Report of the SINGER workshop for the Genebank Database Managers 8-10 December 2010

Bioversity International, Maccarese, Italy

SINGER Knowledge makes the difference

SGRP

Content

Summary of Recommendations

Recommendation 1 – Crop Registers and cross referencing

Cross referencing is not simple, but important and the work on registries, focusing on quality of information and value-adding to collections, creates a body of knowledge which is validated by crop experts. The identification of a duplication may be reflected by the institution holding the accession with the addition of the replicated accessions IDs in the field 'other number (OTHENUMB)') to bring additional information of interest for GENESYS re. cross links between Passport data. The results of the cross referencing could be documented with metadata about who made the cross-referencing, how the result was obtained, who validated the result, when it was made etc. The links from the Crop Registers to the collecting mission data should be established. Descriptors have already been included in the Crop Registry Template (CRT) for creating the links. The information on collecting mission is actually necessary to cross-reference accessions which were collected in the field (during missions).

Recommendation 2 – Collected sample data

The collected samples database and the repository are seen as great resources to be used to complete the quality reports on the basis of the original information. At the system-wide level, the work done on collected samples and their linkage to accessions is partial. SINGER provides the links between the Centres' accessions and the collected samples but there are still a fair amount of samples that are not yet linked.

The group recommends that this work be continued recognizing that it requires a large budget. This is a matter of urgency and Centres should all contribute. The way centers should contribute needs to be discussed by the Inter Center Working Group on Genetic Resources (ICWG-GR)

Recommendation 3 – Revision of the MCPD

At the crop databases level, the MCPD is too limited but MCPD is still valid as a global exchange format because it receives a broad consensus. However, MCPD needs to be revised to include extensions made by other communities and then facilitate the automation at the global level (e.g. EURISCO, Crop Registers. Metadata on passport data need to be added, e.g. who, when, maximum possible completion level. The most important is to have a single identified and documented schema for the MCPD. The appropriate use of the MCPD is based on the proper interpretation of the fields it contains. There is a need to define compulsory fields and optional fields. . The MCPD is online since 2001 as a static reference document and should enable online comments and feedback. A flexible system enabling online discussion is required. The use of a wiki for example will provide ready access to the standards and link explanations. A flexible system enabling online discussion is required.

Recommendation 4 – Facilitate the application of the MCPD

Best practices or guidelines on the way to fill in the MCPD and map the crop data to its format are presently missing and will be necessary to add within the template in order to enhance the data quality, e.g. where data is not available, indicate n/a for descriptors not applicable to an accession instead of leaving a blank field. An annotation tool to add descriptors and metadata will be a useful additional feature.

Recommendation 5 - **Characterization and Evaluation data management at the level of GENESYS**

A drastic revolutionary rethink on how to handle this data, particularly the evaluation data, is required, as it seems that GENESYS will be facing a never ending process which eventually will become unmanageable.

Recommendation 6 - Collating germplasm transfer data on a yearly basis

The Governing Body of the Treaty does not meet every year and will need the system-wide report every two years. However, it is important to compile the data on a yearly basis and fit the report within the calendar year. The delivery of data through SINGER will be ongoing. A template for statistics on breeder's distribution will be developed by Bioversity as opposed to the accession-level template for genebank data. The template will be submitted to the approval of Breeders. However, a long term decision needs to be made by the ICWG-GR about SINGER data sharing process handling or not all breeders' distribution data.

Recommendation 7 - **Awareness of top level management to obtain an institutional commitment to the reporting on genebank and non-genebank material**

Before March 2011, the Inter-Centre Working Group-Genetic Resources (**ICWG-GR)** must inform the Centers' Directors and Directors of Research about the importance of this yearly based system-wide report to the Governing Body and ask their support in obtaining the units organized and support the data collation. If necessary, the Consortium Board can also be alerted to this need.

Recommendation 8 - **Develop a joint proposal with Centres adopting GRIN-Global for a system-wide hands-on workshop for evaluating the data migration possibilities and efforts**

The group suggested that a CGIAR user or open source community for GRIN-Global should be set up with participation of managers of diverse data systems in CGIAR to guide/steer adoption of G-G, sharing resources, etc. Once the final first version of GRIN-Global will be release by USDA, the first step might be a system-wide hands-on workshop for evaluating the data exchange and migration possibilities.

Recommendation 9 - An Online open space for group discussion standards is required

The group needs to have an open space where it can discuss the standards like the SINGER data warehouse dictionary that include the MCPD. Milko Skofic and Luca Matteis (Bioversity) will look at the potential of Google Apps and CGXchange for publishing and commenting the dictionary.

Recommendation 10: Visibility of SINGER and international collections in GENESYS

It was recognized that users need a single door which reveals all the answers; from where they can access everything they need the required information in a consistent manner. SINGER is a network and a community with particular practices that has a model role to play within GENESYS. The group recommends that, in GENESYS, international collections are easily identifiable by the users. The definition and a model of a global system as a single portal composed by several windows need to be clarified and developed. Therefore a decision on whether to keep or abandon an identification some of the SINGER services within GENESYS (e.g. distribution data, collecting missions) will be adequately made by the ICWG-GR and SINGER users.

Recommendation 11 - **Elements for the integration of SINGER into GENESYS**
The group recommends that there is **one common middleware, a single data storage and one upload mechanism**. The upload system should accommodate PaD, characterization (Field/Molecular), evaluation, distribution data. Centres will upload MCPD extended data plus distribution into the middle tier. This will improve the quality of data and data documentation whatever solution regarding the portal is adopted. SINGER and GENESYS must share the same rules for online publishing of the data for users, providing the same quantity and quality of data on both.

DAY 1 – Summary of the present system-wide situation and lessons learned, identification of elements needed

Participants were welcomed by David Williams, SGRP Coordinator, who recalled the importance of the objectives of such a workshop. Performing data sharing in SINGER and therefore GENESYS goes through an agreed and applied exchange mechanism. He encouraged the group to agree on the best way forward to find solutions that will enable the publishing of quality data for the international collections. This workshop was called by the SINGER Task force as indicated in the recommendation 6 of the report: **'***There is a need for Bioversity to promote only one system and to provide SINGER with a system like that of EURISCO that produces quality reports. No concrete decision was made in this regard and it was recommended that the Task Force discuss this issue in a separate meeting with the genebanks' database managers'***.** – (see report available at https://sites.google.com/a/cgxchange.org/genesys/singer-task-force-report-2010)

Session 1- Current status of Passport Data, Georeferences, Collecting missions, Characterization and Evaluation data in the CGIAR Centres' information systems

Overview of available data sources, data types and data sharing technologies available or possible in the centres, issues and constraints for sustainably providing data sets to SINGER and GENESYS

AfricaRice, CIAT, CIP, CIMMYT, ICARDA, ICRAF, IITA, ILRI, IRRI, Bioversity

The genebank database managers gave an overview of the situation in their respective Centres, making it evident that there is a large diversity in terms of data sources, databases and the organization of genebanks and breeders' data. Most of the database managers and curators are multitasking for the data management. In most Centres, the breeding data are maintained in International Crop Information System (ICIS) or an ICIS-like system. It was underlined that Breeders' data are more complex than genebank data and do not fall under SINGER coverage (see the Overview of the Centres' presentations in Annex 1). SINGER cannot accommodate the needs of ICRAF network which is based on famers' evaluation in eco-geographic regions.

The data exchange difficulties within SINGER were highlighted, e.g. the lack of an automatic upload system that rationalizes the data flow, that would allow SINGER to immediately reflect the latest data sent by the Centres and lack of a system to update only new accessions. The use of SINGER standard fields posed some issues such as the incompatibility of categories with the crop databases, lack of flexibility of the descriptors' lists and use of FAO codes. Suggestions were shared regarding the addition of versioning on the data sets, data citation

Session 2- Current system-wide actions: SINGER, GENESYS, Crop Registers, Collected samples database

Current SINGER status in terms of content and web site

The 2010 update of SINGER brought the number of passport data to **746 611**, meaning **an increase of 49 049 accessions**. Collecting missions and germplasm transfers have also been uploaded. However, collections have been merged at the request of the Centres bringing the number of collections from 77 down to 47. All passport data received have been checked and reformatted through collaboration between the SINGER team and the Centres' database managers. The formatted passport data have been provided to GENESYS.

SINGER usage:

- An average of 850 unique visitors consult the site per month;
- 153 requests of germplasm were sent to all the Centres without particular awareness on the existence of the online request gateway
- Requests for system-wide data on the collecting missions, germplasm transfers.

Latest developments:

- Update and redevelopment of the collecting missions database within SINGER and links with the genebanks' accessions;
- The germplasm request gateway is now linked to the official Treaty Registration service that provides users with a permanent identifier.

The current issues on SINGER are:

- Still no upload system including elements like a central registry, quality checking routines and logs;
- Errors are generated along the workflow from the data source to SINGER;
- No versioning of the database;
- No systematic central curation;
- No differential upload, therefore the need to wait for the latest Centre to upload data before refreshing SINGER content;
- Metadata are missing.

The quality of the passport data is crucial in the system as it is pivotal information for field and molecular Characterization and Evaluation (C&E) data, pedigree, for collecting missions and collected samples, etc.

Current GENESYS Status in terms of content and website

Statistics on the last GENESYS updates were given by Fawzy Nawar (see Table 1 below). For the upload of the C&E data, the list of metadata has been provided to the Centres, however this information does not largely exist for legacy data. IITA, CIMMYT and ICARDA mentioned that the data was easy to extract. Data sets were provided to Bioversity and centrally uploaded (see breakdown in Annex 2).

Table 1: Statistics on GENESYS content

Number of accessions	2 333 733
Georeferenced accessions	625 290
Genebanks	356
Accessions with C&E data	3 774 95
Observations	11 337 907
Observations from GRIN	3 340 127
Observations from SINGER	7 997 780

Crop Registers' role as partners for quality data and pedigree information for SINGER and GENESYS

Crop Registers developed during the Global Public Programme phase II use the revised and extended Multi-Crop Passport Descriptors (MCPD) while GENESYS uses the classical MCPD that contains less information. So, Crop registers bring additional important information enabling the identification of duplicates across genebanks and cluster accessions. Crop registers have developed tools to identify the duplicates and create the cross-reference between the passport data of accessions held by different genebanks. The SINGER Task Force recommended that these

tools be made available on the Crop Genebank Knowledge Base website with the indication that the workflow includes a validation by the crop experts.

Recommendation 1 – Crop Registers and cross referencing
Cross referencing is not simple, but important and the work on registries, focusing on quality of information and value-adding to collections, creates a body of knowledge which is validated by crop experts. The identification of a duplication may be reflected by the institution holding the accession with the addition of the replicated accessions IDs in the field 'other number (OTHENUMB)') to bring additional information of interest for GENESYS re. cross links between Passport data. The results of the cross referencing could be documented with metadata about who made the cross-referencing, how the result was obtained, who validated the result, when it was made etc. The links from the Crop Registers to the collecting mission data should be established. Descriptors have already been included in the Crop Registry Template (CRT) for creating the links. The information on collecting mission is actually necessary to cross-reference accessions which were collected in the field (during missions).

Session 3- Improvement of the quality of the passport data and use of the collected sample database

Use of the collected sample database for improving quality of passport and pedigree data

The objective of the Global Public Goods Project Phase 2 (GPG2) activity was to scan and complete data for the collected samples in the collecting missions' database and in the Centres' databases. The work presented during the workshop only referred to 'IBPGR/IPGRI supported collecting missions' in which the Centres were partners. The next step is to turn the collected sample database into a reference resource for the genebanks that wish to complete their passport data and also a product to which Centres can contribute to complete the content.

Table 2: Data Extraction from the IBPGR/IPGRI collecting missions

Collecting Missions	440
Passport Data Records	132 027
Passport Extracted (%)	90%
Samples linked to Accessions	65 000
Georeferenced samples	80 000
Samples with locality	50 000
Samples linked to SINGER	31 825
Samples linked to GRIN	16 152
Samples linked to EURISCO	12 183

There is also an overlap between reports held by Bioversity on the IPGRI/IBPGR missions with the Centres. The repository of pdf files can help Centres to identifying this overlap or provide missing information.

As the work was not completed under GPG2, the database in SINGER only covers a portion of data held by Centres. The most important action is to continue adding the missing links to the accessions and Centres can contribute to the work already initiated by Bioversity.

The check of georeferences for these samples was done but original data cannot be changed. It was recommended to provide feedback mechanisms on the passport quality to avoid issues between Centres' data and the collected samples database.

IRRI verified their Rice data using the scanned Rice mission reports and now the original data can be accessed online in full text. IRRI is experimenting with the Bioversity International web unit the process of adding the url of pdf files in the accessions' passport data of IRIS (International Rice information System). From the passport data, users will be able to open the full text stored on the repository of pdf files. AfricaRice would like to be able to install similar links to their crop database and need to cross-reference with the accession ID. CIAT and ICRISAT have also performed the scanning of their reports and figures are available in the GPG2 project reports. CIAT has published on its genebank web site all the pdf of the collecting mission reports and each pdf is linked back to Pasport data.

The collected sample database will be published online in early 2011 within SINGER.

Recommendation 2 – Collected sample data

The collected samples database and the repository are seen as great resources to be used to complete the quality reports on the basis of the original information. At the system-wide level, the work done on collected samples and their linkage to accessions is partial. SINGER provides the links between the Centres' accessions and the collected samples but there are still a fair amount of samples that are not yet linked.

The group recommends that this work be continued recognizing that it requires a large budget. This is a matter of urgency and Centres should all contribute. The way centers should contribute needs to be discussed by the Inter Center Working Group on Genetic Resources (ICWG-GR)

Debriefing and lessons learned from the last upload of passport data, georeferences and collecting missions data

Milko Skofic presented the results of the 2010 data upload that populated SINGER with more accessions. There are fewer georeferences than before (see table below) and in some cases, accession identifiers have changed which prevent rebuilding the accession-level links with the distribution data and collected samples. A sustainable application of basic principles is therefore necessary. Data sets were received through an upload on the Webdav server (Web-based Distributed Authoring and Versioning) or through email. Milko Skofic provided a feedback report to each Centre but Centres did not all act on their data after receiving the report.

Table 3 : Statistics on the 2010 SINGER upload

Institute	*Old*	*New*	*Difference*	*% change*
BIOVERSITY	1208	1284	76	6
CIAT	72 254	64 721	(7533)	(10)
CIMMYT	120 527	164 326	43 799	36
CIP	15 092	16 061	969	6
ICARDA	140 189	134 741	(5448)	(4)
ICRAF	1785	2005	220	12
ICRISAT	114 865	119 613	4748	4
IITA	27 596	27 280	(316)	(1)
ILRI	20 177	20 229	52	0
IRRI	108 272	117 417	9145	8
WARDA	21 752	26 098	4346	20
Total	696 562	746 620	50 058	7

One common template and a stable upload mechanism are basic elements to achieve an optimal data exchange system.

- The fundamental question is whether we need to change the MCPD standard and go beyond?
- Can we identify what is missing in the MCPD today that would facilitate information exchange if added?

A revised MCPD as the data standard for passport data

SINGER and Crop Registers use an extended MCPD, while GENESYS uses the original MCPD. Collecting missions and distribution data are additional to the MCPD and could be handled separately and be considered as a specific SINGER service. However, the group indicated that a single portal should publish all the data that the Centres send. If GENESYS only uses the MCPD, then where should the additional data be accommodated/located?

The real issues data providers are facing is the coding and mapping of their data to the standard formats. The MCPD needs to take into consideration evolving changes in taxonomy, country, administrative regions etc – How this should be tracked must also be defined.

The MCPD was developed ten years ago to facilitate the exchange of core information and was identified by FAO as a crucial element which should fit almost all crops. However, the current criticism is that the MCPD is limited and eliminates important specific data that genebanks manage and would like to see online. We need to decide what will be the core MCPD and what will be the additional data, as well as what metadata to attach. This will encourage Centres to use the MCPD template. Modification of the current data standard will not affect the Centres' systems and it will just mean to apply the modified standard as the agreed data exchange schema.

Bioversity announced that the revision of the MCPD started with FAO, with the International Treaty for Food and Agriculture (Treaty) to evaluate its current validity and eventually to accommodate emerging issues which may be of relevance (e.g. inclusion into the International Treaty's Multi-Lateral System. MCPD revision was initiated by FAO and Bioversity – the SINGER partners are invited to provide feedback on the MCPD through the survey that will be launched in 2011.

The most important is to have a single identified and documented schema for the MCPD. The appropriate use of the MCPD is based on the proper interpretation of the fields it contains. There is a need to define compulsory fields and optional fields.

The use of a wiki will provide ready access to the standards and link explanations. A flexible system enabling online discussion is required. An annotation tool to add metadata would be a useful addition/enhancement.

Georeferences

Georeferences are part of the MCPD. Precision and accuracy for latitude and longitude are included in the Crop Registry template. Collected sample data bring quality to the georeferences. A decision needs to be made on which precision data is needed and realistic.

Passport data feedback process for curation

The last upload of the Centers' information demonstrated that data were often not formatted according the MCPD. Descriptors and ranges of values used at the Centres-level are not the same as the global descriptors so flexibility, which is in contrast to the standardization concept, is necessary and a new descriptor submission mechanism is needed.

When Centres' values do not match SINGER format, should the standards be enforced or remain flexible? How to balance the curation effort made at the crop database level and the Central level is an issue to be decided. Centres maintain responsibility for data quality. A feedback process between the Central level and crop database managers is needed, providing detailed reports on accepted and rejected data, with the reason for their inclusion or rejection. However, half empty MCPDs are due to lack of data and metadata in the legacy data sets. Database managers can publish only what genebank curators provide and, in the legacy data, part of the missing information can no longer be recovered. The level of completeness of the Passport Data should be part of the data quality check and we need to develop guidelines for best practices on filling in the extended MCPD. Metadata need to be added indicating the maximum level of MCPD completion for one accession. The Generation Challenge Programme (GCP) quality passport guidelines should be promoted.

Passport data upload process with quality check

With regard to direct upload of passport data, it would be advisable to create an intermediary step/section, called the 'data purgatory' during the meeting, where data quality can be checked or curated before going public. The date of the data upload should appear.

Now we must work on an automated system with automatic reporting to confirm the data received, identify quality problems, correct them and then data goes directly to GENESYS. The reason why some data cannot be validated would appear as a feedback to providers. Should this automated system be a website where data can be uploaded and feedback received?

Passport data upload frequency

Centres prefer to have 'on demand' updates through an automatic upload process. It would be up to the data provider to upload on a voluntary basis. If no voluntary upload is performed, then an annual deadline must be set. It should be noted that Centers want to see their validated update appearing on line immediately.

Recommendation 3 – Revision of the MCPD
At the crop databases level, the MCPD is too limited but MCPD is still valid as a global exchange format because it receives a broad consensus. However, MCPD needs to be revised to include extensions made by other communities and then facilitate the automation at the global level (e.g. EURISCO, Crop Registers. Metadata on passport data need to be added, e.g. who, when, maximum possible completion level. The most important is to have a single identified and documented schema for the MCPD. The appropriate use of the MCPD is based on the proper interpretation of the fields it contains. There is a need to define compulsory fields and optional fields. The MCPD is online since 2001 as a static reference document and should enable online comments and feedback. A flexible system enabling online discussion is required. The use of a wiki for example will provide ready access to the standards and link explanations. A flexible system enabling online discussion is required.

Recommendation 4 – Facilitate the application of the MCPD
Best practices or guidelines on the way to fill in the MCPD and map the crop data to its format are presently missing and will be necessary to add within the template in order to enhance the data quality, e.g. where data is not available, indicate n/a for descriptors not applicable to an accession instead of leaving a blank field. An annotation tool to add descriptors and metadata will be a useful additional feature.

Action 1 – Templates and standards will be made available on a wiki to enable comments

Session 4 - The germplasm request gateway on SINGER

Presentation of the workflow and discussion on additional features that genebank curators may wish to add

The SINGER germplasm request process now integrates the registration form on the Treaty website through a link with the official Permanent IDentifier (PID) server located in the United Nations International Computing Center (UNICC, Geneva, Switzerland). Once the user is registered, he/she can perform a request that will send a mail with the requester contact details to the genebanks holding the seeds that were selected and a copy of the email is sent to the requester. The order will be processed offline by the genebank as signature of the Standard Material Transfer Agreement (SMTA) is necessary and because SINGER has no legal status to enable SMTA signature.

The registration page is on the PID server so any site that needs to access the Treaty registration form must be linked to the PID server. The Treaty secretariat provides the code for online ordering systems to be connected.

Questions raised:

- How to share the list of already registered collaborators?
- PID will be tied to an institution with its FAO code or an individual. Will it be possible for Centres to request the type of organization type if they send the PID number to the Treaty?
- The email sent by the request system is convenient for the CGIAR Centres but could it be replaced by another mechanism?

The germplasm requests issued through SINGER need to go into the individual Centre's workflow that locally processes all requests. Email is the most used system but does not sound a fully reliable mechanism so IRRI, ICRAF, CIP have a procedure to check the requests that may have fallen through the cracks. CIAT also has a specific tracking system.

Some Centres would like to have a web service to directly consult the requests posted through SINGER/GENESYS and store information after the requests are processed to keep them on record. In this case, should the requests be accessible at the genebank level or at the crop collection level? Could a simple summary of requests be sent to the Centres once or twice a month?

The requests that are on SINGER are only requests and there is no way to keep track of whether the order was processed or not. SINGER has no legal status with regard to the SMTA and information related to a germplasm transaction and storing the full information on the site, accessible through a password, may pose problems. It seems that even the information stored as part of the request can bear a confidentiality issue and this needs to be discussed with the Treaty secretariat representatives and its legal focus group. The reporting to the Treaty follows a different process and occurs between Centres and the Treaty after the order has been processed, the SMTA signed and the seeds sent.

The SINGER germplasm request system is regarded as an interim system, a proof of concept for GENESYS. However, the process selected for the CGIAR Centres might not work for external partners. At the global level of GENESYS, the email system will require having a contact email address for all genebank disseminating germplasm and keeping track of the changes. A more efficient system than email will be needed i.e. an ordering toolkit.

Action 2 – The germplasm request gateway - The list of accessions will be attached to the request email as an Excel file in order to be easily processed by the genebank curators.

Action 3 – Germplasm request gateway - The registration form developed by the Treaty does not include the type of cooperator as per the MCPD and it makes it difficult to compile the information per category for the reporting to the Treaty. Bioversity will contact the Treaty Secretariat to suggest adding institution type.

DAY 2 – Exchange of quality data

Session 5-Characterization and Evaluation data

The distinction between the different types of data that fall under the 'C&E' (Characterization and Evaluation) definition needs to be made as it entails various measurement methods, data sources, data types and data capture and exchange processes. Characterization of germplasm can be performed on the genebank collection, usually on several plants representing the accession and at the maximum expression of the trait, and also on breeders' trials, carried out on several sites mostly under stress conditions with several seasonal iterations. Characterization can also be based on molecular marker analysis. The GENESYS discussion focused on field characterization and evaluation data, excluding the molecular characterization results.

Debriefing and lessons learned from the last upload of Characterization and Evaluation data

GENESYS has a structure that manages each crop in a separate file, divided by traits and then trait experiment. A total of 1650 experiments were included in GENESYS. Data from the The Germplasm Resources Information Network (GRIN) were easy to obtain as it is freely downloadable. For any given trait, there are differences of data and methods between institutes, and within institutes for the different experiments. The lesson learned is that C&E legacy data must be handled in a particular way as it includes large data sets and there is no standard metadata attached. Genebanks need to obtain experiment metadata from the source which is often not easy.

Identification of the basic elements for quality Characterization and Evaluation data sharing and regular upload[1]

Template and standards for characterization and evaluation data – see table 4 below

The standard to describe the Characterization and Evaluation data at the global level is emerging. The present characterization and evaluation descriptors need refinement and will be evolving with the upload of additional data sets. Trait metadata is associated with trait while other metadata relates to the experiment. All experiments which have the same experiment title were part of the same trial. Standard practices would be to have calendar data, field data, soil data etc. as the proper documentation at the experiment level. **It was suggested to consider adding the name or reference person who took the data as users could reference it with a scientific publication**

The Characterization and Evaluation metadata template will have to be on the wiki for comments.

[1] Recommended by R. Simon for further reading: Jeffrey W. White and Frits K. van Evert. 2008. Publishing Agronomic Data. Agronomy Journal Volume 100, Issue 5.

Table 4: proposed metadata on the characteristics to be included in the Global Portal (Michael Mackay)

Method Metadata Fields	Type	Description
Method_id		*Internal use*
Method	text	Describes the method; e.g. "Salinity tolerance using the method of Munns et al (2003)"
Unit[1]	text	Only used when the measurement is numeric; e.g. g or grams, kg or kilograms, cm etc.
Options[1]	text	Used only if evaluation is done using a scoring/rating or similar system; e.g. "BL, Blue; BK, Black; BR, Brown; GR, Green; YE, Yellow; RD, Red".
Range[1]	text	Only used if the measurement is numeric; i.e. the Unit field
Field Type	integer	0 = character; 1 = numeric; 2 = integer
Field Size	text	*Internal use*

[1] **Either Unit or Unit and Range are used and Options is not used or only Options field is used.**

Location Metadata Fields	Type	Description
Location id		Internal use
Institute	text	WIEWS (FAO) code
Title	text	Text name for experiment; e.g. Ethopian wheat salinity 2004
S_Date	text	Start date of experiment. Format yyyy-mm-dd (leave mm and/or dd as 0 if not known)
E_Date	text	End date. Format yyyy-mm-dd (leave mm and/or dd as 0 if not known)
Location	text	Experiment location; e.g. city, farm; city, country – if no latitude and longitude
Lat	text	Decimal latitude
Long	text	Decimal longitude
Alt	integer	In meters
Citation	text	Citation of publication about the experiment if available; up to 255 characters
Description	text	Description of experiment that people visiting the global portal will see. Could be a nursery name and number. Unlimited length
Environment	integer	0 = field environment; 1 = controlled environment (glasshouse, growth cabinet etc)

Upload mechanism for Characterization and Evaluation data

The proper process is not yet fully identified, particularly for the evaluation data. The last upload of legacy C&E data simply followed the way Centres provided data so it was easy for genebank database managers to extract the data. Bioversity centrally uploaded the data into GENESYS.

The new GENESYS upload tool called Direct Data Control (DDC) currently offers a simple solution to upload small updates; however, there is still a need for upload solutions addressing the large genebanks' data sets. In the DDC, there is one directory per crop and one subfolder per trait. A script slices the trait data into the trait table and the experiment into the trait subfolder. The classification is made by year then by location, by trait and by crop.

The GCP work on a Trait Dictionary and Ontology mapping could help annotating the traits that are harmonized.

It is necessary to increase the communication between the genebank database managers and GENESYS, particularly for new type of data and method. Start date and end date of an experiment is hard to provide because it is not always recorded; most often just the year is indicated. Rainy and post rainy field: same accession will record two seasons for the same experiment.

Characterization and Evaluation data curation

- At what level is the curation for GENESYS needed? The aggregation brings additional ways of checking data against larger sets obtained from various sources. It allows testing of the occurrence of the new trait.
- What data curation will be applied on the controlled term needed and the methodology used? GENESYS will accept the data as they come in. There will of course be a need to curate the data as they are received but the curation will depend on the nature of the crop.
- Do we need to stamp the date of when data is submitted or changed?
- Different types of data are linked to the accession. An accession entered in to GENESYS should not be able to be changed subsequently. Should restrictions in this sense be taken in to consideration?

The type of data to be aggregated at GENESYS level was discussed, particularly for the evaluation. One evaluation is often the result of 50 iterations so GENESYS should only capture the summary data and metadata along with time series data and rank them instead of aggregating raw data. Data received should not be changed before being published on GENESYS. For a certain set of materials evaluated under the same conditions, a ranking or scoring can be applied to enable comparisons. Adding massive amounts of raw data may simply increase confusion for final use so it will be better to attempt to provide consolidated scoring. . In the case of diseases, reporting histogram using averages may cause confusion. Would an annotation tool for the characterization data be useful to help the crop database curator to add metadata for the upload?
At the global scale, Evaluation data is complicated to handle and a social network approach can help to solve this aspect, as it enables a large audience to comment and share views on experimental design and practices. Comments posted will create an evolving knowledge base on around 7.5 million accessions managed by the community itself. Crop groups can provide the ranking on the data.

GENESYS' objective is evolving in Phase II but needs a more clearly defined perspective with regard to the Characterization and Evaluation data. It aims at supporting breeders who are looking for specific characteristics

Recommendation 5 - Characterization and Evaluation data management at the level of GENESYS
A drastic revolutionary rethink on how to handle this data, particularly the evaluation data, is required, as it seems that GENESYS will be facing a never ending process which eventually will become unmanageable.

The following paper should be studied attentively:
Jeffrey W. White and Frits K. van Evert. 2008. Publishing Agronomic Data. Agronomy Journal Volume 100, Issue 5.

Update frequency for Characterization and Evaluation data.

- On-demand and, if not possible, update once or twice a year.
- A reminder once a year would be good.

No central repository needed

The storage of the raw and analyzed data remains the Centers' duty, on institutional repositories or databases. There are certainly problems of storage space and some data are only useful for a limited time. Links with GENESYS can then be created to access the detailed data, as long as the data sets are available online. In most of the Centers, evaluations are performed in distributed sites and data are not always systematically centralized. ICRAF distributed network in eco-geographic regions will need to be addressed.

Session 6 - Germplasm Transfer data

Debriefing of the recent experiences in compiling and analyzing the distribution data needed for preparing the reports to the Governing Body of the International Treaty on behalf of all the CGIAR.

Preparing the report to the Governing Body of the International Treaty

Compiling and analyzing genebank and non genebank transfer data at the system-wide level – presentation-Tom Hazekamp, Michael Halewood

The CGIAR system-wide report is much appreciated by the Treaty signatory countries and was initiated since 2006 when the Treaty came into force. The system-wide report provides transparent information on the international collections' activities regarding the germplasm transfers. The Treaty Secretariat, on the behalf of the Treaty Governing Body (GB), may eventually be in the position to perform these statistical analyzes on the future but not presently, as they do not have the experience and resources to do that.

Each genebank manager presented an overview of their Centre's situation for genebank and breeders' material exchange (See tables and breakdown per center in Annex 3 and 4). These overviews mainly indicated that the time required to collect and provide all data to Bioversity is two weeks, once they have access to the breeders' data.

ICRAF mentioned that the direct distribution of genebank and breeders' material is mainly local. Data are available in many regions and they are in different format. The application of the template will help standardizing.

The use of a summary data template in 2009 was not very much appreciated by Centres. The accession-level template validated in 2010 provides more useful details and filters for the statistics. While the genebanks' data can comply with the accession-level template for germplasm transfers, breeders will find it difficult to complete all the details in the form to be filled in, particularly the section on acquisition. A focus could be given on documenting the international nurseries' distribution.

For CIMMYT, data from breeders are received in SQL but they are not centralized. Breeders have sample level data on acquisition and distribution and not accession-level.

If we look at the entire process, there are different approaches at the Centres level (centralized, decentralized). It is difficult to get the data from breeders and questions were raised on the fact that some breeding material can also just be temporary germplasm material.

The inclusion of the breeders' material into the SINGER report was questioned. There is a need to obtain the Centers' commitment in organizing the process upstream and making sure that Units distributing the gemplasm, breeders and breeding database managers are aware of the reporting and are prepared to contribute. **Breeders need to validate the template.**

Upload system for germplasm transfer data

- Should be an 'On demand' upload.
- Accession-level distribution data are numerous which means that large size files and Excel documents cannot accommodate this amount of data; data are therefore exchanged in SQL /Access format.
- Routines need to be developed to extract and collate data from the relevant sources.

The system-wide report on 2009 germplasm transfers will be sent back to the Centres for their approval before being sent to the Governing Body of the Treaty in March 2011.

Recommendation 6 - Collating germplasm transfer data on a yearly basis

The Governing Body of the Treaty does not meet every year and will need the system-wide report every two years. However, it is important to compile the data on a yearly basis and fit the report within the calendar year. The delivery of data through SINGER will be ongoing. A template for statistics on breeder's distribution will be developed by Bioversity as opposed to the accession-level template for genebank data. The template will be submitted to the approval of Breeders. However, a long term decision needs to be made by the ICWG-GR about SINGER data sharing process handling or not all breeders' distribution data.

Recommendation 7 - **Awareness of top level management to obtain an institutional commitment to the reporting on genebank and non-genebank material**

Before March 2011, the Inter-Centre Working Group-Genetic Resources (**ICWG-GR)** must inform the Centers' Directors and Directors of Research about the importance of this yearly based system-wide report to the Governing Body and ask their support in obtaining the units organized and support the data collation. If necessary, the Consortium Board can also be alerted to this need.

Action 4 – System-wide report to the Governing Body of the Treaty - The acquisition and distribution data for 2010 must be compiled by the end of 2011 and then the statistics will be produced in 2012 for the Governing Body meeting.

Action 5 – System-wide report to the Governing Body of the Treaty - Bioversity will provide a summary template for breeders' data.

Session 7- Data Collation – safeguarding data sets, data upload and data sharing

Large data sets upload - presentation of EURISCO upload system for passport data and questions

Milko Skofic and Sonia Dias provided background information on EURISCO and performed a demonstration of the upload system. The taxonomy used to check taxon names is GRIN.
When the upload is done, the providers perform a full update of their inventory. An update at the accession level will be provided in the future. The automatic transfer to GENESYS exists.
Sixty-six percent of the countries publish their data on their own site, so for 33% of countries, EURISCO is the only way of publishing their data online. These genebanks cannot access reliable connections in their countries, due to lack of support staff. Each collection provides data to the National Focal Points who publish their data. In the future, each curator will be able to upload his/her data and check it. For countries with their own site, once they have received feedback from EURISCO, they also correct their own. A proposal was written to obtain resources to perform the quality check on the millions of records available.

Action 6 – Test of EURISCO like upload - Once the EURISCO upload system is revised, it will be tested by the Centres.

Regular updates -presentation and practice of Direct Data Control (DDC)

One of the issues raised during the presentation made by Fawzy Nawar was on how to deal with trait heterogeneity, both in the way they are named and measured. It might be interesting to know which accessions have been tested with two different methods.

An algorithm can be developed to map the received values against the accepted value within a crop, if all the methodologies are loaded in the same place.

The proposal is to test the DDC during 2011 and provide feedback to the group. Only 50% of the accessions from the CGIAR have C&E data in GENESYS so genebanks need to send the characterization results they have to open the channel to partners and to convince donors.

Action 7 – Characterization data for GENESYS - A first CSV file for the legacy C&E data will be sent by mail to Bioversity (Fawzy Nawar). Once legacy data are uploaded DDC will be used for the subsequent updates and corrections to upload the updates. ICARDA and AfricaRice agreed to test the DDC in 2011.

Web services and other solutions already in use at the Centres level

Matija Obreza

- Individual genebanks are information providers.
- Web services enable to share data more easily and automate.
- Automation reduces errors, unattended integration.
- Requires standards, formats, documentation.
- Improve connectivity of IT systems.
- Mostly applies to legacy systems.
- IITA uses small individual systems, accession database, etc.
- On-site integration of the LAN.
- Off-site integration between different Centres. Message queuing. Reliability of delivery. Mostly open source software.
- Reliable/less reliable integration of different systems. Different types of messages, files.
- Summary. Flexibility in individual components of IT systems. Require backward compatibility or upgrade to related systems.

- Central systems need to provide automation mechanisms as defined requirements and standard mechanisms (MQ, WS).
- Data providers trigger integration.

GRIN-Global is made of web services. SINGER is not web services-based. Web services can be a solution for new evaluation data to send to GENESYS. Some examples on web services did not work well because they were imposed by the data aggregator and not selected by the data provider. Instant updates are the great advantage of the web service and enable automation, and updates use less bandwidth. Complementary methods are offered by web services and the extension on centre-own systems possible. It requires an appropriate code to take comma separated versions of the files. Automation mechanisms require good documentation and clear system requirements. We need to be exact on the standards to apply and the recommendation has to be followed by the Centres. Automation is seen as “nice dream”.

GRIN-Global Golden candidate

Sonia Dias gave a demonstration of GRIN-Global on the following features:

- What can be accessed, added or changed?User set up.
- Data view files and tables.
- Data triggers for the tables.
- Maintenance.
- Easy way to import into GRIN.
- Possibility to publish via website.
- Curator tool.
- Finding accessions and cooperators and saving them in a personal folder.
- Editing accession and cooperator data.
- Customize genebanks.
- Export information on CSV file.

Overview by Michael Mackay

- Data migration (import wizard) is only a month or two old.
- Data server could be in Washington or anywhere in the world.
- All security measures can be added.
- Data views not only relate to client and user but also to the web portal.
- Currently accessions are based on prefix, number and suffix but this can be changed.
- Use administration tool to make changes, added flexibility.
- Some effort may be required for users to learn to use the tool.
- Very flexible, can be tailored.
- Next step is to develop an open source community.
- All software is free.
- Anything can be changed in the middle tier.
- Extra functionality (e.g., pedigree) can be added.
- It is a genebank management information system (not a portal).

GRIN-Global can be installed on a local network and curators can set up the system to store the data they want, e.g. characterization, raw data. With the administrative tool one can add descriptors as needed. Data that are in Excel files can be selected, copied and pasted into GRIN-Global. An automatic selection of the sample for characterization is possible.

GRIN-Global is particularly flexible, even the code in the middle tier can be changed. Wizards have been added, e.g. accession wizard. GRIN-Global has been defined using the USDA parameters and fields but this can be changed through the admin module. There is a dictionary to understand the system.

It was proposed to set up a group for GRIN-Global within the SINGER community. The need to have an open source community or at least, a GRIN-Global user community was discussed again.

- How do you build such a community?
- What are the tools, what are the steps?
- How do you see this community being developed?

	Data sets	type
EURISCO-like system	Large and updates	push
DDC for Genesys	Small updates	push
Repository	any	upload
Web services	any	pull automatic

Discussion on which are the sustainable existing solutions to address needs at the Centres level, SINGER/ GENESYS level

- How can we develop or become a CGIAR open source community for GRIN-Global?
- What do we really mean by an open source community?

In the SINGER meeting in 2009 held at USDA-Beltsville, the group indicated that a CGIAR user or open source community for GRIN-Global should be set up. The release of the first stable version of GRIN-Global was delayed by 18 months but now, we can start considering the version that is just released.

Why a user community or open source community in CGIAR for GRIN-Global?

There is an opportunity for all the CGIAR genebanks to switch to this system, should they decide to do so. CGIAR Centres have specific needs that the present version of GRIN-Global does not accommodate but features can be added as necessary by the CGIAR community. If Centres adopt GRIN-Global, they should be careful not to develop additional features on their own and must share information to keep as far as possible a coherent version. An expert helpdesk that has a strong understanding of the database structure, the technology, the process and can answer questions will be needed. CGIAR Centres are facing problems of resources and duplication of development. It is more cost effective to adapt and fix bugs in GRIN-Global than to maintain or redevelop obsolete systems independently. A user community can provide the solution for sharing resources between Centres. Bioversity will deploy GRIN-Global but funds for development are required.

Requirements for building an open-source community around GRIN-Global

a. A free documented code

If we develop a user community then all modifications made in GRIN-Global have to be 100% documented and defined. The code of GRIN-Global is free. The source code of GRIN-Global is a Public Good like all what USDA produces being a U.S. Federal Agency. There is a commitment from USDA that GRIN-global has been developed to serve the American genebanks and it will be further developed and maintained. USDA will deploy GRIN-Global in the US and will fix any bugs that may develop. GRIN-global does not accommodate all CGIAR needs so extension of the code will be needed, e.g. inclusion of a pedigree system.

b. An Active expert helpdesk

No helpdesk will be provided by USDA outside of the US. A helpdesk and someone able to reply to technical issues is required. There will be a possibility to receive technical support from the US and also for installation.

c. A discussion forum on GRIN-Global and shared tools

This already exists for GRIN-Global but still needs to be developed further.

d. A hands-on workshop for CGIAR developers

How big is the CGIAR learning curve for GRIN-Global? How much time is needed?
A hands-on developer workshop will be necessary to figure out how easy or difficult it is to be used.

e. A group of interest, a critical mass of IT experts

The position of each of the Centres with regard to the adoption of GRIN-Global was solicited during the session to assess which ones could be part of a user community.

- Adopting GRIN-Global
  - CIMMYT
- Presently testing it
  - ICRISAT
- Considering the option & ready to test
  - ICARDA
  - CIP
  - Bioversity-*Musa* genebank
  - AfricaRice
  - ILRI
- Not presently considering
  - IRRI
  - CIAT

f. Breeders and database managers should be part of the community, particularly the International Crop Information System ICIS community, as several Centres manage their breeding data using ICIS.

Action 8 – GRIN-Global user community - Feasibility (see list of agreed action points).

Recommendation 8 - **Develop a joint proposal with Centres adopting GRIN-Global for a system-wide hands-on workshop for evaluating the data migration possibilities and efforts**

The group suggested that a CGIAR user or open source community for GRIN-Global should be set up with participation of managers of diverse data systems in CGIAR to guide/steer adoption of GRIN-GLobal, sharing resources, etc. Once the final first version of GRIN-Global will be released by USDA, the first step might be a system-wide hands-on workshop for evaluating the data sharing and migration possibilities.

DAY 3 – Expanding and strengthening the system-wide collaboration

Session 8 - Expanding the system-wide data standards

The content of this session was modified due to the fact that none of the participants attending the session on Day 3 could represent some projects listed in the programme.

The Crop Ontology (CO)

The use of ontology terms to describe agronomic phenotypes and the accurate mapping of these descriptions into databases is an important step in comparative phenotypic and genotypic studies across species and gene-discovery experiments. The key to data integration (across different sources and disciplines) is to have consensus on the concepts and terms to use along add inter-relationships between terms and definitions that describe data.

The Crop Ontology (CO) is then a system-wide effort to apply a common methodology and is based on existing data sources, as well as on the results of a phenotyping project. The curation and development of the crop-specific terms should be sustained by the individual crop programmes and promoted to National Agricultural Research Systems and Advanced Agricultural Research Systems (NARS/ARIS) through a collaborative project on phenotyping.

The CO currently comprises crop-specific traits for chickpea (*Cicer arietinum*), maize (*Zea mays*), potato (*Solanum tuberosum*), rice (*Oryza sativa*), Rice mutants, sorghum (*Sorghum* spp.) and wheat (*Triticum* spp.). The Cassava (*Manihot esculenta*) Ontology was developed by IITA in 2010. Several plant-structure and anatomy-related terms for banana (*Musa* spp.), wheat and maize are also included. In addition, multi-crop passport terms are included as controlled vocabularies for sharing information on germplasm. Up to now, Crop Ontology terms have been integrated into major crop databases, trait names were mapped and terms are being used to curate several CGIAR Centres' agronomic databases by Centre and map trait names to the to Crop Ontology terms:

- All 163 maize traits in the International Maize Information System (CIMMYT).
- 300 Wheat traits upon 549 included in the International Wheat Information System (CIMMYT) with experimental design factors.
- All 500 Rice traits in International Rice Information System (IRRI), along with experiment and design factors.
- 120 traits were mapped for cassava (IITA) and are included in the new cassava database.

In order to enable a friendly ontology curation and data annotation, a prototype of a distributed tool will be developed and tested by crop breeders' database curators in 2011. A global coordination or a consortium will be needed to maintain the global Crop Ontology and related tools, stimulating its curation. It will also help to sustain the necessary network contacts for the terms to be properly mapped to the global concepts, across crops, across molecular data and phenotypic data and will also provide a basis for prospective research for the use of the Ontology in Web2/3 GENESYS. The Integrated Breeding Platform (IBP) will provide channels for getting new concepts submitted by crop communities.

Session 9 - Data publishing, annotation, citation

Example of the repository of the Collecting missions files

Presentation by Massimo Buonaiuto of the implementation of the crop collecting missions' repository http://www.central-repository.cgiar.org/

The team performed the analysis of types of reports uploaded by IRRI for Rice (Including those from AfricaRice, the Agricultural Research Centre (ARC) of Lao People's Democratic Republic) and Bioversity. The types of documents could be grouped and then metadata analyzed. Metadata describe the content of the documents and Darwincore- germplasm was used with additional fields to describe the content. The repository includes the following technologies: Typo3 to manage the content (db) and to publish easily online, Alfresco DMS to manage live cycle of documents and the upload workflow. A search mask was developed in collaboration with IRRI to provide public access. A URL is given to each file for Centers to display the files as links on their website within a Passport. IRRI is now linking the full text for Rice collecting missions from the repository to the crop registry level and AfricaRice would like to be able to make similar linkages. IITA, ICRISAT and CIAT revised original data and scanned documents but not yet those of the repository. The CIAT scanned documents are available on the CIAT website.

Optical Character Recognition (OCR) was not really a solution as many reports are hand written which makes the automatic reading difficult. Each centre worked independently. Repository is searchable via search engines. In some cases, all information relating to a mission is in one pdf but in future, there might be the possibility to split the files per collecting form so access can be done per accession.

PDF Repository	
Scanned Pages	59 700
PDF Files	3120
GB Scanned	4.7 GB
Pages scanned by Partners	26 000

Example of the GCP central Registry

A demonstration of the GCP central registry, which offers registration, upload, file retrieval and download features, was made. Metadata are added during registration to describe the file. The attribution of rights can be selective to partners. Guidelines are available for upload, as well as a submission template. This repository was developed to be a central resource for the GCP community and provides 255 data sets along with the fully documented data templates to a wide audience.

Data attribution and data citation: Practices in attributing metadata to data sets

Data attribution is essential for accountability and recognition. Research data must be treated like scientific literature. The use of a versioning system and integrity checks are very important. Proper data management demonstrates a good use of public funds. The data management principles published by the Organization for Economic Co-operation and Development (OECD) are a reference for collecting accurate data, organizing data, protecting and safeguarding, archiving and analyzing and communicating data. These OECD guidelines cover sustainability, evaluation criteria, extent of reuse of data and protect data for long-term storage.

Metadata annotation must use format and standardized tools. IP on data and other products often rests with the employer (the center); within that the data generator and curator (may not be

the same person) share a right of recognition as authors. A grey zone exists with respect to which amount of human work into a database merits the phrase 'intellectual input'. It was indicated important that data (passport, characterization, evaluation) retains the documentation on the authors (data generator and curator), not only for recognition but also for credibility of the information. These data authors can be associated to their publications, scientific papers, etc. which can be easily referenced.

We need to look at what is achievable and realistic at the level of a global platform. It is not ideal to cite the aggregator but there is still not a better method so far.

The genebank curators mentioned that in SINGER there is no proper citation. SINGER data can be downloaded by anyone so the citation must be downloaded with the accessions data. All the data collectors must be listed and all modifications or supplementary work performed on the original data must be indicated. A contact person for the data sets should always be provided.

In the case of SINGER, if you are downloading the data, each download should have a citation. But if you download a batch of data from several genebanks, all citations should be included. At least you provide the data. You can cite a summary publication in further steps. All original references can be included in the summary.

Online citation method should be explored as more and more datasets and references are online.

Session 10 - Infrastructure and collaborative tools

What infrastructure and collaborative tools are needed to support the system-wide informatics activities in terms of community development, knowledge sharing and outreach? The AAA framework ICT-KM is a programme of the CGIAR with the mission of developing and promoting tools for Information and Communication Technology and Knowledge Management. ICT-KM provides strategic information and directions to the CGIAR. The CGXchange project aims to include tools for knowledge management, knowledge sharing and opening access to collaboration. The CGIAR should make its research available for the benefit of the international science community. Research outputs should be communicated and used as a public good. Scientific information must be available, accessible and applicable, hereafter referred to as the AAA framework.
Fundamentals of the AAA framework:

- Available = can I find it?
- Accessible = can I access it?
- Applicable=are outputs re-usable?

The challenge is to work collaboratively in different organizations, time zones, etc. How to be more efficient? How to increase research impacts? How can we change toolset and mindset in a context like this? ICT-KM has developed a framework to address this challenge: benchmark studies in collaboration with several institutes: http://ictkm.cgiar.org/what-we-do/triple-a-framework/.

Another way to look at the AAA framework is to see it as knowledge sharing within the research cycle: from the identification of the product to the production of research outputs. For specific questions of collaboration, specific tools have been proposed by ICT-KM. The group took time to look at two different types of communities (stakeholders, target audience, etc.): 1. Internal (within the team); and 2. External (public audience) and identify roles and requirements. The group was then asked to identify what the two communities have in common and where they differ to illustrate that boundaries, in terms of knowledge sharing, are fuzzy between what is internal and what is external. Are we seeing different roles and activities? Then let's identify common elements (roles, types of information and tasks).

Collaborative and social network tools to achieve collaboration and knowledge exchange

Facilitating communication is crucial and the tools available to do so are many. CGXchange is a toolkit that includes collaborative technologies with Google Apps, Google calendar is accessible to http://calendar.cgxchange.org. There are several examples of Google Sites, like the IT managers' meeting site.

Social tools represent another way to communicate (blogs, twitter, etc.) each one with specific characteristics and specific usage. Examples: news story in ILRI blog or a facebook page (http://www.facebook.com/ILRIFanPage), CIFOR, Facebook page (http://www.facebook.com/cifor?v=wall) or blog projects like the Fodder Adoption blog. Most of these Centres use Flickr to share photos; IFPRI has a video channel on Youtube; microblogs on Twitter, etc. Another important tool is Newsfeeds, RSS feeds; Mendelay to share references in academic communities; RSS delicious.com for bookmark Mendeley – Academic social net – reference manager and academic social net; Webinar, goto meeting; Dimdim screen sharing are other examples.

Combining these tools allows communication to be more efficient.

The objective is to facilitate the knowledge sharing, and SINGER as a network, that can also benefit from this. The visibility of the genebanks can be improved by using a combination of these collaborative tools and social network tools:

To make this kind of collaboration happen, we reconsider our daily work:

- What do we need to do? What for and for whom?
- What type of information do we need to share?
- How it can reach other information/informatics professionals?
- What is the scale of impact?

Google makes sense of the content, if it is accessible, and it can help to promote the deep level data that are locked in the database. A RSS can be setup for tracking the germplasm database updates. Anything structured can have an RSS versioning. **The "Deep web" is being made searchable by Google,** e.g. GRIN data were made 'indexable' by Google and the traffic increased tremendously.

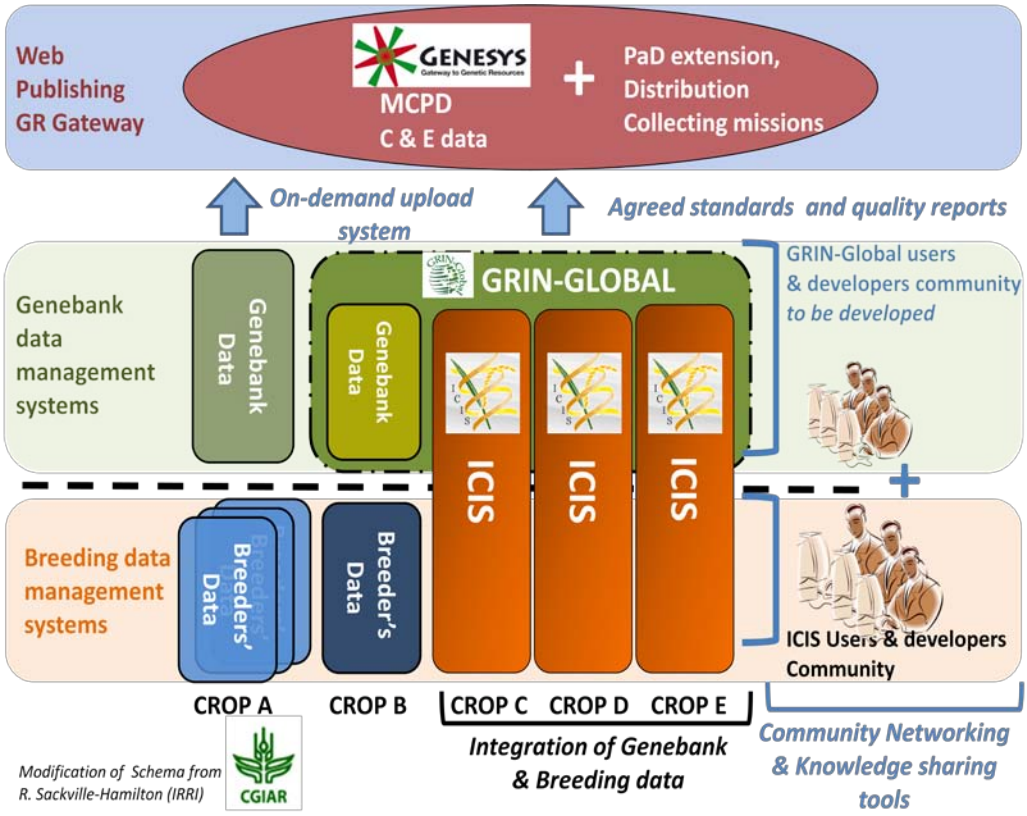
The product documentation must not remain behind a password. The database must be documented and links to data sets should be added to provide examples. Non password protected wikis should be used.

SINGER is a small group and initially it can be kept as it is. ICT-KM team can work with us to put the SINGER group on CGXchange, and is ready to provide training to the group. They need contacts to work with to identify the tools and the training needs.

Recommendation 9 - An Online open space for group discussion standards is required
The group needs to have an open space where it can discuss the standards like the SINGER data warehouse dictionary that include the MCPD. Milko Skofic and Luca Matteis (Bioversity) will look at the potential of Google Apps and CGXchange for publishing and commenting the dictionary.

Day 3 - Summary, conclusions and agreed outputs

The figure below presents the information management elements identified by the group for the CGIAR community. The molecular data management systems are not considered in this schema as it was not discussed.



Session 11- SINGER visibility into GENESYS and access to the data

Discussion was based on the suggestions posted during the meeting on a paper board under the following written question:

How can SINGER best contribute to GENESYS and how can the GENESYS site provide best access to SINGER data?

What visibility?

Web visibility sounds important for the perception of a corporate identity of the international collections. The international collections are owned by the world community and managed by the CGIAR on behalf of this community. Donor countries want to know where the material that was originally collected from their diversity is conserved, check that it is visible, accessible and safely conserved. Furthermore, countries want to be able to access the material and retrieve it, if necessary. A purely 'CGIAR' identification is not desirable on GENESYS and emphasis should be made on the 'international collections' that are held in-trust by the CGIAR Centres on behalf of the donor countries. GENESYS should not give the impression that the CGIAR takes ownership of the

collections. Ideally the access window should be the source or donor countries. Consequently, information on the donor and source of the material has to be as complete as possible and visible.

The group agreed that, in GENESYS, international collections must be easily identifiable by the users because they need to know that these collections are maintained by Centres under agreed conditions with the International Treaty for safe long term conservation and for free distribution along with an SMTA.

A single portal

A plan for GENESYS that can be presented to donors is necessary, but donors will probably be most inclined to support a single global system. Users need a single door which reveals all the answers, from where they can access everything they need in a consistent manner. A Global system is the accumulation of the community of practices and brings more power to raise funds. The SINGER group of genebank database managers must contribute to the development of clear plans for the GENESYS Phase II proposal. Curators must be brought into the discussion about how they want to see the windows within GENESYS II and what services are needed.

SINGER is a network and a community with particular practices that has a model role to play within GENESYS. Not all GENESYS data providers will be in a position to provide the same information in the short term and the SINGER group may need to be dealt with in a specific way with regard to GENESYS. An open discussion on a forum dedicated to GENESYS should be initiated to get the SINGER audience perspective and make more voices heard. Therefore a decision on whether to keep or abandon an identification of the SINGER services within GENESYS (e.g. distribution data, collecting missions) will be adequately made.

GENESYS contains additional C&E data but the Passport data in GENESYS are more limited than in SINGER. So the issue is where to publish the data traditionally maintained and exchanged by Centres with SINGER when the SINGER site no longer exists. The definition of a global system being a single portal composed by several windows needs to be clarified.

CGIAR Germplasm Transfer data

Originally it was not planned for GENESYS to include the distribution data because the global system will not receive this type of data from all data providers. However, SINGER was created to fulfill this particular need of providing transparency on CGIAR germplasm transfers after the specifications of the Convention on Biodiversity (CBD). There is here, once again, a model role in GENESYS for SINGER members that openly publish the distribution data. The Governing Body of the Treaty is expecting a certain amount of data about the germplasm transfers, so data could be entered in GENESYS and the model role of the international collections highlighted. The SINGER website will disappear once GENESYS will fully address the needs of the international collections. Once that happens, then the distribution data must find a place in GENESYS. Of course, GENESYS will need to distinguish the processes between community of practices like SINGER and EURISCO.

Recommendation 10: Visibility of SINGER and international collections in GENESYS
It was recognized that users need a single door which reveals all the answers; from where they can access everything they need the required information in a consistent manner. SINGER is a network and a community with particular practices that has a model role to play within GENESYS. The group recommends that, in GENESYS, international collections are easily identifiable by the users. The definition and a model of a global system as a single portal composed by several windows need to be clarified and developed. Therefore a decision on whether to keep or abandon an identification some of the SINGER services within GENESYS (e.g. distribution data, collecting missions) will be adequately made by the ICWG-GR and SINGER users.

A single middleware and one upload mechanism as necessary first steps

There is a transition phase between the two systems and SINGER will have to remain for a while longer.

A clearly defined workflow for updating the collections which form the backbone of a multilateral system is necessary and was the objective of this workshop. Now, it is urgent to see how quickly we will obtain the data in the middleware after which we will define a way of creating a window that can satisfy our external users. The middleware code is the expensive part of the system and the first step in cost saving is to have only one middleware. The question was raised about choosing one database model but the single database is not the crucial element, while having one data source is**. Websites have to take the data out of the same storage and apply same publishing rules.** At the moment, the GENESYS database takes up just partial data from what is provided by SINGER then all the other data must be stored somewhere else. What is really important is that information in SINGER and GENESYS is consistent for the duration of the transition phase remains. It will be helpful for database managers to submit data to only one portal.

Recommendation 11 - **Elements for the integration of SINGER into GENESYS**
The group recommends that there is **one common middleware, a single data storage and one upload mechanism**. The upload system should accommodate PaD, characterization (Field/Molecular), evaluation, distribution data. Centres will upload MCPD extended data plus distribution into the middle tier. This will improve the quality of data and data documentation whatever solution regarding the portal is adopted. SINGER and GENESYS must share the same rules for online publishing of the data for users, providing the same quantity and quality of data on both.

Addressing particular needs of the international collections like ICRAF and Bioversity-Musa

GENESYS will have to accommodate ICRAF's particular situation, as their information is not centralized and germplasm is located on different sites, in farmers' fields within the eco-regions, in different countries. This situation will probably not be isolated in a global system. ICRAF HQ acts as a hub for data collation from fields in Cameroon, China, Ghana, India, Malawi, Mali, Peru, Sri Lanka and Tanzania. Currently, on SINGER, data from ICRAF Genetic Resources is minimal and does not reflect the reality so the situation must improve when using GENESYS. The ICRAF information system was inherited from Oxford, UK and there is a way forward to see how to use and adapt the current SINGER data dictionary for data on trees. There is some level of field and molecular characterization of fruit trees but there has never been an attempt to centralize this type of data. How can this data be integrated into GENEYS?

Same need applies for the *Musa* network where only the *in vitro* collection is reflected in SINGER while the field characterization performed by NARS on the germplasm could not be published.

There is also a need to add specific descriptors and quality georeferences.

How to increase the visibility of genebanks and their online databases?

The Trust is carrying out a comparative study on the access of the Centres' genebank databases and it appears that there is a regression, and that the access is very variable, not harmonized across the Centres. Suggestions posted on the board:

- Lobby Centres to include a link on their homepage. Web marketing activities.
- GENESYS should provide access to the genebanks' websites.
- Use of GRIN-Global that enables a website for genebank

- Use of social network tools, RSS feed on data upload, mark database content for Google access

Action 9 – A first version of the workshop report will be provided by the end of January 2011 to obtain Centres' comments before the end of February 2011. A list of actions was established (see table below).

Summary of discussion points for further consultation

1. SINGER contribution to the plans for GENESYS phase II
Curators must be brought into the discussion about how they want to see the windows within GENESYS II and what services are needed. An open discussion on a forum dedicated to GENESYS should be initiated to get the SINGER audience perspective and make more voices heard.

2. Particular needs of crop networks in GENESYS
GENESYS will have to accommodate ICRAF's particular situation, as their information is not centralized and germplasm is located on different sites, in farmers' fields within the eco-regions, in different countries. This situation will probably not be isolated in a global system. Currently, on SINGER, data from ICRAF Genetic Resources is minimal and does not reflect the reality so the situation must improve when using GENESYS. Same need applies for the *Musa* network where only the *in vitro* collection is reflected in SINGER while the field characterization performed by NARS on the germplasm could not be published.

3. Web access to the genebank databases
The web visibility of Centres' genebank databases appears to be in regression, and the access is very variable, not harmonized across the Centres. It therefore recommended to lobby the Centres so a link is included on the homepage of the institutional web sites and some web marketing activities should be initiated, like the use of social network tools, RSS feed on data upload, mark database content for Google access. GENESYS should provide access to the genebanks' websites

Summary and validation of the elements to apply for a system-wide quality data sharing process

Recommendation	Action	Who	Deadline
Action 1 – Templates and standards will be made available on a wiki to enable comments			
	Set up a CGXchange website for SINGER group	**Luca**	As soon as possible
The standards with their guidelines need to be shared on Google Docs of CGXchange for SINGER user group comments	Publish the SINGER Data dictionary on Googledocs or wiki to enable comments and questions, votes	**Milko and Luca**	End of January 2011
	Comments will be provided to the FAO/Biodiversity group revising the MCPD	**Elizabeth**	Mid February 2011
	Add on the GoogleDocs the data quality guidelines published by Theo van Hintum as a reference	**Milko/Luca**	End of January 2011
	Comments on the standards posted	**Genebank database managers & curators**	Ad Hoc
Action 2 – Germplasm request gateway - The list of requested accessions will be attached to the request email as an Excel file in order to be easily processed by the genebank curators.			
	Modify the mail sent by the Germplasm request gateway	**Luca**	End of February 2011
Action 3 – Germplasm request gateway -The registration form developed by the Treaty does not include the type of cooperator as per the MCPD and it makes it difficult to compile the information per category for reporting to the Treaty.			
	Bioversity will contact the Treaty Secretariat to suggest adding institution type	**Luca/Elizabeth /Francisco Lopez (Treaty Secretariat)**	End of March 2011
Action 4 – System-wide report to the Governing Body of the Treaty - The acquisition and distribution data for 2010 must be compiled by the end of 2011 and then the statistics will be produced in 2012 for the Governing Body meeting.			
	Collect and format 2010 acquisition and distribution data for genebank and	**Centres' genebanks and breeders,**	August 2011

	breeders' material for 2010	**database managers**	
To bring breeders and their database managers in the picture	Alert the Centers' DGs about the importance of becoming organized at the institutional level and supporting the data collation for genebank and breeding material	**ICWG-GR**	Before March 2011
Action 5 – System-wide report to the Governing Body of the Treaty - Bioversity will provide a summary template for breeders' data.			
	Develop the summary data template for the breeders' material and validate it	**SINGER and Centres' breeders**	Before the end of 2011
Action 6 – **Test of EURISCO like upload -** Once the EURISCO upload system is revised, it will be tested by centres			
Develop one data upload system within a single warehouse with quality control reports and specific formats	New EURISCO upload system as a single system will be tested for large data sets	**Milko**	April 2011
Action 7 – Characterization data for GENESYS - A first CSV file for the legacy C&E data will be sent by mail to Bioversity (Fawzy Nawar). Once legacy data are uploaded the DDC will be used for the subsequent updates and corrections to upload the updates. ICARDA and AfricaRice agreed to test the DDC in 2011.			
	All accessions with characterization data will be sent by ~~mid-February~~ mid-june for the 22 crops of GENESYS in csv file by mail to Michael Mackay, copy Milko Skofic and Luca Mattei	**Centres**	~~Mid-February 2011~~ Deadline modification by Michael Mackay: Mid June
	Central upload in GENESYS	**Fawzy Nawar**	March 2011
	Test of DDC for updates by ICARDA and Africarice	**Fawzy Nawar/Jan Konopka/Albert Tchamba**	End of April 2011
Action 8 – GRIN-Global user community - Feasibility			
	Proposal to organize a hands-on workshop for evaluating the data migration possibilities and	**Centres adopting it (CIMMYT) and ready to test it**	End of 2011

	efforts	**(ICARDA,CIP, ILRI, AfricaRice, ICRISAT, Bioversity-*Musa*)**	
	Proposal for a workshop on how to integrate CGIAR Centres' needs into GRIN-Global	**All Centres**	End of 2011
Action 9 – Workshop report			
	First draft report ready for comments and posted on GoogleDocs	**Elizabeth/Olga /Audrey**	15 January 2011
	Comments sent back by the Centres on the report	**All Database managers**	28 February 2011

Annex 1 - Overview of Centres' presentations

Passport data management

Passport data Management	AfricaRice	Bioversity	CIAT	CIP	CIMMYT	ICARDA	ICRAF	ICRISAT	IITA	ILRI	IRRI
For Genebank material											
Format-templates, metadata used	Data source	MGIS dump + MGIS excel file	Passport data material from CIAT-PGR export into SINGER template	MCPD	CIMMYT + Singer (Maize and Wheat)	Database consists of over 30 main data tables. Definitions of all tables and fields available for users from application interface	Species, accession no, mission identifiers collection site or provenance(with documented, collection date, collection number		Custom solution that fits IITA Genebank needs, separate from Inventory system	ILRI Forage diversity database and FoxPro 8.0 software	Passport Data Dictionary
Center's Database(s) and data sources used	ARGIS database & Data source	IMGBMS & MGIS	Passport information (bean, forages and cassava) and data distribution are in the same schema of GRP database	Records, mission reports	Maize. MGB-DB, Wheat. IWIS2 + GMS	Source data from donors' lists and collecting forms; largely conforms to Descriptor Lists for crops published by IBPGR/IPGRI/Bioversity	AFT, TSSD, GRU sources net, NARs, other gene-banks, other NGOs			Forage Diversity ILRI and HTML software	International Rice Genebank Collection Information System (IRGCIS)
Data files and data exchange formats or protocols used for exchange with SINGER?	No data for breeding columns	SINGER DB dump of banana records, necessity to rewrite the queries made for the 2008 upload exercise	CIAT has used Excel and CSV format to send data to SINGER	Offline; tabular format; problems: species 'numeric identifiers' mixed up; mix up with AVRDC	Wheat: some fields (Selection history, Genus, species) not defined into SINGER format. Maize: some fields (Genus, species) not defined into SINGER format	Up to 2010, SINGER-defined structure (in 1995-96) was used. Last data replication in 2010 used MCPD as per agreement in December 2009. MCPD is too simple for proper documentation of accessions' data exchange format from all systems	Data exchange via excel spreadsheets, mailed to SINGER			We have no major difficulties; however, when we send the data it takes several months to update. It is a static system and the data is out of date. Better to be able to give us our own data in SINGER. As it is, SINGER is not a mirror image of our database, because it is not updated	SINGER Data Dictionary. The data file type submitted is in MS Access – Upload_IRRI_2010.mdb. IRGCIS and SINGER, have the same fields but different categories or codes. Usually IRGCIS has fewer categories and SINGER is more detailed. Match the IRGCIS category with the nearest SINGER category, usually the more general classification, then recode the data. SINGER field/variable can contain one value only but genebank accession may have multiple values. Select only one value. But still doesn't reflect the true/complete data Some fields in the SINGER template are displayed differently on the web

										timely	
How is the database and exchange protocol maintenance done?		Running SQL queries & DB dump zipped	Some data using the SINGER template must be aggregated or updated into CIAT-GRP schema views	Manually	In both Crops, once a year.	Database: daily by genebank staff; responsibilities defined as per competence and responsibilities. Exchange protocols: currently we use three main protocols: MCPD, EURISCO, Crop Register Template for propagation of data to different recipients. Additionally, we can populate customized data files to meet recipients' needs	Updated			The selected fields are converted to excel (or tab delimited tables) and sent as email attachments to SINGER	Retrieval of data from IRGCIS to SINGER template is done for every upload. This means recoding/conversion of codes to fit the codes in the template will always be done since IRGCIS and SINGER template use different sets of codes for similar fields
For Breeders' material										None used	No data submitted yet
Format used	SES and plus???		Institutional formats	MCPD	SINGER-CIMMYT	Differs depending on the crops					
Center's Database(s)& data sources used			A database for breeders' passport data is not formally established, each breeder maintains his/ her own data sources	ICG – International Cooperators Guide for potato; similar for sweetpotato under development	Maize. FieldBook-IMIS, Wheat. IWIS2 + GMS	ICIS, MsSQL, Excel files					
Exchange format/protocols used (indicate for what)				Upload to 'biomart'	SINGER-CIMMYT for accesions added from breeders' material	Exchange of EXCEL files in native format for crop/experiment					
Data files and data exchange format/protocols used					SINGER-CIMMYT for new accesion from breeders' material	See above					

Sites of origin

Centres' Details	AfricaRice	Bioversity	CIAT	CIP	CIMMYT	ICARDA	ICRAF	ICRISAT	IITA	ILRI	IRRI
For Genebank material											
Format-template used	Collecting book format	MGIS dump	Origins and collecting missions data from CIAT-PGR have been exported into the SINGER template	Simple tables	SINGER-CIMMYT	Both are integral part of GRSDB. The relevant tables apply to material conserved in ICARDA genebank, herbaria specimens in ICARDA and also all 'register-type databases'. These data are common resource for all data systems for PGR (along with data tables for taxa, cooperators, climate, soil	Accession no, mission code, mission location or site name, mission country, start date, end date, species collected, number of parent plants		Data types based on data originally collected and crop descriptor booklets	They are incorporated in the passport database	Passport 1
Metadata applied	n.a.	NA	Yes	For georeferences	CIMMYT-SINGER standards	ICARDA Collecting Forms are used for ICARDA organized/sponsored missions. The forms are compatible with IBPGR/IPGRI/Bioversity descriptor lists				See definitions file attached	See Passport dictionary
Centre's Database(s)	ARGIS database	MGIS	CIAT Genetic Resources Programme		Maize: MGB-DB; Wheat: IWIS2 DB – IWIS3 DB	Own format for storing summary information on missions: country, dates, collectors, shord description of mission, availability of report(s) and collecting forms	ICRAF GRU database			Done on accession bases, as part of the main database	International Rice Genebank Collection Information System (IRGCIS)
Data files and data exchange format/protocols used for SINGER?		SINGER dump of banana records	CIAT has used Excel and CSV formats to send data to SINGER		SINGER formats	MCPD template was recently used for replication of data to SINGER	Often Ms Excel spreadsheets			Based on accessions, see files attached	Data file type submitted is in MS Access – Upload_IRRI_2010.mdb. Collecting Site

											information is included in ACCESSIONS table, format is provided by SINGER.
Quality process?		Manual Check + GPG II project	Yes, each curator makes the validation for origin and collecting missions on the different material	DIVA-GIS for georeferences	In progress and under improvement	Range of quality checks are used (conformity to coding, mapping, etc.)	Seed quality process involve, collection, processing, quality testing, storage, distribution and documentation			Done with a code database, to look at the codes	Data was retrieved from IRGCIS Passport table. Data quality check is still on-going as we compare the info in the database with the data in the collection forms.
Scanned reports and collecting forms linked?	Documents scanned and sent to Bioversity	Reports scanned and collecting forms not linked	Yes, they are available on the GRP website: http://www.ciat.cgiar.org/urg	Partially	Maize. Since 1995 the most important were reported and published, before 1995 Year reports of the bank	Mission reports on paper, including collecting forms are organized and stored in dedicated file cabinets – not scanned yet. Recent missions frequently turn computer files as reports and/or collecting forms – stored on the server with database (no direct link to database)	Digitization of seed collection forms via scanning) is underway		Collecting forms scanned, but not linked with Accessions database	Not yet	Can be found in http://www.central-repository.cgiar.org/crop_collecting_missions.html

Georeferences

	AfricaRice	Bioversity	CIAT	CIP	CIMMYT	ICARDA	ICRAF	ICRISAT	IITA	ILRI	IRRI
For Genebank material = site of origin											
–Format-template-metadata used (precision. accuracy, etc)		MGIS dump – No metadata (5 digits after the dot)	georeferences data from CIAT-PGR and exported into SINGER template	Estimate in degree based on highest level of precision in annotation	Maize. Precision format: 00.00, Wheat. Precision: DM	LAT/LON are kept both in 'classic' format (HDD MM SS) and as decimal degrees. LAT/LON precision/accuracy flag in 1-9 scale is used (1 – country level, 5 – Admin1 level, 9 – high precision either from GPS or high confidence georeferencing). Altitude is recorded in 2 fields: original data from collectors and derived from DEM. Exchanged with MCPD format to SINGER; 'classic' or 'decimal degree' for other purposes as required	collection sites for specific species already identified as provenances and are already coded			This format is part of passport; accuracy to the nearest 100m	Part of Passport data
–Metadata		NA		Admin levels, locality and habitat description where available						Part of passport	
–Centres' Database(s)	ARGIS	MGIS	CIAT Genetic Resources Programme		Maize. MGB-db, Wheat. IWIS2		GRU database			Built on passport	International Rice Genebank Collection Information System (IRGCIS)
–Data files and data exchange format/protocols used (indicate for which exchange, where are the data provided)		SINGER dump of Banana records	CIAT has used Excel and CSV formats to send data to SINGER	Part of corporate passport data	Maize & Wheat. Technicians - Collectors					Excel files. We employed one person using arcGIS software, and also geomansa (during the GPG2 project)	Data file type submitted is in MS Access – Upload_IRRI_2010.mdb. Georeference data is included in ACCESSIONS table, format/template was provided by

											SINGER
•For Breeders material = site of evaluation						Minor problem because the number of sites is limited and sites are well known (mainly research stations of which many have their own meteorological data)				None used	
–Format used			Institutional format	ICG	DM	Classic					
–Database(s)			Oracle CIAT database	Under construction, based on African Trial sites by GCP/CIAT and implemented in biomart	Maize. Fieldbook - Maizefinder and IMIS, Wheat. IWIS2	Different but predominantly Excel files					
–Metadata			Institutional format	Admin levels, locality, weather data of growing period, soil data	Structural metadata						
–Exchange format/protocols used (indicate for what)				ICG	Fieldbooks						
–Data files and data exchange format/protocols used (indicate for which exchange, where are the data provided)					IWIN						

Characterization

CHARACTERIZATION	AfricaRice	Bioversity	CIAT	CIP	CIMMYT	ICARDA	ICRAF	ICRISAT	IITA	ILRI	IRRI
Genebank database: if C&E data recorded by genebank											
Format-template used		Characterization: MGIS dump, evaluation: excel file	Institutional template	IPGRI/FAO/AVRDC/CIP descriptor lists for potato, sweetpotato, oca, ulluco for morphological data; otherwise try to conform to tabular data as published by MIBBI	Maize. FieldBooks – MGB, Wheat. FieldBooks – IWIS2	Data from routine C&E experiments stored in consolidated tables for crops. Results of screening for stresses stored in normalized set of tables for all crops. Other, occassional, data stored in separate files connected to database			Data types based on historic data and crop descriptor booklets		see MorphoAgron and Evaluation DataDictionary
Metadata used – data annotation		n/a	Evaluation, person responsible, date of evaluation	Generic standard sheet based on Dublin-core plus various depending on type of C&E data; try always to answer the six classical questions: why, what, who, when, where, how. e.g. using MIAME for gene expression data, MIQAS for molecular data, others .. (see MIBBI.org)	Maize. FieldBooks – MGB, Wheat. FieldBooks – IWIS2	Definition of all descriptors/characters stored in 'table/fields definition files'				data annotation	see MorphoAgron and Evaluation DataDictionary
–Centre's Database(s) and data sources	ARGIS Database on Ms SQL Server 2000	Characterization: MGIS, Evaluation: IMTP	CIAT database	Template based internal archive + biomart	Maize. MGB-DB, Wheat. IWIS2 db – IWIS3 db						IRGCIS (in IRIS not yet complete)

Data files and data exchange formats or protocols used for exchange with GENESYS? Problems and solution	No Data exchange with GENESYS	n/a	GENESYS has not provided a final statement about responsible(s) and date of evaluation	Could be done via simple batch consultation of biomart once data are published.	Maize. Mr F. Nawar's comments (LAMP), Wheat. No problems with data exchange. But some comments about quality data	For GENESYS, batch text files were defined in consultation with GENESYS and replicated to SINGER server. In future, DDC will be used for upload of updates; batch files for first time upload					MS Access table of Morphoagron data from IRGCIS (Oracle database)
How is the database and exchange protocol maintenance done?	Database migration to Ms SQL Server 2008	n/a	The views of database are according to export characterization and evaluation data into CSV format files		At end of season	Scripts for fetching C&E data to GENESYS can be re-used.				Not done; We do not have databases on this. It exists in the centre's own database, analyze data summaries. At present, nothing is available by accession on the web. Evaluation data is available in publications. The row data is all in excel	Morpho agron data is kept in a table in IRGCIS. Data is retrieved for the last Cropyear where the accession was included in the Characteriza tion study.
Breeders' database						As noted for passport, data are kept in variety of ways. The largest sets document International Nurseries trials				None used	
–Format-template used	Excel files used		Institutional template	Same as above plus simple in-house pedigree database for potato and sweetpotato	Maize. FieldBook s – MGB, Wheat. FieldBook s – IWIS2						Study-Variates (Traits,Metho ds,Scale)
–Centre's Database(s) and data sources			Oracle CIAT database		Maize. MGB-DB, Wheat. IWIS2 db – IWIS3 db						IRIS-DMS
–Data files and data exchange formats or protocols used for exchange with SINGER ?					Wheat. No problems with data exchange. But some comments about data						

					quality, only we sent informatio n about internation al trials						
–How is the database and exchange protocol maintenance done?					Daily						No data submitted yet

Conclusions

	Genebank data	Breeding data
AfricaRice	The **rice descriptor** should be revised with more colour images and take into account some particular glaberrima traits	Singer should not handle breeders' data because it is more **complex**
Bioversity	The botanical level called "Section" is not included in the **classification despite the fact that the banana genus needs one.** For banana making a distinction between wild and cultivars is important (i.e Species and Groups) and requires additional fields in the botanical classification part of the MCPD. A change in protocol for exchange is not easy-FAO codes pose problems	
CIAT	Some metadata using the SINGER template must be aggregated or updated into CIAT-GRP schema views. This needs to find a better way to **validate** the information that has been migrated to SINGER and allows to **track changes** or improvements in information	Breeders' information is being migrated to **ICIS** in order to establish the Molecular Breeding Platform. [For more Information: Arturo Franco (a.franco@cgiar.org)]. Consult with the leaders of breeding programmes about the use and benefits of the SINGER, in order to meet their needs to migrate data to SINGER. Much of the breeders' data/material needs to be validated according to non-institutional forms such as SINGER, MCPD, etc. This can take time and involve database curators and database managers
CIP	**Data citation** (visibility of data contributors after so many steps of aggregation/republishing). **Data audit** (who did what to the data when and why): **tools for end-users** e.g. revision history as in source code repositories. Use **versioning of datasets! Data integrity** (how do I know as a contributor that the data is still the same in five years' time or more): MD5, UNF. Huge amounts of longitudinal data coming up: **high resolution GIS surfaces** (at species level). **Next generation sequencing data for genotypes, High resolution images,** What to store and where: **cloud sourcing**: Amazon? Others? Volatility and maintenance costs of digital formats	Data aggregation is highly appreciated by scientists as an added-value both in-house and outside. However, current **backlog on analysis and publication of peer reviewed articles** makes primary data generators reluctant to forego 'low hanging fruits' (CG indicator on publications per institute!):

CIMMYT	Common DB and software for both crops. **GRIN–Global option**	Common DB and software for both crops. **ICIS option**
ICARDA	**More integration with data systems** in other CG genebanks would be desirable to **apply common standards** and to add value to CGIAR collections. **Crop Registers can serve as a model**. **Data standards** need attention (passport and C&E). Define and implement **automated data replication** (to SINGER and others, e.g. GENESYS). Main faults. Too simple /confused documentation of names. No handling of multiple collectors, breeders, etc. – problem with crediting! Solution: **revise MCPD**. Aim at one	**Centre specific** – more integration between genebank and breeding programmes is generally needed
ICRAF	ICRAF requires to set-up live-gene-bank documentation systems to be employed in regions Protocols for greater integration of data from the ICRAF research regions Database management and curation to have more continuity Dedicated computer servers for SINGER data at ICRAF	
ICRISAT		
IITA		
ILRI	**SINGER is not kept updated** (e.g. does not reflect our changes in the last few weeks). Regarding databases for characterization/evaluation data, we should first see what format will be advisable to use and then build up a database with our data! If we put our characterization data on the web, anyone can publish it. We need to publish it first! When publishing, many journals are not fully available, so even these summary data is not freely available. The usage of the warehouse site as the upload portal was a problem so we used tab delimited files and send them as attachments	n.a.
IRRI	Need to work on transfer of IRGCIS georeference data in IRIS which has more or less similar format as the SINGER template. **Modification on the template to handle multiple values** (e.g. collection cooperator, safety storage duplicate, storage type). Make **modification on how data are labeled in the SINGER website**. Should be able to **upload info for new accessions only** (in case no changes were made in the previous data uploaded)	**More effort is needed** if data is to be uploaded to SINGER web, data available is not yet as detailed as genebank materials like passport, georeference, distribution/transfer data. The database is still being enhanced to allow the storage of these data.

Annex 2 - Breakdown of the characterization and evaluation data sent by CGIAR centers to GENESYS in 2010

International Center for Research on the Dry Areas (ICARDA)

Crop	Traits	Accessions	Observations
Barley	30	26175	412733
CHICKPEA	37	11526	189882
FABABEAN	37	4355	142153
LENTIL	43	8585	153914
WHEAT	39	37999	517117
Total		88640	1415799

International Rice Research Institute (IRRI)

Crop	Traits	Accessions	Observations
Rice	30	105667	5180820
Total		105667	5180820

International Institute for Tropical Agriculture (IITA)

Crop	Traits	Accessions	Observations
COWPEA	38	11788	335979
Maize	33	300	9679
Total		12088	345658

International Maize and Wheat Improvement Center (CIMMYT)

Crop	Traits	Accessions	Observations
Barley	10	2921	91227
Maize	5	7277	57643
Wheat	10	10780	904724
Total		20978	1053594

Annex 3 - Centres' report on the collation of germplasm transfer data

Genebank material

DISTRIBUTION OF GENEBANK MATERIAL	AfricaRice	Bioversity	CIAT	CIP	CIMMYT	ICARDA	ICRAF	ICRISAT	IITA	ILRI	IRRI
Acquisition with SMTA, Annex 1 and non-Annex 1											
What are the Center's Database(s) or data sources used to get complete picture?	ARGIS database	Genebank database (MGBMS)	CIAT database and documents of support (LOFA, MOUs)	ADU/SMTA ordering system (in-house)	Maize. FiedlBook + Excel +Acces, Wheat IWIS2 + Excel	GRSDB (main database) – it has distribution module			Excel files, manual reporting	The passport data	International Rice Information System (IRIS). Seed Health Unit's database. International Rice Genebank Collection Information System (IRGCIS)
Was the template version 1.1 used in the last update? How do you rate the use of this template?	Template version 1.0 used	No	No	In general ok	Good	Template version 1.1 was not used in last update. SINGER supplied data (with additional explanation) were used in consultation with Bioversity.				There were no acquisitions done in the last update	Yes, Template v1.1 was used. The template is fine although some fields are difficult to fill up like: 1) Received by centre under SMTA as PGRFA under development with additional conditions (Y/N). 2) Received by centre under SMTA from provider who has previously received the material under SMTA alternative benefit sharing clause 6.11. Need manual checking of documents, which requires more time

What data could not be provided? What data were not properly formatted?	Received by centre under SMTA from provider who has previously received the material under SMTA alternative benefit sharing clause 6.11	SMTA type but now ok	Generally all acquisition data is could be provided(reasons at the previous point)		None					NA	incomplete data on provider institute type, biological status
–How were the data extracted? Per accession or a summary?	summary	accession	accession	accession	Per accession CIMMYT-HQ data, per summary Outreach offices	Database has data extraction functionality (both summary and accession-based modes)				accession	Retrieve all incoming lists under SMTA first, get all the germplasm included in this list, then retrieve information per germplasm/accession
–Data files and data exchange format/protocols used (indicate for which exchange, where are the data provided)		Will be a dump of SINGER DB for banana record	CIAT has used Excel and CSV format to send data to report. data provided are from CIAT GRP Database		SINGER format	MCPD + additional explanations (SMTA, etc.)				NA	data is in MS Access file
–Database manager and database curator?	Tchamba Marimagne	Max Ruas, Ines van den houwe & Els Kempenaers	Yes, one database manager and database curator for each crop (bean, forage, cassava)	Edwin Rojas, (Luis Rojas,Daniel Hiraoka)	DB Manager	GRS and J.Konopka				Alexandra Jorge	Grace Lee S. Capilit (IRGCIS and IRIS).William Eusebio (IRIS). Sally Palmones (SHU and IRIS)

Distribution with SMTA, Annex 1 and non-Annex 1											
–What are the Center's Database(s) or data sources used to get complete picture?	ARGIS database	Genebank database (MGBMS)	CIAT database and documents of support (LOFA, MOUs)	ADU/SMTA database	Maize. Fieldbook IMIS Wheat. IWIS2 IWIS3	GRSDB has Annex 1 flag assigned to each taxa and thus to each accession. 'MLS' flag is assigned to each accession.			Excel files, manual reporting	Datatables (disreq, dispacc, passport, ainvent)	International Rice Information System (IRIS). Seed Health Unit's database. International Rice Genebank Collection Information System (IRGCIS)
–What Format-template was used in the last update?	Template version1.0 used	NA	format suggest by SINGER	Institutional	SINGER format.	Old SINGER format used for transfer of data (accession-based)			Report e-mailed	Provided by SINGER	Template v1.1 was used
–What data could not be provided?		SMTA type but now ok	details of additional conditions.	Some backlog from 'old' requests; meanwhile complemented from linked documents	None					FAO/WIEWS codes	incomplete data on recipient institute type, biological status. not 100% reliable data on "Distributed as PGRFA under development with additional conditions? (Y/N)", info not in the database
–How were the data extracted? Per accession or a summary?	Summary	Will be per accession		accession	Per accession CIMMYT-HQ data, per summary Outreach offices					accession	Retrieve all outgoing lists under SMTA first, get all the germplasm included in this list, then retrieve information per germplasm/accession

–Data files and data exchange format/protocols used (indicate for which exchange, where are the data provided)		Will be a dump of SINGER DB for banana record	CIAT has used Excel and CSV format to send data to report	Institutional	SINGER format	All genebank material distributed with SMTA regardless of 'multilateral status'			XLS Report generated on request from Inventory system	Tab delimited files	data is in MS Access file, since excel can't hold the large number of germplasm/records distributed in a year
–Database manager and database curator?	Tchamba Marimagne	Max Ruas, Ines van den houwe & Els Kempenaers	Yes, one database manager and database curator for each crop (bean, forage, cassava)	Edwin Rojas, (Daniel Hiraoka)	DB Manager					Alexandra Jorge	Grace Lee S. Capilit (IRGCIS and IRIS).William Eusebio (IRIS). Sally Palmones (SHU and IRIS)

Breeders material

Column1	AfricaRice	Bioversity	CIAT	CIP	CIMMYT	ICARDA	ICRAF	ICRISAT	IITA	ILRI	IRRI
Acquisition with SMTA, Annex 1 and non-Annex 1											
–What are the Center's Database(s) or data sources used to get complete picture?	ARGIS database	Genebank database (MGBMS)		ADU/SMTA ordering system (in-house)	Wheat. IWIS2 Maize. Fieldbook + IMIS	Seed Health Lab. (SHL) records show all introductions to ICARDA-summary on incoming seed batches rather than accession-based data			Genebank does not hold breeder's material or data	None	International Rice Information System (IRIS). Seed Health Unit's database. International Rice Genebank Collection Information System (IRGCIS)

–Was the template version 1.1 used in the last update? How do you rate the use of this template?	Template version1.0 used	No		In general ok	Yes	Template 1.1 could not be used (see above)					Yes, Template v1.1 was used. The template is fine although some fields are difficult to fill up like: 1) Received by centre under SMTA as PGRFA under development with additional conditions (Y/N). 2) Received by centre under SMTA from provider who has previously received the material under SMTA alternative benefit sharing clause 6.11. Need manual checking of documents, which requires more time
–What data could not be provided? What data were not properly formatted?	Received by centre under SMTA from provider who has previously received the material under SMTA alternative benefit sharing clause 6.11	SMTA type but now ok	details of additional conditions.		Maize. data from Outreach offices (accesion)	Summary data on acquisitions to crop improvement programs were provided					incomplete data on provider institute type, biological status
–How were the data extracted? Per accession or a summary?	Summary	per accession		accession	Per accession CIMMYT-HQ data, per summary Outreach offices						Retrieve all incoming lists under SMTA first, get all the germplasm included in this list, then retrieve information per germplasm/accession
–Data files and data exchange format/protocols		Will be a dump of SINGER DB			SINGER format						data is in MS Access file

used (indicate for which exchange, where are the data provided)		for banana record									
–Database manager and database curator?	Tchamba Marimagne	Max Ruas, Ines van den houwe & Els Kempenaers	Yes, database manager and database curator for each breeding program	Edwin Rojas, (Luis Rojas,Daniel Hiraoka)	Wheat. DB administrator Maize. Curator	SHL system (+ respective breeders)					Grace Lee S. Capilit (IRGCIS and IRIS).William Eusebio (IRIS). Sally Palmones (SHU and IRIS)
•Distribution with SMTA, Annex 1 and non-Annex 1											
–What are the Center's Database(s) or data sources used to get complete picture?	ARGIS database	Genebank database (MGBMS)	CIAT database and documents of support (LOFA, MOUs)	ADU/SMTA database	Wheat. IWIS2 Maize. Fieldbook + IMIS	Only summary data provided for: International Nurseries and Special Nurseries and other occassional distribution in response to requests			Excel files, manual reporting	None	International Rice Information System (IRIS). Seed Health Unit's database. International Rice Genebank Collection Information System (IRGCIS)
–What Format-template was used in the last update?	Template version1.0 used	NA		Institutional	SINGER format						Template v1.1 was used
–What data could not be provided ?		SMTA type but now ok	details of additional conditions.	Some backlog from 'old' requests; meanwhile complemented from linked documents	Maize. None data per accession from Outreach offices						incomplete data on recipient institute type, biological status. not 100% reliable data on "Distributed as PGRFA under development with additional conditions? (Y/N)", info not in the database

–How were the data extracted? Per accession or a summary?	summary	accession	accession	accession	Per accession (CIMMYT – HQ data)						Retrieve all outgoing lists under SMTA first, get all the germplasm included in this list, then retrieve information per germplasm/accession
–Data files and data exchange format/protocols used (indicate for which exchange, where are the data provided)		Will be a dump of SINGER DB for banana record	Excel files	Institutional	SINGER format						data is in MS Access file, since excel can't hold the large number of germplasm/records distributed in a year
–Database manager and database curator?	Tchamba Marimagne	Max Ruas, Ines van den houwe & Els Kempenaers	Yes, database manager and database curator for each breeding program	Edwin Rojas, (Daniel Hiraoka)	Int, Nurseries Wheat: Dr.Thomas Payne/Efren Rodriguez Int.Nurseries Maize: Dr. Gary Atlin/Efren Rodriguez						Grace Lee S. Capilit (IRGCIS and IRIS).William Eusebio (IRIS). Sally Palmones (SHU and IRIS)

Annex 4 - Data submitted by Centres for the CGIAR report on Germplasm Acquisition and Distribution to the Governing Body IV Meeting

	Genebank Data		Breeding Programme Data	
	Acquisition	Distribution	Acquisition	Distribution
AfricaRice	OK	OK	OK	OK
Bioversity	OK	OK	OK no Breeding Prog.	OK no Breeding Prog.
CIAT	OK	OK	Received most, only beans pending	OK
CIMMYT Maize	Country types missing	OK	OK	2009 data only
CIMMYT Wheat	Country types missing	2009 data only	OK	2009 data only
CIP	OK	OK	OK	OK
ICARDA	OK	OK	OK	OK
ICRAF	OK	OK	OK no Breeding Prog.	OK no Breeding Prog.
ICRISAT	OK	OK	OK	OK
IITA	OK	OK	2009 data only	Not received
ILRI	OK	OK	OK no Breeding Prog.	OK no Breeding Prog.
IRRI	OK	OK	OK	OK

OK: the data were received. Some issues might still be unresolved.
Purple description: Partial data received.
Red description: Data not received.

Annex 5 - Agenda

SINGER workshop for genebank database managers
8-10 December 2010
Hosted by Bioversity International, Rome, Italy
Sakura Room – Ground floor

Agenda (last update: 31 January 2011)

Objective:

Identification of the data types, data standards, technology and agreements needed to achieve seamless data sharing mechanism for genetic resources within CGIAR and provide access to quality, accession-level and system-wide data on the in-trust collections.

Outputs:

- List of agreed data types to be shared
- Data templates and metadata required to be applied
- Revised data dictionary
- Identification of the tools to be used for data sharing, data collation
- Agreement on the periodicity of updates
- Identification of extra efforts in the organization needed at Centres' and SINGER levels.
- Recommendations on SINGER data visibility in GENESYS
- Actions and timeline

Discussions on the identification of the basic elements will cover the following items:
Session 3:

1. Identification of the basic elements for quality data sharing and regular upload
a. Genebank databases, SINGER and GENESYS Data models
b. List of needed Standard data templates
c. Periodicity of data updates/uploads
d. Data repository, quality check, versioning

Session 5:

1. Identification of the basic elements for quality characterization and evaluation data sharing and regular upload considering:
a. Genebank databases, SINGER and GENESYS Data models
b. List of needed Standard data templates, tools and metadata, GENESYS data dictionary
c. Periodicity of data updates/uploads
d. Data repository, quality check, versioning, use of DDC

Session 6:

1. Identification of data sources, basic elements for a data sharing, aggregation and upload mechanisms for germplasm transfer data
a. Genebank databases, breeders or nurseries databases, SINGER and GENESYS Data models
b. Transfer data template version 1.1 and identification of metadata
c. What process at centers' level?
d. Periodicity of data updates/uploads
e. Data repository, quality check, versioning

DAY 1 – WEDNESDAY 8 DECEMBER - Summary of the present system wide situation and lessons learned, identification of elements needed		
8:30-9:00	Welcome and objectives of the workshop Logistics	D. Williams, E. Arnaud A. Chaunac
Session 1	**Current status of Passport Data, georeferences, Collecting missions and c&e in the Centres information systems**	**Facilitator: D. Williams** **Note taking: A. Chaunac**
9:00-10:00	Overview of available data sources, data types and data sharing technologies available or possible in centres as well as issues & constraints for sustainably providing data sets to SINGER and GENESYS *Outline for the presentation will be provided*	10' presentation per Centre- AfricaRice, CIAT, CIP, CIMMYT, ICARDA, ICRAF, ICRISAT, IITA, IRRI, Bioversity, ILRI by videoconference or Skype questions
10:00-10:30	*Coffee Break*	*Cafeteria- Ground floor*
10:30-11:00	Overview from Centres (continued)	10' presentation per Centre
Session 2	**Current system wide actions: SINGER, GENESYS, Crop Registers, Collected samples database**	**Facilitator: E. Arnaud** **Note taking: H. Gaisberger**
11:00-11:15	Current SINGER status in terms of content and web site and SINGER contribution to GENESYS and questions	Milko Skofic
11:15-11:30	Current GENESYS Status in terms of content, web site and questions	Fawzy Nawar
11:30-12:00	Crop registers' role as partners for quality data and pedigree information for SINGER and GENESYS - Discussion with Jan Konopka, Grace Capilit, Alexandra Jorge, Max Ruas, Angela Hernandez	Jan Konopka (Barley registers), Grace Capilit (Rice registers)
12:00-13:00 Lunch discussion	*Knowing the GENESYS and SINGER will become one portal, participants are requested to discuss during lunch time on how centers through the SINGER network can contribute at best to GENESYS and how GENESYS site can provide access at best to SINGER data?*	*Staff room- 1st floor*
Session 3	**Improvement of the Quality of the Passport data and use of the collected sample database**	**Facilitator: M. Ruas** **Note taking: A. Chaunac**
13:00-13:10	Use of the collected-sample database for improving quality of passport data, pedigree data	Hannes Gaisberger/ Federico Mattei

13:10-13:30	Discussion with Grace Capilit, Albert Tchamba, Matija Obreza, Hannes Gaisberger, Federico Mattei	
13:30-14:00	Debriefing and lessons' learned from the last upload of passport data, georeferences and collecting missions data	Milko Skofic and discussion with genebank database managers
14:00-15:00	Identification of the basic elements for quality data sharing and regular upload	See items page 1
15:00-15:30	*Coffee Break*	*Cafeteria- Ground floor*
Session 4	**The germplasm request gateway on SINGER**	**Facilitator: E. Arnaud** **Note taking: A. Chaunac**
15:30-16:15	Presentation of the workflow Discussion on additional features that genebank curators may wish to add to the workflow	Luca Matteis/Elizabeth Arnaud

DAY 2 – THURSDAY 9 DECEMBER - Exchange of Quality data		
8:30-8:45	**Presentation of additional attendees and Summary of DAY 1**	**Facilitator: E. Arnaud**
Session 5	**Characterization and evaluation data**	**Facilitator: M. Mackay** **Note taking: F. Mattei**
8:45-10:20	Debriefing and lessons' learned from the last upload of characterization and evaluation data *Discussion panel with Fawzy Nawar, Matija Obreza (IITA), Efren Rodriguez (CIMMYT) and Jan Konopka (ICARDA)*	Fawzy Nawar
	Identification of the basic elements for quality characterization and evaluation data sharing and regular upload	See items page 1
10:20-11:00	*Coffee break*	*Staff room – 1st floor*
Session 6	**Germplasm Transfer data - Debriefing of recent experiences in compiling and analyzing the distribution data needed for preparing the reports to the Governing Body of the International Treaty on behalf of all the CGIAR Centres**	**Facilitator: T. Hazekamp** **Note taking: S. Lambiase**
11:00-12:00	Preparing the report to the International Treaty Governing Body - Compiling and analyzing genebank and non genebank	Tom Hazekamp, Michael Halewood

	transfer data at the system wide level – presentation	
	Presentation of data sources used , constraints/issues and applied solutions at Centers' level - *Outline to be provided*	5 mns per center
	Discussion with genebank database managers to identify the process for compiling at Centres level and system Wide level	Led by Michael Halewood and Tom Hazekamp
12:00-13:00	*Lunch*	*Staff room-1st floor*
13:00-13:30	Identification of data sources, basic elements for a data sharing, aggregation and upload mechanisms for germplasm transfer data	See items page 1
Session 7	**Data Collation – safeguarding data sets; data upload and data sharing mechanisms**	**Facilitator: R. Simon** **Note taking: M. Ruas**
13:30-14:15	Large data sets upload - presentation of EURISCO upload system for passport data and questions	Milko Skofic/Sonia Dias
14:15-15:00	Regular updates - presentation and practice of Direct Data Control (DDC)	Fawzy Nawar
15:00-15:30	*Coffee break*	*Cafeteria- Ground floor*
15:30-16:00	GRIN-Global Golden candidate – presentation and questions	Michael Mackay
16:00-16:15	Web services and other solutions already in use at centers' level	Matija Obreza
16:15-17:25	Discussion on which are the sustainable existing solutions to address needs at Centers' level, SINGER/ GENESYS level	
20:00	*Social Dinner* *Restaurant 'Il Torchio Sardo' – Via Fabio Numerio, 30/34* *Tel: 06 785757140*	Located 500m from the hotel. Please confirm your presence no later than 10am Thursday

DAY 3– FRIDAY 10 DECEMBER – Expanding and strengthening the system wide collaboration		
Session 8	**Expanding the System-wide data standards to molecular data**	**Chair: Max Ruas** **Note taking: A. Chaunac**
8:30-8:45	Summary DAY 2	Elizabeth Arnaud
8:30-9:00	Expending the range of shared data set to molecular data: Identification of the sources and practices at centers' level – example of the GCP project ' Establishing a Genetic Resource Support Service (GRSS) for the plant breeding community', Crop ontology	Elizabeth Arnaud
Session 9	**Data publishing, annotation, citation**	**Facilitator: S. Reinhard** **Note taking: A. Chaunac/F. Mattei**
9:00-9:15	Example of the Repository of the Pdf files of collecting forms and collecting mission reports	Massimo Buonaiuto with Grace Capilit
9:15-9:30	Example of the GCP central Registry	Milko Skofic / Elizabeth Arnaud
9:30-10:00	Data attribution and data citation: Practices in attributing metadata to data sets – Center level, SINGER level, GENESYS level	Simon Reinhard
10:00-10:30	*Coffee break*	*Cafeteria – Ground floor*
Session 10	**Infrastructure and collaborative tools**	**Facilitator: A. Pastore** **Note taking: M. Buonaiuto**
10:30-12:00	What infrastructure and collaborative tools are needed to support the system-wide informatics activities in terms of community development, knowledge sharing and outreach? *Discussion and selection of tools and infrastructure type according to the needs*	Nadia Manning Videoconference
12:00-13:00	*Lunch*	*Staff room – 1st floor*
Session 11	**Summary, conclusions & agreements**	**Facilitator: D. Williams** **Note taking: F. Mattei**
13:00-14:00	What data will SINGER collate and how? Summary of suggestions from the board and discussion on	Elizabeth Arnaud

	how SINGER can contribute at best to GENESYS and how GENESYS site can provide access at best SINGER data ? How to Keep SINGER identity among the other existing system?	
14:00-15:00	What kind of centre-level and system-wide team organization is needed to support the data sharing process? Definition of roles	
15:00-15:30	*Coffee break*	*Cafeteria – Ground floor*
15:30-16:45	Summary and validation of the elements to apply for a system-wide quality data sharing process	
16:45-17:00	Conclusions and closure	D. Williams

Annex 6 - List of participants

SINGER workshop for the genebank database managers

8-10 December 2010

Bioversity International, Maccarese, Rome, Italy

List of participants (updated 31 January 2011)

	Centre / Project	Participant contact details	EMAIL
1.	**Africa Rice Center/ SINGER IT focal point**	**Marimagne TCHAMBA**, Data Manager, WARDA/ADRAO 01 BP 2031, Cotonou, Benin, Tel: (229)21350188 Fax: (229)21350556	ATchamba@CGIAR.ORG
2.	**Bioversity/ SINGER IT focal point**	**Max RUAS**, Information System Analyst, Bioversity France, Parc Scientifique Agropolis II, 34397 Montpellier Cedex 5, France,Tel: +33 (467) 61 9939	m.ruas@cgiar.org
3.	**CIAT/ SINGER IT focal point**	**Angela HERNANDEZ,** System Analyst, Centro Internacional de Agricultura Tropical (CIAT), Apartado Aéreo 6713, Cali, Colombia, Tel: (57-2) 445-0000 Extension 3617	a.hernandez@cgiar.org
4.	**CIMMYT / SINGER IT focal point**	**Efren RODRIGUEZ**, Head of Data Processing and Seed Distribution, CIMMYT (Centro Internacional de Mejoramiento de Maíz y Trigo), Km. 45 Carretera Mexico Veracruz, El Batan Texcoco Estado de Mexico, C.P. 56130, Tel: +52(55) 58042004, Fax: +52 (55) 5804 7558	e.rodriguez@cgiar.org
5.	**CIP/ SINGER IT focal point**	**Reinhard SIMON**, Centro Internacional de la Papa (CIP), Apartado 1558, Lima 12, Peru, Tel: +51 (1) 3496017 x3025	r.simon@cgiar.org
6.	**ICARDA/ SINGER IT focal point**	**Jan KONOPKA**, Germplasm Documentation Officer, Genetic Resources Section (GRS), Biodiversity and Integrated Gene Management Program (BIGM), P.O. Box 5466, Aleppo, Syria, Tel: +963-21 29612682 Fax: +963-21 2213490	j.konopka@cgiar.org

	Centre / Project	Participant contact details	EMAIL
7.	**ICRAF/ SINGER IT focal point**	**Calleb ORWA**, Database manager, World Agroforestry Centre, Box 30677-00100, United Nations Avenue, GIGIRI, Nairobi, Kenya	C.ORWA@CGIAR.ORG
8.	**ICRISAT/ SINGER IT focal point**	**M THIMMA REDDY**, Documentation Officer, Genetic Resources, International Crops Research Institute for the Semi Arid Tropics (ICRISAT), Patancheru 502 324, Andhra Pradesh, India, Tel : +91 40 30713581(Office), +91 9949665440 (Mobile), Fax : +91 40 30713074	t.reddy@cgiar.org
9.	**IITA/ SINGER IT focal point**	**Matija OBREZA**, Software development manager, International Institute of Tropical Agriculture (IITA), PMB 5320, Ibadan, Oyo State, Nigeria, Tel: +234 2 7517472	m.obreza@cgiar.org
10.	**IRRI/ SINGER IT focal point**	**Grace CAPILIT**, Senior Specialist – Database Administration, TT Chang Genetic Resources Center, International Rice Research Institute (IRRI), Los Baños, Laguna, Tel: 845-0563, +63(049)536-2701 to 05 local 2368	G.Capilit@cgiar.org
11.	**Bioversity/ SINGER coordinator**	**Elizabeth ARNAUD**, SINGER Coordinator, Bioversity International, Via dei Tre Denari, 472/a, 00057 Maccarese, Rome, Italy, Tel: +39 066118 323, Fax: +39 0661979661	e.arnaud@cgiar.org
12.	**Bioversity / EURISCO**	**Sonia DIAS**, EURISCO Coordinator, Bioversity International, Via dei Tre Denari, 472/a, 00057 Maccarese, Rome, Italy, Tel: +39 066118 204, Fax: +39 0661979661	s.dias@cgiar.org
13.	**Bioversity / GENESYS**	**Michael MACKAY**, Principal Investigator, Bioversity International, Via dei Tre Denari, 472/a, 00057 Maccarese, Rome, Italy, Tel: +39 066118 235, Fax: +39 0661979661	m.mackay@cgiar.org
14.	**Bioversity / GENESYS**	**Fawzi NAWAR**, Java Web Developer, Bioversity International, Via dei Tre Denari, 472/a, 00057 Maccarese, Rome, Italy, Tel: +39 066118 321, Fax: +39 0661979661	f.nawar@cgiar.org
15.	**Bioversity / GPG2**	**Massimo BUONAIUTO**, Multimedia/Web Specialist, Bioversity International, Via dei Tre Denari, 472/a, 00057 Maccarese, Rome, Italy, Tel: +39 066118 406, Fax: +39 0661979661	m.buonaiuto@cgiar.org

	Centre / Project	Participant contact details	EMAIL
16.	**Bioversity / SINGER technical coordinator**	**Milko SKOFIC**, Database Programmer, Bioversity International, Via dei Tre Denari, 472/a, 00057 Maccarese, Rome, Italy, Tel: +39 066118 286, Fax: +39 0661979661	m.skofic@cgiar.org
17.	**Bioversity/ GPG2**	**Hannes GAISBERGER**, Consultant, Bioversity International, Via dei Tre Denari, 472/a, 00057 Maccarese, Rome, Italy, Tel: +39 066118 244, Fax: +39 0661979661	h.gaisberger@cgiar.org
18.	**Bioversity/ GPG2**	**Federico MATTEI**, Consultant, Bioversity International, Via dei Tre Denari, 472/a, 00057 Maccarese, Rome, Italy, Tel: +39 066118 468, Fax: +39 0661979661	f.mattei@cgiar.org
19.	**Bioversity/ Policy & Law**	**Michael HALEWOOD**, Head, Policy & Law, Bioversity International, Via dei Tre Denari, 472/a, 00057 Maccarese, Rome, Italy, Tel: +39 066118 294, Fax: +39 0661979661	m.halewood@cgiar.org
20.	**Bioversity/ SGRP**	**David E. WILLIAMS**, SGRP Coordinator, CGIAR System-wide Genetic Resources Programme (SGRP), c/o Bioversity International, Via dei Tre Denari 472/a, 00057 Maccarese, Rome, Italy, Tel:+39 066118 225, Fax: +39 0661979661	d.williams@cgiar.org
21.	**Global Crop Diversity Trust/**	**Luigi GUARINO**, Senior Science Coordinator, Global Crop Diversity Trust, c/o FAO, Viale delle Terme di Caracalla 00153 Rome, Italy, Tel: +39 06 570 55142, Fax: +39 06 570 55634	luigi.guarino@croptrust.org
22.	**ICT-KM**	**Antonella PASTORE**, Project Coordinator, ICT-KM, Bioversity International, Via dei Tre Denari, 472/a, 00057 Maccarese, Rome, Italy, Tel: +39 066118 314, Fax: +39 0661979661	a.pastore@cgiar.org
23.	**SINGER consultant**	**Thomas HAZEKAMP**, Consultant, c/o Bioversity International, Via dei Tre Denari, 472/a, 00057 Maccarese, Rome, Italy	t.hazekamp@cgiar.org
24.	**SINGER consultant**	**Luca MATTEIS**, Web Developer, Bioversity International, Via dei Tre Denari, 472/a, 00057 Maccarese, Rome, Italy, Tel: +39 066118 351, Fax: +39 0661979661	l.matteis@cgiar.org

Observers: Adriana Alercia, John Michael, Imke Thormann

Annex 7 - List of Abbreviations/Acronyms

A

AfricaRice – Africa Rice Center
ARGIS – AfricaRice Germplasm Information System

B

Bioversity – Bioversity International

C

CBD – Convention in Biodiversity
CGIAR - Consultative Group on International Agricultural Research
C&E – Characterization and Evaluation data
CIAT – Centro Internacional de Agricultura Tropical/International Centre for Tropical Agriculture
CIMMYT – Centro Internacional de Mejoramiento de Maíze y Trigo/International Maize and Wheat Improvement Center
CIP – Centro Internacional de la Papa/International Potato Center
CO – Crop Ontology

D

DDC – Direct Data Control

E

EURISCO - European Plant Genetic Resources Search Catalogue

F

FAO - Food and Agriculture Organization (of the United Nations)

G

GB – Governing Body of the Treaty
GCP – Generation Challenge Program
GENESYS – Gateway to Genetic Resources
GPG2 - Global Public Goods Project Phase 2
GRIN – Genetic Resources Information Network

I

IBPGR/IPGRI – International Board for Plant Genetic Resources/International Plant Genetic Resources Institute
ICARDA – International Center for Agricultural Research in the Dry Areas
ICG – ICG – International Cooperators Guide for potato
ICIS – International Crop Information System
ICRAF – See World Agroforestry Centre
ICRISAT - International Crops Research Institute for the Semi-Arid Tropics
ICT-KM - Information and Communications Technology and Knowledge Management

ICWG-GR - Inter-Centre Working Group on Genetic Resources
IITA – International Institute of Tropical Agriculture
ILRI – International Livestock Research Institute
MGBMS – Musa Genebank Management Syetem
IMIS – International Maize Information System
IRGCIS – International Rice Genebank Collection Information System
IRIS – International Rice Information System
IRRI – International Rice Research Institute
IWIS – International Wheat Information System

M

MCPD – Multi-Crop Passport Descriptors
MGIS - Musa Germplasm Information System

N

NARS/ARIS - National Agricultural Research Systems/ Advanced Research Institutions

O

OECD - Organization for Economic Co-operation and Development

P

PaD – Passport data
PID – Personal IDentifier

R

RSS - Really Simple Syndication

S

SGRP - System-wide Genetic Resources Programme
SINGER – System-wide Information Network for Genetic Resources
SMTA – Standard Material Transfer Agreement

T

The Treaty- the International Treaty on Plant Genetic Resources of Food and Agriculture

U

UNICC - United Nations International Computing Centre

W

WARDA – See AfricaRice
WebDAV – Web-based Distributed Authoring and Versioning
World Agroforestry Centre – formerly International Council for Research in Agroforestry (ICRAF)

Annex 8 – Recommendation of the SINGER Task Force meeting, June 2010

Recommendation 1: SINGER and Genesys will share the same database. SINGER information management will be handled by Genesys; consequently, the database function will be lost and taken over by Genesys.

Recommendation 2: The future of the SINGER website needs to be further discussed by the Task Force and a way forward agreed, particularly for the transition period while Genesys is getting up and running.

Recommendation 3: The recommendations relating to the cost-benefit analysis for adopting GRIN-Global, the data attribution proposal and governance issues are still valid and should be considered by the SINGER Task Force in the ongoing implementation of the network.

Recommendation 4: Both cross-referencing tools mentioned above should be made available to the crop networks, acknowledging that expert validation is required in the process. The tools could be inserted into the Crop Genebank Knowledge Base.

Recommendation 5: The importance of pedigree information has been once again stressed to identify the parent of sample. There is a need for information on neighbourhood/duplicate/parental trees to be included in Genesys.

Recommendation 6: There is a need for Bioversity to promote only one system and to provide SINGER with a system like that of EURISCO that produces quality reports. No concrete decision was made in this regard and it was recommended that the Task Force discuss this issue in a separate meeting with the genebanks' database managers**.**

Recommendation 7: An additional chapter should be added to the CGKB on data management and the upload mechanism could also be described here. The Generation Challenge Programme (GCP) would certainly be ready to publish their methodologies, such as the tools for crop registries, genotyping and phenotyping protocols or guidelines for core collections. An updated manual for collecting could be loaded on the CGKB. Increased awareness about this product is needed.

Recommendation 8: It was noted that pedigree management systems would serve as a key element for the integration of the GR management system and the IBP. It might also be further developed by GRIN-Global in Phase II. This could also be an additional proposal for the Gates Foundation, as both projects are currently funded this donor.

Recommendation 9: The Task Force needs to list all information components that already exist and outline the elements currently missing in order to produce a revised schema based on the one developed by Ruaraidh for the SINGER consultation meeting.

Recommendation 10: Termination of collective actions would constitute a step backwards, and it must be put into perspective considering the new information needs of the world. One key action is to raise awareness among ICWG-GR and the traditional SINGER audience and donors about the newly named portal Genesys (collectively developed). Genetic resources activities should be balanced against the breeding approach of MPs. The SINGER Task Force needs to demonstrate the advantage of global access to germplasm information in comparison to single, independent genebank databases.

Recommendation 11: The Task Force and the SINGER network members should provide key talking points and agree on a strong message to collectively convey when approached by consultants of the scoping study. We could combine the ISC vision (Attached in Annex 2) with the Task Force recommendation.