Task Group on Data fitness for Use for Agrobiodiversity

Proposed Use cases

Use case A: Find gaps in conservation in crop wild relatives - proposed by Nora P. Castañeda-Álvarez

1. Describe the objective

Background: Plant genetic resources are the building blocks of agriculture. Crop wild relatives, as a class of plant genetic resources, are used in plant breeding to produce varieties containing novel traits, such as improved nutritional quality, and tolerance to biotic and abiotic stresses. Plant breeders and researchers obtain samples of crop wild relatives from genebanks, however initial assessments have detected that one in ten accessions correspond to wild relative taxa (FAO, 2010). Sympatry of cultivated and related wild forms is also often observed in centres of origin of domesticated plants (Jarvis & Hodgkin 1999). Such a situation is very favourable to genetic exchange between individuals displaying different phenotypes, and may therefore lead to the production of new genotypes and phenotypes on which human and natural selection can operate. (Elias et al.,Molecular Ecology (2001) 10, 1895 –1907). In addition, threats to biodiversity like habitat destruction, invasive species and climate change are expected to hamper the survival of crop wild relatives in their natural habitats (Jarvis *et al*., 2008, Brummit *et al*., 2015).

Objectives: Given the previous background, a study to detect the extent of representativeness of crop wild relatives in genebanks, establish the ex situ conservation priorities of the crop wild relatives analyzed, and understand the patterns of global richness where future field collections could be conducted.

2. Who are the actors ?

Data managers, modelers, genebank curators, herbaria curators, sample collectors, breeders

3. Data/information products to be produced

In order to be able to achieve these objectives, it was necessary to produce potential distribution models, and to apply a gap analysis methodology to assess the sufficiency of CWR samples in genebanks (Ramirez-Villegas *et al*., 2010).

4. Data sources the most used

The analyses conducted heavily relied in geographic explicit information, thus, occurrence records obtained from GBIF.org, germplasm mobilization platforms [e.g., the CGIAR’s System-wide Information Network for Genetic Resources (SINGER), the European Plant Genetic Resources Catalogue (EURISCO), and the United States Department of Agriculture’s Genetic Resources Information Network (GRIN), mainly], data provided by researchers, peer-reviewed and gray literature, and herbaria.

5. Tools the most used

- GRIN Taxonomy (http://www.ars-grin.gov/~sbmljw/cgi-bin/taxcrop.pl?language=en-) and the Harlan and De Wet Crop Wild Relative Inventory (http://www.cwrdiversity.org/checklist/) were largely used to identify the relationship between crop wild relatives and their associated crops, confirmed and potential uses in breeding, and the native distributions of crop wild relatives.
- An API to GRIN Taxonomy, the Taxonomic Name Resolution Service (http://tnrs.iplantcollaborative.org/), and TaxonStand (Cayuela *et al*., 2012) were used as references to identify possible misspellings in the taxonomic names of crop wild relatives, and to standardize them.
- GEOLocate and the Google Maps Geocode API were used to georeference records.
- MaxEnt was used as the algorithm to produce potential distribution maps (Phillips *et al*., 2006).
- An in-house tool was designed to process batches of occurrence records in the

following order: 1) standardize fields, 2) check and standardize taxonomic names, 3) check and re-calculate geographic coordinates, 4) final standardization. The tool consists of a stand-alone application prepared in java and available here: www.github.com/CIAT-DAPA/cwr_occurrencesvalidation

- In terms of the gap analysis, the code used works with R and is adapted to analyze a single taxon or a complete crop genepool (https://github.com/npcastaneda/gap-analysis-maxent).

6. Describe the data flow into steps ("to be able to...") indicating who is doing the indicated step (a scientist, a developer, the system?)

To be able to …

7. Identify GBIF role and improvements

i. status of data (quality, coverage)

Major gaps in occurrence records are persistent, especially in countries like DRC, China, Russia, Argentina, and regions like the Amazon and central Africa. Improved access to occurrence records of crop wild relatives in such regions will help to detect populations that might not be represented yet in genebanks (and therefore require further conservation actions).

ii. Additional attributes in demand

links to CWR checklists (e.g., the Harlan and De Wet Crop Wild Relative Inventory, GRIN Taxonomy, and Mansfeld's World Database of Agricultural and Horticultural Crops -http://mansfeld.ipk-gatersleben.de/apex/f?p=185:3::::::-)

iii. Additional sources - > what are the possible connectors between GBIF and the sources

- Genesys
- European Plant Genetic Resources Catalogue (EURISCO)
- United States Department of Agriculture's Genetic Resources Information Network (GRIN)

iv. What data mobilization is needed? By whom?

- other initiatives (e.g., GRIN global, Genesys) support the digitization and mobilization of passport data associated to germplasm accessions. It has been estimated that less than a third of the existing genebanks in the world had organized their collections and made their passport data digitally and openly available to the public. Establishing a close collaboration between GBIF and these initiatives (GRIN global and Genesys) will increase the visibility of plant genetic resources occurrence data (including crop wild relatives) via GBIF.

- The project “Adapting agriculture to climate change: collecting, conserving and preparing the crop wild relatives” has produced a CWR occurrence database, including information not yet mobilized through GBIF (i.e., researchers archives, gray literature). This information can also help to improve the completeness of CWR data mobilized by GBIF.

→ Is this a priority? if not implemented how will it block the progress?

Link to recommendations #

6.2.1, 6.2.2, 6.2.3, 6.3.1, 6.3.2, 6.3.3, 6.6.1, 6.6.2, 6.6.3, 6.6.4, 6.6.7, 6.7.2, 6.8.1, 6.8.2, 6.8.3, 6.8.4, 6.9.1, 6.9.2, 6.9.3, 6.10.1, 6.10.2, 6.11.1, 6.11.2, 6.11.3, 6.11.5, 6.12.3, 6.14.1

Bibliographic references

Brummitt, N. A., Bachman, S. P., Griffiths-Lee, J., Lutz, M., Moat, J. F., Farjon, A., … Nic Lughadha, E. M. (2015). Green plants in the red: a baseline global assessment for the IUCN Sampled Red List Index for plants. Plos One, 10(8), e0135152. doi:10.1371/journal.pone.0135152

Cayuela, L., Granzow-de la Cerda, Í., Albuquerque, F. S., & Golicher, D. J. (2012). taxonstand: An R package for species names standardisation in vegetation databases. Methods in Ecology and Evolution, 3(6), 1078–1083. doi:10.1111/j.2041-210X.2012.00232.x

FAO. (2010). The second report on the state of the world’s plant genetic resources for food and agriculture. Rome, Italy: Commission on Genetic Resources for Food and Agriculture, Food and Agriculture Organization of the United Nations.

Jarvis, A., Lane, A., & Hijmans, R. J. (2008). The effect of climate change on crop wild relatives. Agriculture, Ecosystems & Environment, 126, 13–23. doi:10.1016/j.agee.2008.01.013

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. Ecological Modelling, 190(3-4), 231–259. doi:10.1016/j.ecolmodel.2005.03.026

Ramírez-Villegas, J., Khoury, C., Jarvis, A., Debouck, D. G., & Guarino, L. (2010). A gap analysis methodology for collecting crop genepools: a case study with Phaseolus beans. PloS One, 5(10), e13497. doi:10.1371/journal.pone.0013497

Use case B: Conservation plan for Crop Wild Relatives - by Dag Terje Endresen

1. Describe the objective

Background: Genetic diversity from crop wild relatives (CWRs) can be exploited as gene donors to provide needed new properties for food crops emerging from new climate regimes, demand for more food to feed growing world population, to keep up in arms race with evolving crop pests, etc. Novel molecular breeding and genetic modification methods allow for genetic diversity and alleles to be transferred between crops of different species and open for more effective use of the CWRs as a genetic resource.

Objective: Develop and implement a conservation plan for CWR (national --> regional --> global). Start with a national CWR conservation plan, collaborate with neighbouring countries to build a regional CWR conservation plan. A global CWR conservation plan based on both a global assessment and available national and regional CWR conservation plans.

2. Who are the actors ?

CWR expert team to develop a national list of priority CWR taxa. National protected area managers to include selected priority CWR taxa into their monitoring plans. This activity will gather the required information to support conservation policy decisions, including assigning designated CWR conservation populations together with the national CWR expert team. Students and researchers to collect molecular evidence on genetic diversity in and between the designated CWR conservation populations. This information will support conservation policy decisions (strengthen when unique genetic diversity, weaken when genetic diversity is similar across designated CWR populations in different locations).

3. Data/information products to be produced

- Global checklist of CWR, what species are classified as CWR (same genus as a crop)
- National checklist of CWR, national conservation status, national conservation priority
- Occurrences for nationally designated conservation populations of CWRs
- Connecting monitoring data on CWRs collected from national activities
- Abundance and quantities to verify that the CWR designated conservation population is ok
- Connecting genetic monitoring data to verify CWR designated population is ok

4. Data sources the most used

- Taxon checklists, GRIN Taxonomy, Mansfeld's World Database (for CWR checklist)
- GBIF occurrence data (for Distribution Modelling, SDM)
- Environment data from WorldClim and the Norwegian mapping agency (for SDM)
- External data include spatial extents for protected areas in the country
- A global checklist of crop species (that could have CWRs in the country)
- Data on the economic and other value for society for the crops that related to the CWRs

5. Tools the most used

- Species distribution modeling software (Maxent, R)

6. Describe the data flow into steps ("to be able to...") indicating who is doing the indicated step (a scientist, a developer, the system?)

- To be able to develop a national CWR checklist first the ABD crops and respective species need to be identified. The Mansfeld's world database of crop plants have a comprehensive list of crops.
- Next the CWRs in a country are identified. The general rule of species belonging to the same genus as a crop species can be used.
- A national checklist can be used to extract all species of the same genus as a crop on a global crop species checklist information resource.
- The next step is to identify nationally prioritized CWR species. We need to develop a set of prioritization criteria.
- Data on the economic and other value for society for the crops that are related to the CWRs can provide important input to the prioritization of CWR checklist species.
- Direct economic or societal value of the CWR species itself can provide another prioritization criteria.
- With a national prioritized CWR checklist the conservation strategy can be developed.
- One objective is to ensure *in situ* conservation for designated conservation populations.
- First step is to identify prioritized and stable CWR populations within existing protected areas. Comprehensive occurrence data from GBIF can be used.
- Without comprehensive occurrence data on CWRs species distribution modeling can be used as a predictive tool. Species predicted to occur inside protected areas must of course be verified by an expert to be present before any designated conservation populations are decided.
- If a prioritized CWR species is not conserved within any existing protected areas, then it is possible to start to gather evidence for proposing modification of the protected areas.
- However, more often threatened prioritized CWR species should anyway be

collected for *ex situ* conservation to avoid them to be lost.
- When designated CWR conservation populations are decided, an *ex situ* backup copy is warranted both as part of the conservation goal and to provide more rapid and easier access to this genetic resource for use and research purposes.

7. Identify GBIF role and improvements

i. status of data (quality, coverage)

ii. Additional attributes in demand

- Indexing CWR and crop species status into the GBIF portal to allow for these attributes to be used in user initiated search.
- GBIF portal to implement solutions to identify and link occurrence level information published from different data owners in different datasets.

iii. Additional sources - > what are the possible connectors between GBIF and the sources

iv. What data mobilization is needed? By whom?

- Mobilisation of taxon checklist information resources for CWRs and ABD crops.
- Mobilisation of occurrence data for CWR species, gaps analysis and strategies for filling gaps for these species

Recommendation: Publish in GBIF national, regional, and global CWR checklist with conservation priority assessment status. GBIF nodes to assist national CWR expert to publish checklists and monitoring data for designated CWR populations. Provide and USE persistent identifiers that identify the designated CWR populations (new term needed?) and conservation sites (dwc:locationID)

→ Is this a priority? if not implemented how will it block the progress?

Link to recommendations #

6.2.1 MCPD to be indexed by the GBIF portal
6.2.3 Extension with attributes to describe CWR species
6.2.5 Agrobiodiversity community governance for the Darwin Core germplasm extension, including a process to cover in situ CWR
6.3.1 CWR species checklist will for the starting point for national CWR conservation strategies
6.3.2 Taxon level extensions for CWRs to be indexed by the GBIF portal taxon backbone
6.3.3 Training for nodes on CWRs to participate in CWR data mobilization and use

6.6.1 The global CWR occurrence dataset will guide the national CWR conservation strategy
6.7.1 Training for nodes on CWRs to participate in CWR data mobilization and use
6.8.1 In particular taxon backbone information on CWRs would be useful
6.8.2 Georeferencing data cleaning tools and services embedded in the GBIF portal
6.8.4 Taxon names cleaning tools and ABD checklist resources in the portal
6.9.1 Improved tools to identify occurrence level duplicates will improve modeling prerequisites
6.9.2 Solutions and methods to allow refined versions of CWR occurrences into the portal
6.9.3 Cross-linking duplicated occurrence level information between datasets
6.14.1 Combining occurrence data with other data

Recommendation: An extension to the GBIF Taxon core for CWR attributes/information. Minor additions to the taxon data models implemented by GBIF will enable data exchange of attributes to support CWR conservation by using the existing GBIF infrastructure - replacing the need for a new and parallel data flow mechanism.

- Perhaps simply the taxon extension for description might provide sufficient functionality(?): http://rs.gbif.org/extension/gbif/1.0/description.xml
- The species profile taxon extension also provide similar functionality, that could be explored: http://rs.gbif.org/extension/gbif/1.0/speciesprofile.xml

Bibliographic references

Asdal Å, Philips J, and Maxted N (2013). Boost for crop wild relative conservation in Norway. *Crop wild relative* 9:20-21.

Brehm JM, Maxted N, Martins-Loução MA, and Ford-Lloyd BV (2010). New approaches for establishing conservation priorities for socio-economically important plant species. *Biodiversity and Conservation* 19(9): 2715-2740. doi:10.1007/s10531-010-9871-4

Fielder H, Brotherton P, Hosking J, Hopkins JJ, Ford-Lloyd B, and Maxted N (2015). Enhancing the Conservation of Crop Wild Relatives in England. *PLoS ONE* 10(6): e0130804. doi:10.1371/journal.pone.0130804

Harlan JR, and de Wet JMJ (1971). Towards a rational classification of cultivated plants. *Taxon* 20(4): 509–517. doi:10.2307/1218252

Iriondo JM, Maxted N, and Dulloo ME (eds.) (2008). Conserving Plant Genetic Diversity in Protected Areas: Population Management of Crop Wild Relatives. Wallingford, UK: CAB International. ISBN 9781845932824. doi:10.1079/9781845932824.0000

Maxted N, Ford-Lloyd BV, and Hawkes JG (eds.) (1997). Plant Genetic Conservation. The *in situ* approach. Chapman and Hall, London, UK. ISBN 978-94-009-1437-7. doi:10.1007/978-94-009-1437-7

Maxted N, Scholten M, Codd R, and Ford-Lloyd B (2007). Creation and use of a national inventory of crop wild relatives. *Biological Conservation* 140(1-2): 142-159. doi:10.1016/j.biocon.2007.08.006

Maxted N, Dulloo E, Ford- Lloyd BV, Iriondo JM, and Jarvis A (2008). Gap Analysis: A tool for complementary genetic conservation assessment. *Diversity and Distributions* 14(6): 1018-1030. doi:10.1111/j.1472-4642.2008.00512.x

Maxted N, and Kell S (2009). Establishment of a global network for the in situ conservation of CWR: Status and needs. Background study paper no. 39. FAO Commission on Genetic Resource for Food and Agriculture, Rome, Italy. Available online at http://www.fao.org/docrep/013/i1500e/i1500e18a.pdf

Maxted N, Kell S, Ford- Lloyd B, and Toledo Á (2012). Towards the systematic conservation of global crop wild relative diversity. *Crop Science* 52(2): 1-12. doi:10.2135/cropsci2011.08.0415

Maxted N, Dulloo ME, Ford-Lloyd BV, Frese L, Iriondo JM, and Pinheiro de Carvalho MAA (eds.) (2012). Agrobiodiversity conservation: Securing the diversity of crop wild relatives and landraces. Pp. 72-77. CAB International, Wallingford, UK. ISBN 9781845938512. doi:10.1079/9781845938512.0000

Maxted N, Brehm JM, and Kell S (2013). Resource book for the preparation of national conservation plans for crop wild relatives and landraces. University of Birmingham, UK. Available at http://www.fao.org/fileadmin/templates/agphome/documents/PGR/PubPGR/ResourceBook/TEXT_ALL_2511.pdf

Phillips J, Kyratzis A, Christoudoulou C, Kell S, and Maxted N (2014). Development of a national crop wild relative conservation strategy for Cyprus. *Genetic Resources and Crop Evolution* 61(4): 817-827. doi:10.1007/s10722-013-0076

Vincent H, Wiersema J, Kell S, Fielder H, Dobbie S, Castañeda-Álvarez NP, Guarino L, Eastwood R, Leon B, and Maxted N (2013). A prioritized crop wild relative inventory to help underpin global food security. *Biological Conservation* 167: 265–275.

Use case C: Find agrobiodiversity hotspots and monitor changes for decision on conservation - proposed by Yves Vigouroux

1. Describe the objective

Humanity relies on a few crops for food supply. Genetic diversity of these crop and their wild relatives is a key asset for adaptation of agriculture to future pest and climate conditions. We still do not know enough about the genetic diversity of our crop and wild relatives, how this diversity built up across times, how it evolves and changes. Understanding pattern of diversity and evolution of this diversity is a key information to develop informed conservation strategy.

The objectives are to identify the hot spots of agrobiodiversity, understand how these hot-spot appears and changes across times. The final aim is to develop conservation strategy of agrobiodiversity in situ.

2. Who are the actors ?

Actors are researchers at the descriptive level and local community with the assistance of national conservation services (government, NGOs, Universities and other research centers…) for preservation of specific areas. The analysis of the origin of the hotspots is also an important research question: co-occurrence with wild relatives, historical origin of crop and known diffusion path, human/farmer cultural diversity**.**

3. Data/information products to be produced

- Map of agrobiodiversity across wild relatives and cultivated species and main factors explaining such pattern
- Map of agrobiodiversity changes across times (few or several decades)

4. Data sources the most used

Occurrence data (GBIF, GENESYS, CGIAR genbank, …) and associated fields mainly latitude/longitude and varietal names

Environmental database (environmental data like BIOCLIM current, past and future climate; soil data, ..) and using plant names with statistical database (FAO stats for estimation of cultivated areas, yields with the possibility to use temporal variation of these datasets)

Genetic/genomics depository system (NCBI Genbank, Gramene, DRYAD, ..) or plant specific system (MaizeDB, Coffee Genome hub, Rice databases, …)

Cultural datasets (anthropological, linguistic map, …)

5. Tools the most used

- Genetic diversity analyses tools i.e. structure, estimation of diversity
- Niche modeling (Maxent)
- Maps of hotspots of agrobiodiversity with a high resolution

6. Describe the data flow into steps ("to be able to...") indicating who is doing the indicated

step (a scientist, a developer, the system?)

a. To be able to identify the different species occurrences, we need :

a) occurrences with geographical coordinates with validated data for the occurrence location (latitude/longitude and village/country) and date of sampling

b) completeness of the occurrence of the different species, with valid species names

b. To be able to link between database /datasets, accessions names (fields) need to allow connection with

a) Phenotypic variation of the studied species (CGIAR Genebank or other data sources where phenotypic data might be available);

b) Genetic/genomic diversity work (NCBI Genbank, Gramene, DRYAD, .. or plant specific system MaizeDB, Coffee Genome hub, Rice databases, …).

The connection could be at the level of the accession name/occurrence name (ideal) or could be from closely geographic samples (less optimal as statistical inference is needed based on latitude/longitude).

c. To be able to retrieve environmental dataset (environmental data like BIOCLIM current, past and future climate; soil data...) and with cultural datasets (anthropological, linguistic map), a good and checked latitude and longitude dataset is needed.

d. To be able to monitor the temporal change in agrobiodiversity, supplementary information in crop like varietal names and date of sampling (precision day) is needed. Annual update of status is only necessary.

7. Identify GBIF role and improvements

i. Status of data (quality, coverage)

1) Quality of geo-reference datasets (latitude/longitude coordinates need to be checked) is crucial and good coverage is also important

2) Quality in species determination i.e. up-to-date species name and new fields like name of the variety.

3) Ability to link GBIF occurrence with the same occurrence found in other database (CGIAR Genebank or other data sources where phenotypic data might be available; NCBI Genbank, Gramene, DRYAD, .. or plant specific system MaizeDB, Coffee Genome hub, Rice databases, etc)

ii. Additional attributes in demand

Completeness of the supplementary information associated with crop taxa: varieties names are necessary

iii. Additional sources - > what are the possible connectors between GBIF and the sources

Keep or share unique identifier across databases. Either GBIF kept the other database unique identifier (best solution for entry from gbif on agrobiodiversity data) or a unique identifier is shared across database (necessitate coordination)

iv. What data mobilization is needed? By whom?

- Mobilization of new occurrence data about agrobiodiversity is needed by the agrobiodiversity community into GBIF
- Centralization and completeness of occurrence at a given entry point (GBIF) by deposit into GBIF of data occurrence from other sources

→ Is this a priority? if not implemented how will it block the progress?

- Latitude and longitude validity is very important and access to tools that insure this quality are necessary. Data without quality should not be considered.
- Up to date species names are also very important because it allow to discuss of a given and precise category
- Link between databases and unique identifiers are of the utmost importance since it allow to link very different datasets useful for this objective.

Link to recommendations #

6.1.1, 6.1.2, 6.1.3, 6.1.4, 6.1.5, 6.2.1, 6.2.2, 6.2.3,6.2.4, 6.3.1, 6.4.1, 6..6.1,6.6.2,6.6.3, 6.6.4, 6.9.1, 6.9.2, 6.9.3, 6.11.1, 6.14.1, 6.14.2, 6.14.3

Bibliographic references

Use case D: Crop modelling - proposed by Ebrahim Jahanshiri

1. Describe the objective

Objective: To include/combine crop specific descriptors and vocabularies in addition to the common GBIF protocols/data standards to help the crop performance and modelling community use the most relevant crop cultivars/landrace and their associated information.
Background: Prediction and modelling of the crops is of outmost importance.
Crop specific data are crucial for specific modelling and yield projection exercises. Such data are normally calibrated for specific crop-cultivars and are available per crop model. Variety of standardisation is underway to improve the data flow from trials to modelling.

2. Who are the actors ?

Crop modelling community can converge more on using the standards to describe their field trials in addition to describing the range of environmental conditions. GBIF could consider extending the protocols. International Benchmark Sites Network for Agrotechnology Transfer (IBSNAT), The Agricultural Model Intercomparison Project (AgMIP) and Bioversity are the players in the standardisation descriptors and variables.

3. Data/information products to be produced

GBIF can provide assistance for this data in two ways. First, it can store the crop specific variables such as the plant descriptors and crop ontology terms that are published by Bioversity International in its internal database and relate them to the proper taxonomic information and second, assist with the agriculture diversification through recognising the inter-relationship of alternative crops and crop wild relatives that are currently mediated in relation to all species. GBIF should also assist in data completeness by encouraging gap analysis at national and global levels to assure reliable quality of data with regards to valid attributes of geographic coordinates, time and taxonomy. The users will be able to filter, relate and find information on the new crops for niche modelling, crop potential mapping and etc

4. Data sources the most used

1-The standards developed by the International Benchmark Sites Network for Agrotechnology Transfer (IBSNAT) project and subsequently revised by the International Consortium for Agricultural Systems Applications (ICASA) were of considerable value for describing experiments (White, et.al 2013). The new ICASA v2 can be implemented as extensions. ICASA v2 contains 600 variables and therefore only a core set of it related to experiments can be implemented. Bioversity Int. is currently mapping Crop Ontology to the ICASA variables for trial management.

2- Also relating to the other types of species that have relationships with the plants/crops of interest for example the parasitic relationship (insect, pest, etc.) with the species of interest. Darwin Core has an extension that can be used to relate the two species and describe their relationships. Implementing the additional information in the Darwin Core for host and parasitic information, is very valuable (Ex. "host: Quercus alba", "parasitoid of:Cyclocephala signaticollis | predator of Apis mellifera" Darwin Core terms, 2015).

3- Another section (or extending the overview section result of the specie search) related to physiological and agronomic aspects. As an example the following are the basic information that are needed for one of the popular crop models (FAO AquaCrop, 2015).

- Number of plants per hectare
- Time from planting to emergence
- Maximum canopy cover,
- Time from emergence to start senescence
- Time from emergence to maturity, i.e. length of crop cycle
- Time from emergence to flowering
- Length of the flowering stage
- Maximum effective rooting depth
- Time from sowing to maximum rooting depth
- Reference Harvest Indexes
- Water Productivity

It will be interesting to enrich GBIF data with links to efforts like Crop Ontology that strive to compile and validate standard agricultural concepts and trait dictionaries.

4- GBIF could act as mediator for the linked information on the new food crops through filters and search functions that are designed to accommodate standard agricultural terms related to agronomy. This facility would be very helpful in agricultural diversification projects that are ongoing (for example Crops For the Future mandate on introducing alternative crops). Filter and search mechanism within the GBIF could be extended to include the abovementioned physiological characteristics that can be used for research on the crop modelling using the information on the related crops.

Linking to data standards in agriculture (White, et. Al 2013), the invasive species databases like CABI CropWise database will be useful to distinguish the invasive species and a link to a database describing pathogens either identified in the accessions or in the locality where the accessions were collected could greatly enrich GBIF data. Interactions between insects and crops: e.g. the sensitive stage is when the crop is at the stage of seedling.

5. Tools the most used

The tools that are already in use by the agricultural community and modellers to facilitate the inter-comparison of the models can be cited in the GBIF section for agrobiodiversity tools (Porter, et. al 2014).New database like CFF database is under development for the underutilised crops will be able to link easily with GBIF backbone for taxonomy checking and other related information to food crops that can be used for agriculture diversification.

6. Describe the data flow into steps ("to be able to...") indicating who is doing the indicated step (a scientist, a developer, the system?)

To be able to …

7. Identify GBIF role and improvements

i. status of data (quality, coverage)

Rate of adoption of standards in ABD is increasing and this is a very good time to add GBIF data sources to the big picture of ABD research.

ii. Additional attributes in demand

Crop Ontology, ICASA standard variables, IBSNAT,Darwin core extension

iii. Additional sources - > what are the possible connectors between GBIF and the sources

Bioversity International, AgMIP, CFF

iv. What data mobilization is needed? By whom?

GBIF task force on ABD could create a list of possible relevant data and their associated ID to be linked to GBIF data.

→ Is this a priority? if not implemented how will it block the progress?

Link to recommendations #

6.1.1, 6.1.2, 6.1.5, 4.2.4, 6.2.6, 6.4.1, 6.8.2, 6.10.1, 6.10.2, 6.11.1,6.11.2, 6.11.3, 6.12.2, 6.12.3, 6.14.1,6.14.2,6.14.3

Darwin Core Terms: A quick reference guide. (n.d.). Retrieved October 11, 2015, from http://rs.tdwg.org/dwc/terms/#locationindex

FAO AQUACROP. (n.d.). Retrieved October 11, 2015, from http://www.fao.org/nr/water/infores_databases_aquacrop.html

Porter, C. H., Villalobos, C., Holzworth, D., Nelson, R., White, J. W., Athanasiadis, I. N., … Jones, J. W. (2014). Harmonization and translation of crop modeling data to ensure interoperability. *Environmental Modelling & Software*. http://doi.org/10.1016/j.envsoft.2014.09.004.

Raes, D., Steduto, P., Hsiao, T. C., & Fereres, E. (2009). AquaCrop the FAO crop model to simulate yield response to water: II. Main algorithms and software description. *Agronomy Journal*, *101*(3), 438–447.

White, J. W., Hunt, L. A., Boote, K. J., Jones, J. W., Koo, J., Kim, S., … Hoogenboom, G. (2013). Integrated description of agricultural field experiments and production: The ICASA Version 2.0 data standards. *Computers and Electronics in Agriculture*, *96*, 1–12. http://doi.org/10.1016/j.compag.2013.04.003

Use case E: Predict distribution of use of [currently] orphan crops- proposed by Ebrahim Jahanshiri

1. Describe the objective

Objective: To increase known diversity of crop species by predicting the distribution of orphaned crops that will eventually provide enough diversity and choice to improve the traits of the available crops. To broaden the available crops, provide the alternative crops for changing climate, new diseases, etc, providing the suitability index for crops.

Background: To secure food supplies, one solution is to diversify the food crops. Currently there are 7000 species are listed as neglected or underutilised crops (Williams, 2002). These crops have the potential to substitute the current food crops or provide additional sustenance. A common hindrance to spread of these crops is the lack of knowledge and distribution of their species and their closed relatives. One important question is that given the basic characteristics of the species and other related information regarding available data on farming communities that grow these crops, where it can be found, so that the diversity that is needed to improve these crops can also be found. Crops For the Future

2. Who are the actors ?

Distribution modellers that specifically work on food crops to use the occurrence data more efficiently. GBIF to provide filters and hierarchy of the relationships with orphaned crops bearing specific tags to recognise the crops and their immediate and wild relatives. Also ABD community providing feedback on the use of GBIF data for their prediction modelling exercise.

3. Data/information products to be produced

The same methodology for mapping the distribution of the species for conservation, landscape restoration can be used specifically to map the distribution of the orphaned species. The focus here however should be on the eco-geography and trait distributions. Ethnobotanic data on the use of the crops will be necessary. Therefore, crop mapping for their traits is a valuable information for the breeders to relate the geography to traits distribution and make inferences on the availability of crop relatives for breeding purposes.

4. Data sources the most used

Inclusion, integration and linking of databases such as Germplasm Resources Information Network (GRIN), National Plant Germplasm System (NPGS), System-wide Information Network for Genetic Resources (SINGER), EURISCO, Genesys, SpeciesLink, JSTOR Plant Sciences, Botanical Research and Herbarium Management System (BRAHMS) and other with emphasis on the crop species is important (Crop Genebank knowledge base, 2015). There are some ethnobotany databases such as international and national ethnobotany databases that could be useful to link to.

5. Tools the most used

GIS and mapping software.

6. Describe the data flow into steps ("to be able to...") indicating who is doing the indicated step (a scientist, a developer, the system?)

- To be able to shortlist the crops and their relatives based on the location (ABD, scientists)
- To be able to retrieve the occurrence crops that are currently orphan and underutilised under at various levels (genera, cultivar/landrace) (GBIF to tag species based on this attribute and ABD scientists (Bioversity Int. and Crops For the Future currently can help with recognising the list of these crops)

7. Identify GBIF role and improvements

i. status of data (quality, coverage)

ii. Additional attributes in demand

Orphaned, underutilized species, checklists of species and landraces, traits, use of the plant, nutritional value

iii. Additional sources - > what are the possible connectors between GBIF and the sources

Bioversity Int., Crops For the Future.

iv. What data mobilization is needed? By whom?

Mediators can be contacted from both of these organisations to further retrieve information on these crops. The mediators will be able to provide the information and start collaboration with GBIF.

→ Is this a priority? if not implemented how will it block the progress?

Link to recommendations #

6.1.1, 6.1.4, 6.1.5, 6.2.4, 6.3.1, 6.4.1, 6.8.1, 6.8.3, 6.14.1

Bibliographic references

Crop Genebank Knowledge Base. (n.d.). Retrieved October 11, 2015, from http://cropgenebank.sgrp.cgiar.org/index.php?option=com_content&view=article&id=662
Williams, J. T. (2002). Global Research on Underutilized Crops. An Assessment of Current Activities and Proposals for Enhanced Activities.

Use case F: Restoration of degraded landscapes and ecosystem services in a given country (Ethiopia) - proposed by Elizabeth Arnaud

1. Describe the objective

'As global population continues to rise, forests and agricultural land must be sustainably managed and more effectively used to satisfy increasing food demands and mitigate carbon emissions (World Resource Institutes Web Site, http://www.wri.org/our-work/project/global-restoration-initiative)'. When land is degraded, many of the benefits that ecosystems provide to local communities and agricultural production are also degraded, food security is compromised and resilience is reduced. Agrobiodiversity can be used to restore important services in agroecosystems e.g. soil quality, nutritional options at landscape level, pest and diseases control, pollination.

The objective is to enable researchers and local authorities to access regionally-relevant data sets on agrobiodiversity, soil, water, pest & diseases, gender-sensitive socio-economy, and policies to enable the identification of a mix of plant species and varieties maximizing the diversity for traits matching the needs of the restoration strategy along with information on seed availability - A major output of the research project will be a **participatory agricultural ecosystem restoration toolkit** supported by information systems. First product will be a Species database storing traits useful for restoration and nutrition-functional diversity in Nile region, available in standards APIs.

2. Who are the actors ?

Data managers, crop modelers, species modelers, restoration experts, local authorities, local communities in charge of the restoration

3. Data/information products to be produced

- Published national species checklist
- Mix of species and Landraces/cultivars lists with traits of interest for restoration of the target area and ecosystems services , and uses by the communities
- Predictive ecogeographic distribution of species

- Identification of data gaps and guidance for targeted inventories/field survey
- Identification of the species dispersal potential/barriers
- Localization of the seed sources

4. Data sources the most used

- Occurrences : GBIF, Collecting mission database, Genesys, national inventories, Kew databases
- Taxon name: GBIF taxonomy backbone, PlantList (Kew), Checklist of CWR, checklist of Neglected and underutilized species, National checklist of endemic species (Ethiopia)
- Traits for restoration : Treedatabase (ICRAF), useful plants (Kew), TRY, FAO databases for Food
- Species distribution, land cover, land use : Geonetwork (FAO)
- National Red list of Threathened species (Ethiopia)

5. Tools the most used

- Taxonomy curation : using Taxonomic Nomenclature Resolution Service v3.0 (TNRS) , The Plant List
- Georeference curation : Geolocate
- Ecological niche modeling : DIVA-GIS, Maxent, Floramap
- Ecogeograhic modelling tools : BioVel, Capfitogen (http://www.planttreaty.org/content/tools-capfitogen International Treaty)
- Ecosystem services : tools of Natural Capital (Invest, Cost)
- Google Earth maps

6. Describe the data flow into steps ("to be able to...") indicating who is doing the indicated step (a scientist, a developer, the system?)

a. To be able to compile a base list of endemic and introduced species, including trees, and a list of landraces/cultivars

1. Download from global, regional and national sources and compile available list of endemic and introduced species lists, cultivars names checklists for the given country where the restoration will take place.
2. Compare with the validated results of community surveys already performed in the area : preferred species/landraces, uses of the plants
3. Resolve taxonomic names using the GBIF taxonomic backbone - Taxonomic reconciliation is a time-consuming but essential step.
4. Classify the species into shrub, tree, grass, succulent, etc
5. Add principal uses: food, fodder, fuelwood, apiculture, intercropping, etc
6. Get traits: ecological, functional, agronomic, inter-species relations

7. Add
    - Crop Wild Relative Status
    - Neglected and Underutilized Species status
    - Invasive species status
    - Nutrition value
    - Threat status (IUCN)

b. **To be able to use this checklist to regularly extract data from GBIF , and other prefered sources, and getting occurrences, kml file and any complementary information included in Darwin Core Germplasm**

1. Upload the checklists in the GBIF taxonomy Backbone
2. Get the data sets
3. go to a simple pipeline to check the coordinates and improve quality
4. Need to save this checklist in myGBIF space for :
5. reuse on GBIF later and get the updates on data sets
6. for searching additional sources linked to GBIF to get more data (e.g. CWR, NUS, Invasive species, IUCN status, Nutritional value …)
7. Save the search results in myGBIF space

c. To be able to produce predictive mix of species, landraces and cultivars

1. Get quality occurrences for the species distribution in the country, at the site level. Compile occurrence data of key species in the restoration process and link them to environmental data (climate (WORLDCLIM), soil…) to Perform a Species Distribution Model in targeted areas
2. Get time series to show the evolution of the land cover and species distribution from e.g. the last 20 years
3. Run a species eco-geographic model to predict species with traits of interest for the restoration needs and species with an adaptive potential for the targeted restoration.

d. To be able to check the seed availability of the predictive mix of species, landraces and cultivars

1. Locate where seeds are conserved and available
2. Identify if methods for regeneration, seedlings are available

e. To be able to identify the threats and beneficial services in the target regions

1. Get occurences for the target regions for of pest and diseases,
2. Occurrences useful diversity (pollinators, underground diversity)
3. Occurrences of livestocks

7. Identify GBIF role and improvements

GBIF role

Provide an entry point for starting a compilation of species and landraces/cultivars for a target region for restoration. GBIF portal must be a source of quality occurrences and relevant additional attributes (e.g. Darwin Core germplasm) along with a complete taxonomy that includes infraspecies levels and propose taxon name resolvers. Propose a profile that guides the users seeking for taxon attributes that are relevant for restoration (traits, species status, land cover). Resulting checklist can be uploaded in GBIF and stored in a user's specific space on the portal 'myGBIF'. This user's space should enable storing access to the preferred external sources and running on demand stored queries.

i. status of data (quality, coverage)

- not enough occurrences at national level - Usually a minimum of 20–50 records is needed to produce accurate species distribution models based on 'presence-only' data
- Need a species distribution map per country of higher granularity
- Taxonomy backbone is incomplete: add Mansfield taxonomy, PlantList, etc, authoritative lists of landraces/cultivar names
- Provide taxonomic name resolvers
- Classification of the species into shrub, tree, grass, succulent, etc currently appears through the load of the wikipedia page and not as a searchable filter

ii. Additional attributes in demand

- National checklists of species, authoritative lists of landraces/cultivars
- Darwin Core germplasm extension with all attributes of the Multi Crop Passport Data
- Crop Wild Relative Status
- Neglected and Underutilized Species status
- Invasive species status
- Nutrition value
- Threat status (IUCN)
- Traits useful for restoration: ecological traits, agronomic traits, nutritional value, usage of the plant
- Seed availability
- Add or link to appropriate controlled vocabularies (traits from the TRY thesaurus, Crop Ontology, Planteome, AGROVOC, CABI, etc)
- add shapefiles for download
- Enable routine to get updates and resolves new names
- Propose cleaning pipelines (e.g. BioVel) and storage of the resulting maps in 'myGBIF'

iii. Additional sources - > what are the connectors

- Genesys – Darwin Core germplasm attributes
- Collecting mission database – taxon name, country, location
- CWR checklist of the Crop wild relatives and climate change Portal – (http://www.cwrdiversity.org/) - taxon name, country
- Treedatabase (ICRAF), Useful Plants database (Kew) – taxon name, country, administrative boundaries, traits
- TRY – taxon names, country, traits
- Geonetwork – country

We need also a discovery system for identifying other relevant sources of data

iv. What data mobilization is needed? By whom?

- Data mobilization on species occurrences at the national level by the GBIF nodes and scientists from herbaria, research institute
- Regional data mobilization by countries sharing borders and similar restoration project
- Publish quality data - Currently, one-third of the names entered into online databases are estimated to be incorrect and about 15% of species names in herbaria specimens are misspelled (Whitfield 2011).
- Collect of relevant maps and remote sensing data by experts
- Collect of community preferences and usage done of the plants and trees by NGOs, extensionists, conservationists engaged in restoration

→ Is this a priority? if not implemented how will it block the progress?

Link to recommendations #

6.2.1, 6.2.2, 6.2.3, 6.2.4, 6.2.5, 6.3.1, 6.3.2, 6.4.1, 6.5.1, 6.5.2, 6.5.3, 6.5.4, 6.6.1, 6.6.2, 6.6.4, 6.7.1, 6.8.1, 6.8.2, 6.8.3, 6.8.4, 6.10.1, 6.10.2, 6.11.1, 6.11.5, 6.11.6, 6.12.3, 6.13.1, 6.13.2, 6.13.3

Bibliographic references

Vivero J.L., Kelbessa E. and Demissew S. (2005) - The Red List of Endemic Trees & Shrubs of Ethiopia and Eritrea – Fauna & Flora International, Cambridge, UK, booklet, 28 p.

Awas T. - Endemic plants of Ethiopia (2009)- Preliminary working list to contribute to National plant conservation target - Institute of Biodiversity Conservation (IBC) - http://www.ibc.gov.et/wp-content/uploads/downloads/Endemic_plants_of_Ethiopia-Reported.pdf

Climate-Resilient Green Economy (CRGE). (2011). Ethiopia's Climate-Resilient Green Economy, Green Economy Strategy, brochure.

Government of the Federal Democratic Republic of Ethiopia (2013) National Nutrition Programme June 2013-June 2015

Maundu, P.; Bosibori, E.; Kibet, S.; Morimoto, Y.; Odubo, A.; Kapeta, B.; Muiruri, P.; Adeka, R.; Ombonya, J. (2013) UNESCO Safeguarding Intangible Cultural Heritage, 17p

Use case G: Access and benefit sharing (ABS) mechanism (FAO International Treaty) - proposed by Dag Terje Endresen

1. Describe the objective

International agreements and treaties prescribe access and benefit sharing for designated crops (Treaty Annex 1 list of crops) (FAO 2009a). The Convention of Biological Diversity (CBD) establishes biodiversity resources as a national sovereign property. Need to develop mechanisms to assess the use of this genetic resource in commercial activities such as commercial crop improvement breeding companies - so that breeding companies can pay the 1.1% fee on profits (sales gross income) from protected products as prescribed by the multilateral system (MLS) agreement (FAO 2009b).

2. Who are the actors?

- Genebank with germplasm material assigned to be part of the multilateral system (MLS) need to issue a so-called Simple Material Transfer Agreement (SMTA) when distributing MLS material. The SMTA must be registered and shared with the International Treaty (ITPGRFA FAO).
- Germplasm users (such as seed companies, plant breeders and researchers) requesting living germplasm material in the MLS need to sign and comply with the terms and conditions of the SMTA and the ITPGRFA.
- The International Treaty secretariat (ITPGRFA FAO) needs to maintain a global information system (GLIS) with information about all the material in the MLS and all the SMTAs issued in response to respective seed request.

3. Data/information products to be produced

- A global information system (GLIS) with information about all germplasm material assigned into the multilateral system (MLS), and including in particular the DOI for the genebank germplasm accession and the DOI assigned to the recipient's copy of the living germplasm material.
- SMTA transaction information including information about all (MLS) germplasm material shipped by a genebank in response to a legible seed request from a germplasm user. All germplasm accessions (or similar germplasm material) shipped under SMTA conditions must include the DOI of the genebank accession and the DOI assigned to the recipient's copy of this germplasm material. The recipient will need to keep track of this assigned DOI for reporting use and final products.

4. Data sources the most used

- GeneSys
- EURISCO

5. Tools the most used

- The International Treaty on Plant Genetic Resources for Food and Agriculture (FAO 2009a)
- SMTA guidelines (FAO 2009b)
- Easy SMTA (FAO 2012)

6. Describe the data flow into steps ("to be able to...") indicating who is doing the indicated step (a scientist, a developer, the system?)

To be able to …

7. Identify GBIF role and improvements

i. Status of data (quality, coverage)

ii. Additional attributes in demand

iii. Additional sources - > what are the possible connectors between GBIF and the sources

iv. What data mobilization is needed? By whom?

→ Is this a priority? if not implemented how will it block the progress?

Link to recommendations #

Bibliographic references

FAO (2009a). International Treaty on Plant Genetic Resources for Food and Agriculture. Food and Agriculture Organization of the United Nations, Rome, Italy. Available at http://www.fao.org/docrep/011/i0510e/i0510e00.htm

FAO (2009b). Illustrated version of the Standard Material Transfer Agreement. Food and Agriculture Organization of the United Nations, Rome, Italy. Available at http://www.planttreaty.org/node/1646

FAO (2010). The Second Report on the State of the World’s Plant Genetic Resources for Food and Agriculture. Available at: http://www.fao.org/docrep/013/i1500e/i1500e.pdf

FAO (2012). Easy-SMTA: User manual. Food and Agriculture Organization of the United Nations, Rome, Italy. Available at http://www.planttreaty.org/sites/default/files/Easy-SMTA_Manual_en.pdf

Use Case H: Identifying gaps in data for Agrobiodiversity to guide targeted data collect proposed by Jean Cossi Ganglo

1. Describe the objective

The process of data curation reveals important gaps in GBIF mediated data and depending on data uses important data are lost with regards to original data downloaded from GBIF site. It is important to fill the gaps usually identified on data attributes, mainly taxonomic names, geographic coordinates, time collection... Gaps are also noted with regards to spatial / environmental coverage and taxonomic groups.

2. Who are the actors ?

GBIF experts in capacity building, node managers and other data publishers

3. Data/information products to be produced

At national and global levels, increase significantly the percentage of curated data so that at least 80% of GBIF mediated data can be used in Ecological Niche Modelling

4. Data sources the most used

- GBIF data portal
- National inventories
- Checklist of species and landraces/cultivars

5. Tools the most used

1. Google refine
2. Geolocate
3. Taxonomic name resolver
4. Maxent, Floramap, etc

6. Describe the data flow into steps ("to be able to...") indicating who is doing the indicated step (a scientist, a developer, the system?)

To be able to increase the quality data in GBIF

1. Identify with the Agrobiodiversity community the gaps in species occurrences, time series, taxonomic groups, national coverage
2. Capacitate node managers and other data publishers to perform gap analyses on GBIF data at national levels
3. Set priorities on data collections to fill in the gaps identified
4. Train the node managers in collecting the data relevant for Agrobiodiversity

7. Identify GBIF role and improvements

i. status of data (quality, coverage)
ii. Additional attributes in demand
iii. Additional sources - > what are the possible connectors between GBIF and the sources
v. What data mobilization is needed? By whom? Mobilize occurrences of species of importance for agrobiodiversity and sample-base data, national checklist of species and landraces, trait data
→ Is this a priority? if not implemented how will it block the progress?
Link to recommendations #
6.3.3, 6.4.1, 6.11.1, 6.11.2
Bibliographic references