriate Material Transfer Agreements (MTA) (including the SMTA for Annex 1 material); the ordering system will be linked to the PID Server and 'data store' under development by the ITPGRFA Secretariat. The data store will enable the ITPGRFA Governing Body – and only this entity – to perform summary data analysis. "ALIS" will thus allow plant breeders and other users to search through genebank holdings worldwide for particular traits, select the accessions they need, order this germplasm and generate and register MTAs, through a single portal. The various elements in this process have been provisionally called 'the ITPGRFA module' until an appropriate name is selected.

Following development of the "ALIS" schema and concept name server for passport and other critical data (such as characterization and evaluation data), a core database structure and data replication or harvesting system will be constructed. SINGER, EURISCO and USDA data will be replicated into the central data structure as the first contributions to "ALIS"; the data managed by the 25 GRIN-Global rollout genebanks will follow. As the respective collective databases of the CGIAR genebank curators and European genebank curators, SINGER and EURISCO will be linked to crop registries storing harmonized characterization data sets from various genebanks.

---

[3] The acronym "ALIS" is a temporary name that is likely to be reformulated during the project. For this reason, it appears in quotation marks throughout the document.

The visibility of these two major networks of data sources – each representing a well-established community of practice – will be maintained within "ALIS".

Information in GRIN-Global and other germplasm-documentation systems will be mapped to "ALIS" with the most appropriate interoperability technology available and using Biodiversity Information Standards (TDWG). The crop descriptors database developed through SGRP/SINGER will be updated by the SINGER team using the agreed minimum list (Component 1). A data-exchange schema will be designed by the "ALIS" development team (cf. Section E, Component 3 – Collaboration) to include these minimum data standards. The development of a multi-crop platform like "ALIS" necessitates the identification of multi-crop concepts (ontology) to facilitate cross queries and information retrieval – this will be implemented in collaboration with the Component 1 team.

Development of "ALIS" web-based user interface

This powerful and user-friendly search, retrieval and ordering system will be built in consultation with curators and breeders under the guidance of an international steering committee (ISC). The project team will build the "ALIS" ordering system to be compliant with the ITPGRFA requirements. A prototype 'shopping-cart' function has already been added to SINGER and will be submitted to users for evaluation in the framework of the GPG2 project at the request of CGIAR centers. The CGIAR genebanks holding germplasm in trust under the ITPGRFA are the priority collections for implementing an ITPGRFA-compliant ordering system. This ordering system and related guidelines will then be promoted by the CGIAR centres through GPG2 Activity 2.1.3 on Best practices in genebank management.

Outreach and capacity building for data providers

Help-desk support will be provided along with user documentation and training workshops for genebanks adopting "ALIS" and the ITPGRFA ordering tool kit. Online tutorials and outreach material will be developed and translated by Bioversity's regional offices and regional plant genetic resource networks in collaboration with the Secretariat of the ITPGRFA.

Quality data acquisition and checking

GPG2 activities will provide the mechanism for assessing data quality from the CGIAR community and its direct partners. Existing quality-checking tools, including GCP tools and Bioversity-promoted tools, will be included in the data flow or will be available online for data providers. Reference documents promoted by GBIF on data quality will be recommended when appropriate: http://www.gbif.org/prog/digit/data_quality:

The data quality checking procedures will also be promoted through GRIN-Global.

4. Expected outputs

Through consultations with crop expert groups, a set of agreed characterization and evaluation data standards will be produced for 22 crops, with a focus on traits and characteristics of high priority to users. In addition to publication and dissemination via the Internet and through regional and crop networks, these standards will be integrated into GRIN-Global.

GRIN-Global is being developed by USDA through a complementary project. During the development and rollout of GRIN-Global, a help desk and deployment team will be established through this project to facilitate the adoption of GRIN-Global in at least 25 genebanks worldwide. Support and training of genebank staff will be provided throughout this process by the GIG project team.

Information on the holdings of at least 65 genebanks worldwide will be harvested and shared through "ALIS", a powerful and user-friendly online search, retrieval and ITPGRFA-compliant

ordering system. Technological standards, tools and protocols will be developed to enable genebanks – including the 25 genebanks included in the initial GRIN-Global rollout – and existing information systems like SINGER and EURISCO to share accession-level information with "ALIS". Help-desk and training support to genebank data providers will also be provided, and project experts will visit critical locations to ensure that genebank staff have the capacity for sharing data and using the ITPGRFA ordering system.

5. Expected outcomes

It is expected that "ALIS" will provide access at least 2.5 million records of accessions from SINGER, EURISCO, the GRIN community and other key partners. With "ALIS", breeders will be able to cross-search for information on morphological and other characteristics relevant for breeding for at least ten crops. For material in the MLS, users will be able to search for records on accessions held worldwide to locate germplasm of interest, and exchanges of germplasm will be recorded by the ITPGRFA Secretariat.

Within four years of project completion, "ALIS" is expected to record a 300% annual increase in searches over the current CGIAR system (SINGER) and a 400% increase in downloads of accession information over the baseline measured at project startup. The information systems that contribute to "ALIS", including SINGER, EURISCO and GRIN-Global, will continue as independent gateways; the partnerships supporting these systems will ensure the sustainability of "ALIS". Bioversity is committed to the further development of "ALIS" into a truly global system.

6. Target beneficiaries and impact

This initiative represents a direct and tangible response to the needs of germplasm managers and users as expressed in the Global Plan of Action and crop and regional strategies. Breeders and other users will be able to search across genebank collections worldwide with "ALIS" for the germplasm they need for crop improvement and research.

In addition, genebank curators and breeders will be able to better document germplasm by applying agreed characterization and evaluation standards to capture data in the field, and the information they collect will be managed through an easy-to-use genebank data-management system. The capacity of genebank curators to share their information will be reinforced through the use of the new standards and information system. Genebanks' capacities in informatics will also be strengthened.

7. Assumptions

For the effective implementation of GIG, crop experts and other stakeholders must be available and actively participate in the proposed crop consultation groups, and agree on common standards for each of the 22 target crops. Genebank curators must also apply the agreed standards for characterization and evaluation.

It is assumed that GRIN-Global will receive long-term support for assisting genebanks with deployment, including a helpdesk, user support and technical training. It is further assumed that the 25 genebanks and regional networks selected for initial rollout will provide the institutional support necessary to ensure the migration of their data.

Collaborating institutions must be willing to share their data through "ALIS" as global public goods. Web services technology (including BioCASE and TAPIR) is assumed to provide the long-term approach to data sharing for "ALIS". It is assumed that "ALIS" and the ITPGFRA module can be

developed for maximum interoperability and that "ALIS" will become the appropriate data source. Finally, plant breeders and other germplasm users must be made aware of the system.

8. Consequences of not funding

Lack of funding for this work would result in a continuation and exacerbation of the current situation in which the potential of crop germplasm conserved *ex situ* is not effectively used in breeding programmes because of limited and fragmented access to needed information.

9. Additional information and comments

Collaboration with SGRP/GPG2 activities

The collaboration between GIG and SGRP will build on a series of activities already being undertaken through the GPG2 project:

- **[2.1.3] Readily accessible knowledge base on best practices in genebank management [Ehsan Dulloo]** – A database and online training tools will be developed to support capacity development in CGIAR and NARS genebanks. A group of NARS scientists (two in each region) will be trained as trainers for further dissemination of best practices.
- **[3.2] One-stop entry point for information on and access to the in-trust collections [Rajesh Sood]** – SINGER is being upgraded and request procedures are being harmonized into a system-wide common protocol, incorporating the SMTA and an order-tracking system. This activity will be closely linked to the work of GIG and SINGER, and all its functionalities, will be accessible through "ALIS". The workplan for the integration of SINGER into "ALIS" will be aligned with the GPG2 SINGER workplan to the greatest extent possible.
- **[3.3] Crop registries [Jan Konopka]** – This activity will support the design of global registries for target crops in order to catalogue accessions held in common among CGIAR centres and other large collections, providing central access to distributed data sources through the global information system. Through this activity and a complementary GCP project, passport descriptors are being revised. These descriptors will be validated and integrated into the global information system being developed through GIG.
- **[4.1.2] Location data quality [Robert Hijmans]** – All spatial data on samples held by CGIAR genebanks will be corrected and validated using the latest GIS tools, concepts and digital gazetteers. These will include automatic geo-referencing, estimation of missing data and validation of existing data for inclusion in the global system.
- **[4.1.4] Identification of eco-geographic gaps in the in-trust collections [Andrew Jarvis]** – Strategies and procedures are being developed for analyzing genetic diversity and gaps in collections with regard to genetic, morphological, agronomic and eco-geographic/passport data. The information on diversity yielded through this activity will be accessible through "ALIS".
- **[4.2.1] Review of characterization standards and strategies [Hari Upadhyaya]** – A task group will review the crop-specific characterization standards developed by Bioversity for seven crops. A revised strategy will be developed for more efficient and effective characterization, based on combining the optimum choice of descriptors with the optimum choice of accessions. The output will be a strategy, standards and methodologies for goal-oriented characterization, which may contribute to Component 1.

The work of GPG2 and other ongoing initiatives, particularly the USDA-led project to develop GRIN-Global, will be closely linked to the work of GIG to ensure common methodologies and standardization of outputs. GPG2 activities will help to ensure the provision of quality data to "ALIS" through SINGER.

Collaboration with ECPGR/EURISCO activities

Following consultation with the ECPGR Documentation Network and Steering Committee, GIG and ECPGR will build on activities being undertaken for the development of EURISCO, including the following:
- **National Inventories** – Strengthening and developing European National Inventories of genebank holdings
- **Uploading system** – Further improvement of the automated uploading system of National Inventory data into EURISCO
- **Location data quality** – Reporting and estimation of missing spatial data regarding samples in EURISCO

- **Taxonomy data quality** – Reports related to taxonomic data in EURISCO, highlighting suggested corrections
- **One-stop entry point for information on and access to the European *ex situ* National Inventories** – ECPGR is discussing the use of EURISCO for the implementation of a registry of MLS material, for SMTA reporting procedures and for the establishment of a registry of material in an ongoing project called 'A European Genebank Integrated System for Plant Genetic Resources for Food and Agriculture' (AEGIS). The possibility of installing a germplasm ordering system can be proposed to the community.

All these activities will be closely linked to the objectives of GIG; it will be proposed that all the functionalities of EURISCO be accessible through "ALIS".

Strategic collaboration

Strategic collaboration with other institutions and activities will maximize the impacts of GIG by aligning it with similar initiatives and mobilizing additional resources; important strategic linkages include those with SINGER, EURISCO and GRIN. Others could include sponsorship by the Secretariat of the ITPGRFA of outreach activities for participants in "ALIS" and OTK training workshops. In addition, projects being undertaken by other CGIAR centres outside the context of GPG2, including the GCP projects 'Implementation of web services technology in the GCP Consortium' and 'Management of the GCP Central Registry and the creation and maintenance of templates for data storage in repositories', will provide data sets that are important to GIG. CGIAR centres are currently undertaking evaluation projects yielding valuable data sets (e.g. on breeding lines). An assessment will be undertaken of concurrent work in this area and collaboration will be initiated with these centres (see Table 2).

The Global Crop Diversity Trust is providing support for upgrading collections, and in some cases for long-term conservation of collections. The GIG team will consult and collaborate with the Trust in order to make the most of this support and meet the project's objectives.

A strategic collaboration with GBIF/TDWG will be discussed during the IT consultation for developing web service technology and outreach activities, following approval by the ISC. This proposal was developed with the idea that "ALIS" would become a GBIF-type platform specialized in genebank holdings. GBIF (http://www.gbif.org/) primarily focuses on herbarium and botanical garden holdings. However, GBIF is opening its scope to a wider range of data sets and data providers, and supporting the development of TDWG standards. GBIF gathers data through the use of web services, a challenging, advanced technology. The GBIF network relies on active national nodes that it supports financially and technically. GBIF web services are being improved for the indexing of larger data sets but they are still not a fully stable technology. In collaboration with TDWG, GBIF has built a large community of practice for exchanging technical developments and elaborating data-standards; passport data from USDA, SINGER and EURISCO are published through GBIF. However, data providers do not yet provide characterization and evaluation data sets to these platforms. This is one of the objective of "ALIS" within the next three years; exchange schemas need to be developed for this purpose.

GBIF is not well represented in Africa and Asia, and collaboration with CGIAR centres through SGRP would be of interest to them for that reason. The EURISCO upload system is among the alternative solutions being considered by GBIF national nodes for low-technology data providers.

Table 2. Other strategic partnerships

	Trust	Treaty Secretariat	Bioversity	SGRP/GPG2	USDA GRIN-Global	CGIAR projects outside SGRP	ECPGR
Provides funds	Cash	Cash/ In kind	in kind	in kind	-	-	-
Provides Project modules		- PID server - Data store -OTK	Deployment of GRIN-Global to national genebanks	- Crop registries - Char. standards - OTK	-	-	-
Provides expertise to other modules	-	-Ordering systems	Input to TSC of GRIN-Global project	- SINGER - Technical Guidance to Treaty	-	-	-
Inter-dependence with other projects	Regeneration project	Materials, tools and PA to encourage inclusion and sharing of materials under the Treaty	-	See previous page on GPG2 activities	Bioversity leads deployment	CGIAR Centre-based projects, e.g. GCP projects and breeding projects	EURISCO