Data and metrics on the conservation and use of genetic resources of crops
Extended methods and materials

Background

This document describes the sources of information and methods used to prepare concise, data-driven reports on the status of conservation and use of genetic resources of specific crops. These methods have been developed under the *Mainstreaming Global Crop Conservation Strategies in Plant Treaty Processes* project, and are currently being used to prepare concise reports for approximately 30 crop groups. The methods can serve as guidelines for authors and contributors of the Global Crop Conservation Strategies (GCCS) when analyzing publicly available data in support of future GCCS and related initiatives. The format can also serve as a guideline for creating concise, data-driven updates to the GCCS and related documents.

These concise reports provide a current overview of the global *ex situ* conservation status for the crop and its wild relatives, including key metrics on:

- the identity and composition of genebank collections;
- the Multilateral System (MLS) status of accessions in these collections;
- storage, regeneration, and safety duplication status;
- documentation, information systems, and research resources;
- germplasm distribution; and
- varietal registrations and releases.

The report also includes global statistics on crop production, trade, and availability in food supplies (as appropriate), as well as information about crop networks and partnerships. The report is meant to provide an update to some of the information presented in the previous Global Conservation Strategy for the crop (if one has been published), but is primarily based on publicly available datasets, rather than a survey of genetic resource collections institutions and expert consultation meetings.

The development of this document was funded by the German Federal Ministry of Agriculture, Food and Regional Identity (BMLEH) as part of the project *Mainstreaming Global Crop Conservation Strategies in Plant Treaty Processes* led by the Crop Trust. The Crop Trust cooperated with the Secretariat of the the International Treaty on Plant Genetic Resources for Food and Agriculture in the development of this document.

Data sources and acquisition

Primary data sources for the metrics reported in this summary included:

- Genesys Plant Genetic Resources portal (Genesys) - for *ex situ* collections holdings, composition, and conditions; Multilateral System (MLS) status; and safety duplication status. Data retrieved in October 2024 and reporting status at that date. Genesys PGR was also the source of characterization and evaluation datasets, with data retrieved June 2025.
- World Information and Early Warning System on Plant Genetic Resources for Food and Agriculture of the Food and Agriculture Organization of the United Nations (FAO WIEWS) - for *ex situ* collections holdings, composition, and conditions; Multilateral System (MLS) status; regeneration status; safety duplication status; and germplasm distributions and varietal releases. *Ex situ* collections data retrieved in October 2024 and reporting information from 2023 or previous country level reports, with data for regeneration status reporting the period 2012 to 2014 (retrieved in 2024), for germplasm distributions reporting the period 2014 to 2019 (these metrics retrieved in March 2022 - Khoury *et al.* [2023] and Khoury *et al.* [2025]), and for varietal releases reporting the period 2015 to 2019 (these metrics retrieved in 2022 - Khoury *et al.* [2023]).
- Botanic Gardens Conservation International PlantSearch database (BGCI PlantSearch) - for *ex situ* collections holdings and composition. Data retrieved in November 2024 and reporting status at that date.
- Global Information System of the International Treaty on Plant Genetic Resources for Food and Agriculture (Plant Treaty GLIS) - for Multilateral System (MLS) status. Data retrieved directly from the Plant Treaty in October 2025 and reporting status at that date.
- Svalbard Global Seed Vault portal (SGSV portal) - for safety duplication status. Data retrieved in October 2024 and reporting status at that date.
- Data Store of the International Treaty on Plant Genetic Resources for Food and Agriculture (Plant Treaty Data Store) - for germplasm distributions. Data retrieved in October 2024 and reporting information for years 2019 to 2021, with data also retrieved in February and June 2022, reporting information for years 2015 to 2018 (from Khoury *et al.* [2023] and Khoury et al. [2025]). Data assessed included all transfers within these years, including those initiated both from international (CGIAR) collections as well as all other providers.
- International Union for the Protection of New Varieties of Plants (UPOV) PLUTO database - for varietal registrations. Data retrieved in 2019, reporting information for years 2014 to 2018 (from Khoury *et al.* [2023]).
- FAOSTAT - for global production, trade, and food supplies statistics. Data retrieved in 2021, reporting information for years 2015 to 2018 (from Khoury *et al.* [2023]).
- Wikipedia - for public interest metric. Data retrieved in 2022 and reporting on public views of the wikipedia webpage of the crop taxon, over the entirety of one year (2019) (from Khoury *et al.* [2023]).
- National Center for Biotechnology Information's Entrez database (NCBI Entrez) - for research resources. Data retrieved in 2019 and reporting status at that date (from Khoury *et al.* [2023]).

- Google Scholar - for research resources. Data retrieved in 2019 and reporting information on articles published between 2009 and 2019 with the crop taxon name in the article title (from Khoury *et al.* [2023]).
- PubMed Central - for research resources. Data retrieved in 2022 and reporting on articles published to that date with the crop taxon name within the article text (from Khoury *et al.* [2023]).
- Global Biodiversity Information Facility (GBIF) - for research resources. Data retrieved by application programming interface (API) in August 2025 and reporting status at that date.

Target taxa and taxonomic cleaning and standardization

Metrics on *ex situ* collections holdings, composition, and conditions; Multilateral System (MLS) status; and pertinent research resources (i.e., NCBI and GBIF) were typically developed to be inclusive of the entire crop genepool (generally defined as all taxa within the genus).

For the edible aroids summary, we developed metrics for the main aroid crops covered in the Edible Aroid Conservation Strategies (2010), including the genera *Alocasia, Amorphophallus, Colocasia, Cyrtosperma,* and *Xanthosoma*.

For the tropical and subtropical forages summary, to identify a pragmatic number of forage crops to assess, we surveyed the genera with the most forage crop accessions in Genesys, removing accessions of forage crops clearly only pertinent to temperate environments. We then further edited this list based on inputs from the International Center for Tropical Agriculture (CIAT) and the International Livestock Research Institute (ILRI), which maintain the largest international tropical and subtropical forage collections in the CGIAR, as well as from the author of the 2015 Global Conservation Strategy for tropical and subtropical forages. This process yielded a prioritized list of forage taxa across 24 genera [*Aeschynomene, Alysicarpus, Andropogon, Cenchrus (*syn. *Pennisetum), Centrosema, Chloris, Chrysopogon, Clitoria, Cynodon, Desmanthus, Desmodium, Digitaria, Lablab, Leucaena, Lotus, Macroptilium, Megathrysus, Panicum, Paspalum, Sesbania, Setaria, Stylosanthes, Urochloa (*syn. *Brachiaria),* and *Vigna*]. To avoid imbalanced results due to several food crop taxa also being within these genera, we removed accessions of food crop taxa identified in the dataset, including under the names: *Cenchrus americanus* (L.) Morrone (pearl millet), *Digitaria exilis* (Kippist) Stapf (white fonio), *Digitaria iburua* Stapf (black fonio), *Lablab purpureus* (L.) Sweet (lablab), *Panicum miliaceum* L. and *Panicum miliaceum* subsp. *miliaceum* (proso millet), *Panicum sumatrense* Roth (little millet), *Paspalum scrobiculatum* L. (kodo millet), *Setaria italica* (L.) P. Beauv. (foxtail millet), *Vigna aconitifolia* (Jacq.) Maréchal (mat bean), *Vigna angularis* (Willd.) Ohwi & H. Ohashi (adzuki bean), *Vigna mungo* (L.) Hepper (black gram), *Vigna radiata* (L.) R. Wilczek and *Vigna radiata* var. *radiata* (mung bean), *Vigna subterranea* (L.) Verdc. (bambara groundnut), *Vigna umbellata* (Thunb.) Ohwi & H. Ohashi (ricebean), and *Vigna unguiculata* (L.) Walp., *Vigna unguiculata* group sesquipedalis, and *Vigna unguiculata* subsp. *unguiculata* (cowpea).

Taxonomic standardization was conducted using both World Flora Online and GRIN Global Taxonomy via the Global Names Verifier application programming interface (API). For taxa where the accepted names differed between these two taxonomic resources, the final taxonomic name was determined based on crop-specific knowledge and literature review.

Genebank collections dataset annotation and processing

The main genebank collections dataset was compiled by merging the Genesys and FAO WIEWS datasets, both of which report information at the accession level. For each institution with data available from both datasets, either Genesys or FAO WIEWS was selected as the data source, based on our assessment of the recency of updates and the completeness of datasets. In general, data from CGIAR were compiled from Genesys, while the dataset containing the greater number of records for a given institution was selected for other institutions.

Institutes in the accession level dataset were assigned institutional types (CGIAR, international, national, etc.) based on the FAO WIEWS Organizations database, using the institute code and manual revision. Annex 1 status of crop was assigned according to the FAO (2009) listing. Primary and secondary regions of diversity were listed as per Khoury *et al.* (2023) with additions based on crop-specific literature review and subject matter expert inputs. Data columns were aligned as relevant with the Multi-Crop Passport Descriptor (MCPD) standard. Country codes were standardized using 3-letter ISO 3166-1 country codes. We retained the originally listed country codes of former countries that now correspond to multiple countries (e.g., Yugoslavia).

Metrics calculations

Introduction and background on crop

Status of production, trade, and food supply metrics were retrieved from Khoury *et al.* (2023), with a full explanation of the calculation process for those metrics included in that document. In short,

- Production, trade, and food supply statistics sourced from FAOSTAT, averaged across years 2015 to 2018.
- The "Number of countries where significant contributor" metric refers to the count of countries where the crop is reported as within the top 95 percent of crops in that country in terms of production, trade, or availability in food supplies.
- The "Evenness of contribution across world regions" metric quantifies evenness of production, trade, or availability in regional food supplies across world regions, using a GINI (inequality) index methodology, where in this case 0 equals highly uneven and 1 equals completely even production, trade, or availability in regional food supplies.

- The "Estimated international interdependence" metric quantifies the degree of production, trade, or availability in food supplies outside of the primary region of diversity of the crop, as an estimated proxy of international interdependence regarding plant genetic resources of the crop, where 0 equals low estimated international interdependence on crop (i.e., production, trade, or contribution to food supplies only in the primary region of diversity of the crop and not elsewhere around the world) and 1 equals high estimated international interdependence (i.e., crop produced, traded, and available in food supplies widely around the world.

Wikipedia pageviews also source from Khoury *et al.* (2023), reporting over one year (2019) the number of public pageviews on Wikipedia based on the name of the crop taxon.

Identity and composition of ex situ *collections*

The main dataset for these metrics is the combined Genesys and FAO WIEWS genebank collections dataset.

- The "Number of institutions holding genebank collections" metric was calculated by counting distinct institutions based on the institute code field (INSTCODE).
- The "Number of distinct taxonomic names in genebank collections" metric was calculated after taxonomic cleaning and standardization; the final names generally reflect accepted taxonomic names but are also inclusive of categories such as "*Genus* sp.", "*Genus* hybr.", "*Genus species* x *species*" (i.e., hybrids), and other general categories and variations thereof.
- The number of accessions of different crop improvement categories (i.e. crop wild relatives (CWR), weedy materials, landraces, etc.) metrics were based on sample status (using the SAMPSTAT field) reporting in the combined dataset, aligned with the Multi-Crop Passport Descriptor (MCPD) standard (available on Genesys). Crop wild relatives (CWR) were counted using SAMPSTAT values in the 100s range; weedy materials when SAMPSTAT was equal to 200; landraces when SAMPSTAT was equal to 300; breeding materials when SAMPSTAT values were from 400 to 499; improved varieties when SAMPSTAT was equal to 500; and other materials when SAMPSTAT was equal to 999. Metric calculation for GMO (SAMPSTAT equal to 600) was excluded from this study as there was a minute amount of data (the total number of records with STAMPSTAT equal to 600 across the 15 crops accessed was only 13 records). Accessions missing a SAMPSTAT value were also counted as a metric. All metrics include a percentage value, calculated as the count of accessions in the category divided by the total number of accessions in all categories, including those not marked with an improvement type.
- The "Number of countries where germplasm has been collected for genebank collections" metric was calculated by counting the number of unique countries (using the ORIGCTY field) from which germplasm had been collected, based only on records where SAMPSTAT = 100s, 200, 300, or 999.
- The "Number of accessions in genebank collections from the primary region(s) of diversity" were calculated by matching the ORIGCTY of accessions (excluding breeding materials and improved varieties) with those countries included in the primary regions of diversity identified in Khoury *et*

al. (2023). For the “Number of accessions in genebank collections from the primary and secondary region(s) of diversity” metric, secondary regions were identified per crop through literature review and are available on this repository. Both metrics include a percentage value, calculated as the count of accessions in the category divided by the total number of accessions of the crop.

- Botanic garden metrics: data sourced from BGCI PlantSearch (2024). Calculated metrics include both a count of distinct taxa in those collections (after taxonomic cleaning and standardization) as well as a count of distinct institutions (based on institution name as reported by BGCI GardenSearch ID).

Multilateral System status of accessions in ex situ *collections*

These metrics are estimated independently from two datasets: 1) the combined Genesys and FAO WIEWS genebank collections dataset; 2) the Plant Treaty GLIS portal (2025). Most of these metrics are self-evident, with data processing described above.

- The “Number of accessions in genebank collections in international institutions”, “Number of accessions in genebank collections in national or other institutions”, and “Number of accessions in genebank collections in Annex I” metrics were based on the combined Genesys and FAO WIEWS genebank collections dataset. International collections included CGIAR collections, non-CGIAR A15 collections, and other international non-CGIAR non-A15 collections; these are listed in the edited FAO WIEWS list of institutions we used. All metrics include a percentage value, calculated as the count of accessions in the category divided by the total number of accessions of the crop.
- The “Number of accessions with DOI (Plant Treaty GLIS 2025)” and “Number of accessions included in the Multilateral System (MLS) (Plant Treaty GLIS 2025)” metrics were based on the Plant Treaty GLIS portal (2025). For the second metric, the MLSSTAT field in the Plant Treaty GLIS portal was used. GLIS records without INSTCODE values were dropped from the GLIS dataset. Only active (not historical) records were counted.
- The “Number of accessions included in the Multilateral System (MLS) (genebank collections database)” and “Number of accessions not included in the Multilateral System (MLS) (genebank collections databases)” metrics were based on the MLSSTAT field in the combined genebank collections dataset. The “Number of accessions included in the Multilateral System (MLS) that are in international collections (genebank collections databases)” metric was calculated similarly, but counting only accessions maintained in international collections. All metrics include a percentage value, calculated as the count of accessions in the category divided by the total number of accessions of the crop.

Storage conditions, regeneration status, safety duplication

The main dataset for these metrics is the combined Genesys and FAO WIEWS genebank collections dataset.

- Storage type and seed storage condition (short-, medium-, and long-term) metrics were calculated based on the type of germplasm storage field (STORAGE), aligned with the Multi-Crop Passport Descriptor (MCPD) standard (available on Genesys) (i.e., 10=Seed collection; 11=Short term; 12=Medium term; 13=Long term; 20=Field collection; 30=In vitro collection; 40=Cryopreserved collection; 50=DNA collection; and 99=Other. Accessions not marked with a germplasm storage type were also counted as a metric. Metrics include a percentage value, calculated as the count of accessions in the category divided by the total number of accessions in all categories, including those not marked with an germplasm storage type.(i.e. the total number of accessions of the crop). For the seed storage conditions percentages, the values were divided by the total number of accessions stored as seed.
    - (For Table 2): For primarily seed-propagated crops, a metric was calculated for the number of accessions in long term cold storage based on 13=Long term.
    - (For Table 2): For primarily vegetatively propagated crops, a metric was calculated for the number of accessions conserved in vitro or in cryopreservation. Accessions that are conserved both in vitro and in cryopreservation were counted only once to avoid double counting.
- Regeneration status information is from FAO WIEWS (2024); data is from years 2012 to 2014 (cumulative).
- The "Number of accessions safety duplicated out of the country in genebank collections" was assessed using Genesys data only. Accessions were considered duplicated abroad if their reported duplication sites as reported in the Genesys dataset were located in countries other than the holding institution's country, determined by comparing the three-letter country code in INSTCODE with duplication site codes (using the DUPLSITE field), excluding SGSV and known internal entries. Percentage is the number of accessions duplicated out of the country divided by all accessions pertinent to the crop in the combined genebank collections dataset.
- Svalbard Global Seed Vault data is from the SGSV portal (2024). Percentage is the count of accessions in SGSV divided by the total accessions pertinent to the crop as listed in the combined genebank collections dataset.

Documentation, information systems, and research resources

- The Passport data completeness index (PDCI) metric utilizes accession-level data from Genesys (2024) only, and is calculated as described in van Hintum *et al.* (2011).
- Metrics on gene, genome, nucleotide, and protein resources for the crop source from the National Center for Biotechnology Information's Entrez database (NCBI Entrez), as reported in Khoury *et al.* (2023). Data was retrieved in 2019 and the metric reports status at that date.
- The Google Scholar metric is as reported in Khoury *et al.* (2023). Data was retrieved in 2019, with the metric reporting information on articles published between 2009 and 2019 with the pertinent crop taxon name in the article title.
- The PubMed Central metric is as reported in Khoury *et al.* (2023). Data was retrieved in 2022, with the metric reporting on articles published to that date with the pertinent crop taxon name within the article text.

- Global Biodiversity Information Facility data is from GBIF (2025), and includes all records listed in GBIF pertinent to the crop (i.e. any records of taxa within the crop genus and synonyms of the genus).

Germplasm distributions and varietal registrations and releases

Germplasm distributions and varietal registrations and releases metrics largely follow those developed and reported in Khoury *et al.* [2023].

- FAO WIEWS distributions data is the annual average count of distributions over years 2014 to 2019 for any accessions or samples distributed of taxa within the pertinent crop genus/genera, as reported in Khoury *et al.* (2023) and Khoury et al. (2025).
- Plant Treaty Data Store distributions data is the annual average count of distributions over years 2015 to 2021. This dataset was based on data from 2015 to 2018 from Khoury *et al.* (2023) and Khoury et al. (2025), combined with new data provided from the Plant Treaty Data Store for 2019 to 2021, which were merged for this metric.
- The "Number of countries receiving germplasm" metric counts the distinct number of countries receiving germplasm as an annual average over years 2015 to 2021, as recorded in the Plant Treaty Data Store distributions dataset described above. This metric is equivalent in methodology to that reported in Khoury *et al.* (2023) but includes the new data from years 2019 to 2021.
- The "Evenness of distributions across world regions" metric quantifies evenness of germplasm distributions across world regions based on the annual average distributions data over years 2015 to 2021, as recorded in the Plant Treaty Data Store distributions dataset described above. The evenness metric uses a GINI (inequality) index methodology, where in this case 0 equals highly uneven and 1 equals completely even distributions across all world regions. This metric is equivalent in methodology to that reported in Khoury *et al.* (2023) but includes the new data from years 2019 to 2021.
- The International Union for the Protection of New Varieties of Plants (UPOV) PLUTO data is the annual average count of registrations as recorded in PLUTO over years 2014 to 2018, from Khoury *et al.* (2023).
- FAO WIEWS varietal releases data is the annual average count of varietal releases over years 2015 to 2019, from Khoury *et al.* (2023).

Data and code availability

Data processing, metric calculation, and table generation were conducted in R, with code available on this GitHub repository.

References

FAO (2009) *The International Treaty on Plant Genetic Resources for Food and Agriculture*. Food and Agriculture Organization of the United Nations (FAO), Rome, Italy.
https://openknowledge.fao.org/server/api/core/bitstreams/a9d0de2a-8e98-4f75-98a8-673078841030/content

Khoury CK, Sotelo S, Amariles D, and Hawtin G (2023) *The Plants That Feed the World: baseline data and metrics to inform strategies for the conservation and use of plant genetic resources for food and agriculture*. International Treaty on Plant Genetic Resources for Food and Agriculture Rome: Food and Agricultural Organization of the United Nations. doi: 10.4060/cc6876en. https://www.fao.org/documents/card/en/c/cc6876en

Khoury CK, Sotelo S, Hawtin G, Halewood M, Lopez Noriega I, and Lusty C (2025) Germplasm exchange: Thematic Study for *The Third Report on the State of the World's Plant Genetic Resources for Food and Agriculture*. Rome: Food and Agricultural Organization of the United Nations. doi: 10.4060/cd4850en.
https://doi.org/10.4060/cd4850en

Van Hintum T, Menting F, and Van Strien E (2011) Quality indicators for passport data in *ex situ* genebanks. *Plant Genetic Resources* 9(3): 478–485.
https://www.cambridge.org/core/product/identifier/S1479262111000682/type/journal_article.
https://doi.org/10.1017/S1479262111000682

Acknowledgements

The development of this document was funded by the German Federal Ministry of Agriculture, Food and Regional Identity (BMLEH) as part of the project *Mainstreaming Global Crop Conservation Strategies in Plant Treaty Processes* led by the Crop Trust. The Crop Trust cooperated with the Secretariat of the the International Treaty on Plant Genetic Resources for Food and Agriculture in the development of this document.