



Elizabeth Arnaud¹, Nora Castañeda²,
Jean Cossi Ganglo³, Dag Endresen⁴,
Ebrahim Jahanshiri⁵, Yves Vigouroux⁶

Bioversity International¹, CIAT and Univ. Birmingham²,
University of Abomey-Calavi & GBIF Benin³, GBIF Norway &
Natural History Museum, University of Oslo⁴, Crops for the
Future⁵.

Institut de Recherche pour le Développement (IRD) ⁶



GBIF Data Quality for Agrobiodiversity

Report of the Task Group



UNIVERSITY OF
BIRMINGHAM



IRD
Institut de recherche
pour le développement

GBIF Norway

Task Group on GBIF Data Fitness for Use for Agrobiodiversity

- 2015
- To capture the best available experiences, document limitations in existing GBIF services, and suggest improvements in the functionality of GBIF.org for domain-specific needs.
- 51 experts responded to the survey



UNIVERSITY OF
BIRMINGHAM



GBIF Norway



Agrobiodiversity

BIODIVERSITY

Agrobiodiversity

Mixed agro-ecosystems
Crop species/varieties ←
Livestock and fish species
Plant/animal germplasm
Soil organisms in cultivated areas
Biocontrol agents for crop/livestock pests
Wild species as landraces or with breeding
Cultural & local knowledge of diversity

- Diversity of crops, wild relatives, trees, animals, microbes and other species that contribute to agricultural production, resilience to climatic events, to environmental and economic changes, restoration of ecosystem services (pollination), provides a genetic reservoir of new traits and species for farming



Agrobiodiversity – focus on plants

950 species of cultivated plants are threatened globally (*Khoshbakht and Hammer, 2007*).

- ❑ Crop Wild Relative Species
- ❑ Landraces/cultivars
- ❑ Neglected and underutilized species



Agrobiodiversity conservation, management and monitoring

- **Large range of data:** *Taxon, vernacular names, occurrences, geospatial distribution, genotype, phenotype, environmental factors, agronomic traits, functional traits, species interactions, socio-economic factors, local knowledge...*
- **Discovery, access and adequate use** of primary biodiversity data to inform decision making
- Bring enough data **of comparable scale and granularity**
- Identify gaps, find **proxys** and **interpolate data**
- Develop **metadata** and **controlled vocabularies** for describing data sets (taxon, traits, etc)
- Multilingual and multidisciplinary knowledge

Agrobiodiversity-based Restoration in Ethiopia

Figure 1. Sites location and land cover
(Data Source: GlobeLand30/NGCC)

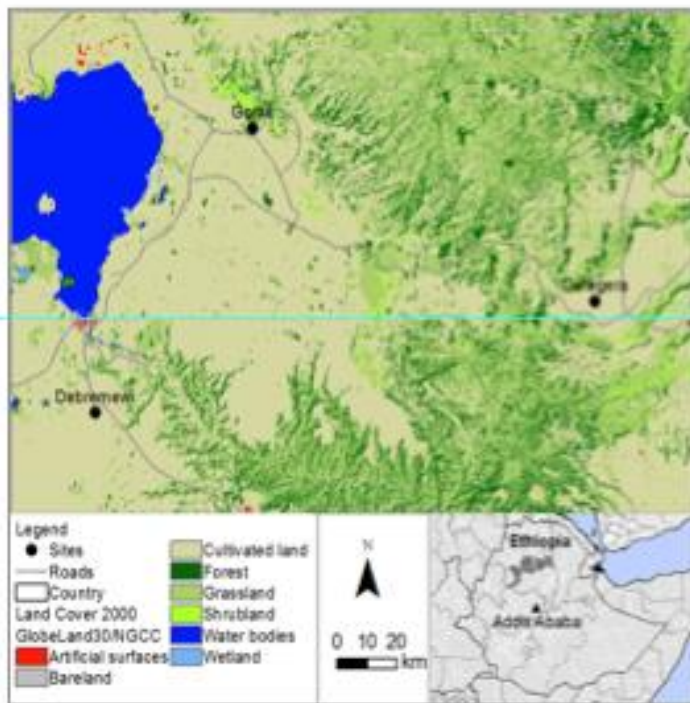
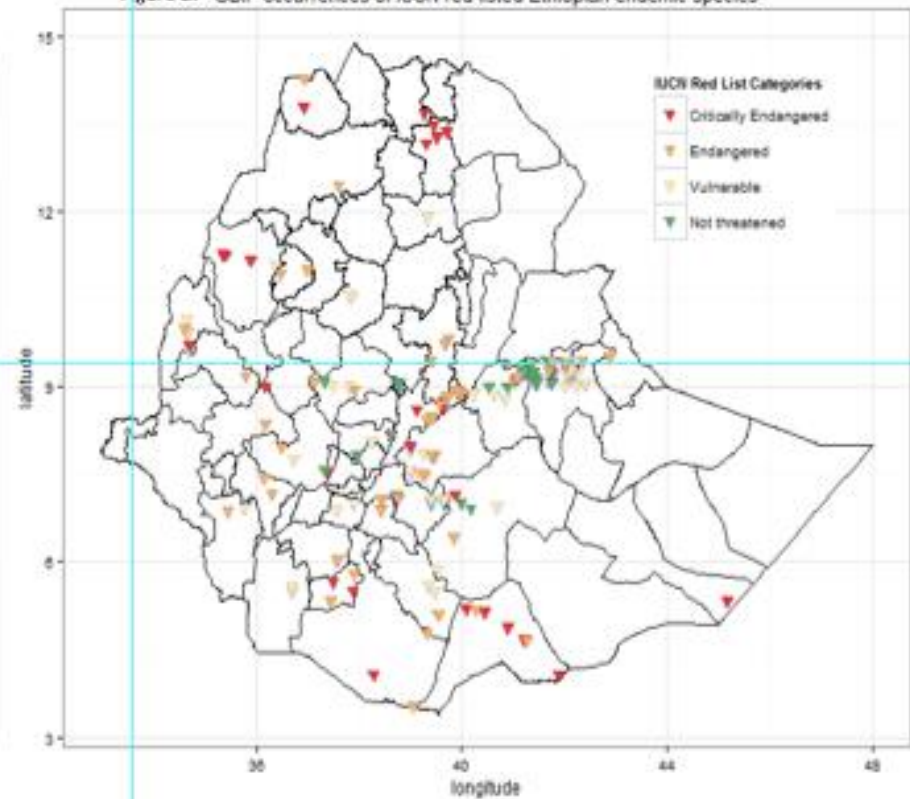


Figure 2: GBIF occurrences of IUCN red listed Ethiopian endemic species



Agrobiodiversity Data



- Data Standards exist for *ex situ* conservation and genetics
- Unstructured data for *in situ* wild and on farm
- Not easily linked

- GBIF can play an important role by mobilizing and connecting biodiversity and agrobiodiversity datasets

Ex situ conservation – Genebanks

- 7,4 million samples of crop varieties and wild relatives in genebanks worldwide (FAO, SOW-PGRFA)
- Multi Crop Passport Data (MCPD)** – core metadata describing living samples – **Unique Identifier** for accession management and exchange
- Integrate MCPD into Darwin Core terms Darwin Core germplasm extension**
- Add attributes for pre-breeding and breeding data, and characterization & evaluation of trait data



Term	MCPD (2015)	Darwin Core (dwc), germplasm (g)
NA	(not applicable)	dwc:datasetID
0	PUID	dwc:occurrenceID
1	INSTCODE	dwc:institutionCode
2	ACCENUMB	dwc:catalogNumber
3	COLLNUMB	dwc:recordNumber
4	COLLCODE	g:collectingInstituteID
4.1	COLLNAME	dwc:recordedBy
4.1.1	COLLINSTADDRESS	(dwc:recordedBy)
4.2	COLLMISSID	dwc:collectionCode
5	GENUS	dwc:genus
6	SPECIES	dwc:specificEpithet
7	SPAUTHOR	dwc:scientificNameAuthorship (if SUBTAXA is empty)
8	SUBTAXA	dwc:infraspecificEpithet
9	SUBTAUTHOR	dwc:scientificNameAuthorship (if SUBTAXA is not empty)
10	CROPNAME	dwc:vernacularName

Classification & identification of Crop Wild Relatives species (CWR)



- Publish **Crop Wild Relatives Species checklists** to GBIF portal and Integrated into the GBIF taxonomy backbone:
 - Global crop wild relative species checklist (www.cwrdiversity.org/checklist/) and the SADEC-CWR checklist.
- Add Taxon-backbone-level attributes and search filters to:
 - Identify the conservation priority and status of crop wild relative species at the global level
 - Provide information on the relationship of crop wild relatives (CWR) and their associated crops (e.g. gene pool and taxon group).

Mobilizing data on cultivated plants



- Taxonomy and checklists of traditional names in relevant languages.
- Geospatial distribution information in cultivated areas. (remote sensing, crowd sourcing)
- Morpho-taxonomy, agronomic traits (farmers and breeders), functional traits,
- local uses, characterization and evaluation data.
- Use, agronomic practices, cultural practices, seed conservation and exchange

Interactions between species



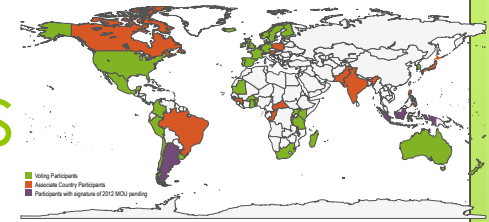
- **Additional attributes needed :**
 - species relation to crops or between species at the taxon name level like pest, predator, pollinator, etc.
 - 'pathogen' with the scientific name of the pathogens and vernacular names of the diseases should be made available.
- GBIF portal to enable the selection and download of crop occurrences along with occurrences of pests, diseases, pollinators, livestock, etc.
- Information on the risk of a species becoming invasive - a link between the GBIF portal CABI database on invasive species



Improving the mobilization of new data sources

- Publish on GBIF existing digitized ABD data sources & collections,
- Stimulate the digitization of relevant collections through small competitive grants
- Simplify the upload of Agrobiodiversity data records, requiring a very low technical expertise,
- Promote the use of citizen data portals such as the EarthSkySea and iSpot as a mean for publishing data to GBIF.org.

Data Mobilization targets for GBIF Nodes



- **Input of national experts for national inventories and checklists**

- Specific objectives should be to:
 - Revise or complete the taxonomy of lesser known plant genera
 - Evaluate the species' genetic diversity
 - Gather detailed information on species distribution and ecology
 - Collect ethnobotanical data on folk knowledge and traditional uses
 - Assess the nutritional value and potential for commercialization of NUS
 - Data on crop wild relatives and on species traits useful for crop improvement and for landscape restoration.



GBIF services and tools for data processing and cleaning

- GBIF = point of access for **the most reliable and up-to-date taxonomy**
- GBIF to **support the integration of popular data cleaning tools** such as GEOLocate, OpenRefine, and workflow services from BioVeL, Galaxy or Taverna compliant protocols
- **Increase visibility of existing taxonomic name reconciliation tools** : Global Names Architecture (GNA), access to the Plant List
- Implement or improve tools for cross-checking and validating nomenclature of records published by different collections -to resolve naming issues

Improving fitness for use through data quality

Develop or adapt existing tools to:

- ❑ **Identify quality improvement thresholds** based on the decided weighting of scores e.g. issues with taxon names regarding completeness of name-strings and up-to-date nomenclature;
- ❑ **Develop standard suite of tests** to be applied at record level to the fields of Darwin Core and germplasm extension.
- ❑ Suspect or in-error status be tabulated and accumulated into a score for the record, but the **weighting** given to the tests would be **use-dependent**.
- ❑ Evaluation at the **dataset-level** is necessary as some quality assessments can only be inferred from aggregates of data records

Improving fitness for use through data quality

- ❑ **Check the completeness of the data**, provide percentage of records with actual data reported for each attribute
- ❑ Highlight attributes in a search result **with no actual data reported.**
- ❑ Provide **statistics on the percentage of completeness and issues over time**
- ❑ Add a **quality stamp** provided by the data owner and the publisher.
- ❑ **User-friendly & adequate tools to perform data cleaning** at national levels before data publication.
- ❑ GBIF portal to **display data quality issues as standard annotations** to the respective occurrence or taxon entity.
- ❑ Data quality annotations available **in data downloads** from the GBIF Portal



Combining GBIF-mediated data with external data sources

- **Making GBIF API more interconnected and accessible** to the agricultural information providers (AGRIS, FAOSTATS, CABI).
- GBIF and partners to explore connecting GBIF-mediated data **with Linked Open Data (LOD)** and testing adequate formats for semantic web services
- **Occurrence-level data** most often link to external data sources using geographic coordinates, by taxon name, and taxon ID. Consequently quality of these data is paramount
- Expand current data types (taxon, occurrence and event) to include **experimental data** such as crop trait information and characterization and evaluation data on crop disease resistance, and other economically useful trait measurements required in crop improvement activities



Agrobiodiversity user profile access

- **A hierarchy of data-profiles and user-profiles**, thematically designed search widgets, and tools enabling thematic users to increase efficiency of their use of the GBIF portal to find the range of data they need.
- Implementing a hierarchy of thematically designed data/user profiles **at the API level** would also help with simplified access.

Use Cases

1. **Find gaps in conservation in crop wild relatives**
2. Conservation plan for Crop Wild Relatives
3. Find agrobiodiversity hotspots and monitor changes for decision on conservation
4. Crop modelling
5. Predict distribution of use of [currently] orphan crops
6. **Restoration of degraded landscapes and ecosystem services in a given country (Ethiopia)**
7. Access and benefit sharing (ABS) mechanism (FAO International Treaty)
8. Identifying gaps in data for Agrobiodiversity to guide targeted data collect

Report, Survey results and Use cases:

<http://www.gbif.org/resource/82283>

Thank you !

