

Databases and information systems

Where plant genetic resources are recorded and how to query them

Identifier	SAMEN-PROV0001-P12
Author	Thomas Schüller, Syntagma Forschung Wien · ORCID 0009-0003-9799-6747
Contact	forschung@tschueller.com
Version and date	2.3 · 2026-09-28
Rights	Text all rights reserved; tools and data layers MIT
Disclaimer	Interfaces and query patterns according to the operators' documentation (retrieved 24.09.2026), not according to our own queries.

Abstract

Paper P04 describes the data chain of conservation, paper P11 the record at its beginning. This paper describes its end: the information systems in which plant genetic resources are recorded, and the way to query them. It arranges fourteen systems by scope and type of data, shows the identifiers that connect them, and sets out query patterns, interfaces and terms of use. It is based on the data layer `datenbanken.json`, on which the Database Guide also rests. All fourteen entries there carry the status “partly verified”: the information comes from the operators' documentation, not from our own queries.

1. The landscape

1.1 By scope

Figure 12 ‘Database landscape’ arranges the systems on three levels. **Nationally**, genebanks keep their holdings: in Austria genbank.at of AGES, in Germany GBIS/I of IPK Gatersleben, in the USA GRIN-Global. **At European level**, EURISCO brings the national inventories together; AEGIS is the virtual European genebank built on it. **Globally**, there are Genesys as the passport data portal, WIEWS as FAO's reporting and code system, GLIS as the issuing body for persistent identifiers, and several neighbouring resources that were not built for genebanks but are needed for conservation.

1.2 By type of data

System	Scope	Type of data	Interface
Genesys PGR	global	passport data	yes, with login
FAO WIEWS	global	reports, institutes	not documented
GLIS	global	identifiers (DOI)	yes, login only for registering
Seed Information Database	global	seed traits	not documented
GBIF	global	occurrences, taxonomy	yes, search without login
Crop Ontology	global	traits, vocabulary	yes
NCBI GenBank / ENA	global	sequences	yes, without login
IUCN Red List	global	threat status	yes, with token

System	Scope	Type of data	Interface
World Flora Online	global	taxonomy	yes, without login
EURISCO	Europe	passport data, traits	none documented
ECPGR AEGIS	Europe	network	no
GRIN-Global	USA	passport data, taxonomy, traits	not documented
IPK GBIS/I	Germany	passport data, ordering	BrAPI, details not verified
genbank.at	Austria	passport data, traits, ordering	not documented

“Not documented” means that the data layer records no documented interface, not that none exists.

2. What the systems do and do not do

2.1 Passport data catalogues

Genesys PGR of the Crop Trust is the largest record of ex situ accessions in the format of the Multi-Crop Passport Descriptors. On 24 September 2026 its home page gave around 4.54 million accessions; the Crop Trust's annual report 2025 speaks of over 4.46 million samples, around three quarters of the world holdings of about 5.9 million (paper P04). Genesys holds no material but forwards requests to the holding genebank.

EURISCO, operated by IPK Gatersleben on behalf of ECPGR, records over two million accessions from over 400 institutes, with trait data and AEGIS labelling. According to ECPGR, **AEGIS** comprises around 122,000 accessions that are maintained according to agreed standards and supplied under the terms of the Seed Treaty; the details are in EURISCO.

Nationally, **genbank.at** supplies the Austrian holdings to EURISCO and offers passport, characterisation and evaluation data as well as sample requests. **IPK GBIS/I** records around 151,000 accessions of the German federal ex situ genebank and supplies material under a material transfer agreement. **GRIN-Global** holds over 600,000 active accessions of the US genebank system and the taxonomy GRIN Taxonomy.

None of these catalogues says anything about whether a sample is viable or available for distribution.

2.2 Identifiers and reports

FAO WIEWS assigns the institute codes (INSTCODE), records over 17,000 institutes and carries the State of the World reports and sustainable development indicator 2.5.1. **GLIS**, the information system of the International Seed Treaty, assigns DOIs for plant genetic material free of charge; according to its home page at the time of retrieval, around 2.1 million.

2.3 Neighbouring resources

- **Seed Information Database** (Society for Ecological Restoration with Kew): storage behaviour, seed weight, germination conditions; according to the operator, around 182,000 records on around 54,800 taxa. It answers whether a species can go into the seed bank (papers P02 and P07).
- **GBIF**: occurrence and specimen data, including crop wild relatives.
- **IUCN Red List**: version 2026-1 with around 176,000 assessed species, of which over 49,500 are threatened (paper P18).
- **World Flora Online**: global consensus classification with name matching (paper P13).

- **Crop Ontology:** uniform terms for traits, methods and scales per crop, so that trait data from different genebanks become comparable.

3. Identifiers as a bridge

The systems are not linked technically but through identifiers.

Identifier	Assigned by	Unique	Used in
INSTCODE, e.g. AUT001	FAO WIEWS	globally	all passport data catalogues
Accession number (ACCENUMB)	holding institute	only with INSTCODE and genus	Genesys, EURISCO, national systems
DOI, prefix 10.18730	GLIS	globally, persistent	Genesys, EURISCO (if reported), IPK
Taxon identifiers (WFO ID, taxonKey, NCBI Taxonomy ID, IUCN taxon ID)	each system	only within its own system	WFO, GBIF, INSDC, IUCN

The institute code consists of the three-letter country code under ISO 3166 and a number and is the key of the data model (paper P04). The DOI via GLIS replaces the composite key with a single persistent identifier, which is entered in the passport data field **PUID** (paper P11). For names there is no such bridge: the transition between the taxon identifiers runs via the scientific name and is prone to synonyms.

4. Querying step by step

The data layer assigns nine questions to the systems. This results in a sequence:

1. **Clarify names** in World Flora Online or GRIN Taxonomy, note synonyms.
2. **Search for accessions**, first nationally (genbank.at, IPK GBIS/I), then in EURISCO and Genesys.
3. **Resolve identifiers:** institute code via WIEWS, DOI via GLIS.
4. **Check the context:** storage behaviour in the Seed Information Database, occurrences in GBIF, threat status in the Red List.
5. **Request material** from the holding genebank, not from the catalogue.

The data layer gives query patterns for nine systems. Only patterns that were checked against the official documentation are marked as verified; they, too, are not based on a query of our own.

System	Purpose	Pattern	Verified
Genesys	overview of a genebank	https://www.genesys-pgr.org/wiews/{INSTCODE}	yes
GLIS	resolve DOI	https://doi.org/10.18730/{DOI-SUFFIX}	yes
WIEWS	institutes and codes	https://www.fao.org/wiews/data/organizations/en/	yes
GRIN-Global	taxonomy search	https://npgsweb.ars-grin.gov/gringlobal/taxon/taxonomysearch	yes

System	Purpose	Pattern	Verified
World Flora Online	match names	<code>https://list.worldfloraonline.org/matching_rest.php?input_string={NAME}</code>	yes
GBIF	occurrences, machine-readable	<code>https://api.gbif.org/v1/occurrence/search?scientificName={NAME}&country={LAND2}</code>	yes
GBIF	occurrences in the web portal	<code>https://www.gbif.org/occurrence/search?q={NAME}</code>	no
IUCN	look up species, with token	<code>https://api.iucnredlist.org/api/v4/taxa/scientific_name?genus_name={GATTUNG}&species_name={ART}</code>	yes
Crop Ontology	variables of an ontology	<code>https://croponontology.org/brapi/v1/variables/{ONTOLOGIE}</code>	yes
ENA	sequence data of a taxon	<code>https://www.ebi.ac.uk/ena/portal/api/search?result=read_run&query=tax_tree({TAXID})</code>	yes
NCBI	nucleotide search	<code>https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=nuccore&term={NAME}</code>	no

For EURISCO, AEGIS, genbank.at, IPK GBIS/I and the Seed Information Database, the way is through the search form.

5. Interfaces and licences

Genesys offers an interface at `api.genesys-pgr.org` with authentication via OAuth2, as a login or with client credentials via the helpdesk; BrAPI also requires authentication. For R there is the package `genesysr`. The terms of use require attribution and prohibit passing the data on to third parties; a publication refers to the portal instead of to accompanying raw data.

EURISCO has no documented interface (as also stated in *Frontiers in Plant Science* 2026); search and export run via the web interface, and the terms of use are worldwide, royalty-free and non-exclusive. **IPK GBIS/I** documents BrAPI endpoints; the passport data snapshot 2025 is available under CC BY-SA 4.0. **GRIN-Global**, as a work of the US government, is in the public domain. **GLIS** requires a login only for registering and updating.

GBIF permits searching without a login, downloads with an account. The licence applies per dataset: CC0, CC BY 4.0 or CC BY-NC 4.0. If a query mixes these, the most restrictive determines what may be done with the result. **World Flora Online** is under CC0. The sequence archives are freely accessible; **NCBI** permits three requests per second, ten with a key. The **IUCN** requires a token and permits only non-commercial use with an obligation to cite.

6. Data quality

Duplicates. The same material is often held in several genebanks. The fields `DONORCODE`, `OTHERNUMB` and `DUPLSITE` record this, provided they are filled in (paper P11). In addition there is the duplication across systems: a record from genbank.at also appears in EURISCO and Genesys. Whoever adds up counts from different portals counts it several times.

Outdated taxonomy. Genesys relies on the taxonomy of the US genebank system (paper P04), other systems on their own lists. A name assigned years ago may be a synonym today; a search only for the current name misses it.

Missing coordinates. Collecting site and coordinates are recommended in the passport data format, not mandatory. Entries without them are counted but cannot be assigned to an area of origin. No measurement of their share is available to this strand.

Genebank data and occurrence data. A passport data record describes a sample that an institute holds. An occurrence record in GBIF describes that an organism was observed or documented at a place at a time. The one documents available material, the other a population that may be extinct today. For wild relatives, an occurrence without a genebank sample indicates a collecting gap.

Number of accessions and viable accessions are different quantities (paper P04): the Bavarian wild plant genebank found that in 21 per cent of accessions more than half of the seeds were empty, infested or dead. No catalogue shows this.

7. Sequence data

The archives of the International Nucleotide Sequence Database Collaboration (NCBI, EMBL-EBI, DDBJ) hold sequences, reads and assemblies and, under their policy, are freely accessible. This makes them the core of the debate on digital sequence information: benefit-sharing was designed for physical material, while the sequence is freely available (paper P03). The Governing Body of the Seed Treaty reached no consensus on this in November 2025. Under Decision 16/2 of the Convention on Biological Diversity, large users are to pay, as an indicative rate, 1 per cent of profits or 0.1 per cent of revenue into the Cali Fund; the mobilisation of funds remained far below expectations until 2026 (data layer `rech.json`).

The archives work with their own identifiers, the INSDC accession number and the NCBI Taxonomy ID. The data layer records no common identifier between genebank sample and sequence; the entry point is the taxon, as in the pattern for ENA.

8. The tool

The Database Guide starts from the question, not from the system: one selects one of the nine questions, obtains the relevant systems with interfaces, identifiers and terms of use, and has query addresses built from name, genus, species or institute code, for instance for Genesys, GBIF, ENA and the IUCN. It also explains the classification systems (paper P13). Own holdings in the passport data format are kept in the Collection Manager; the Collections Atlas shows collections and recording systems with data status and state of evidence.

Figures

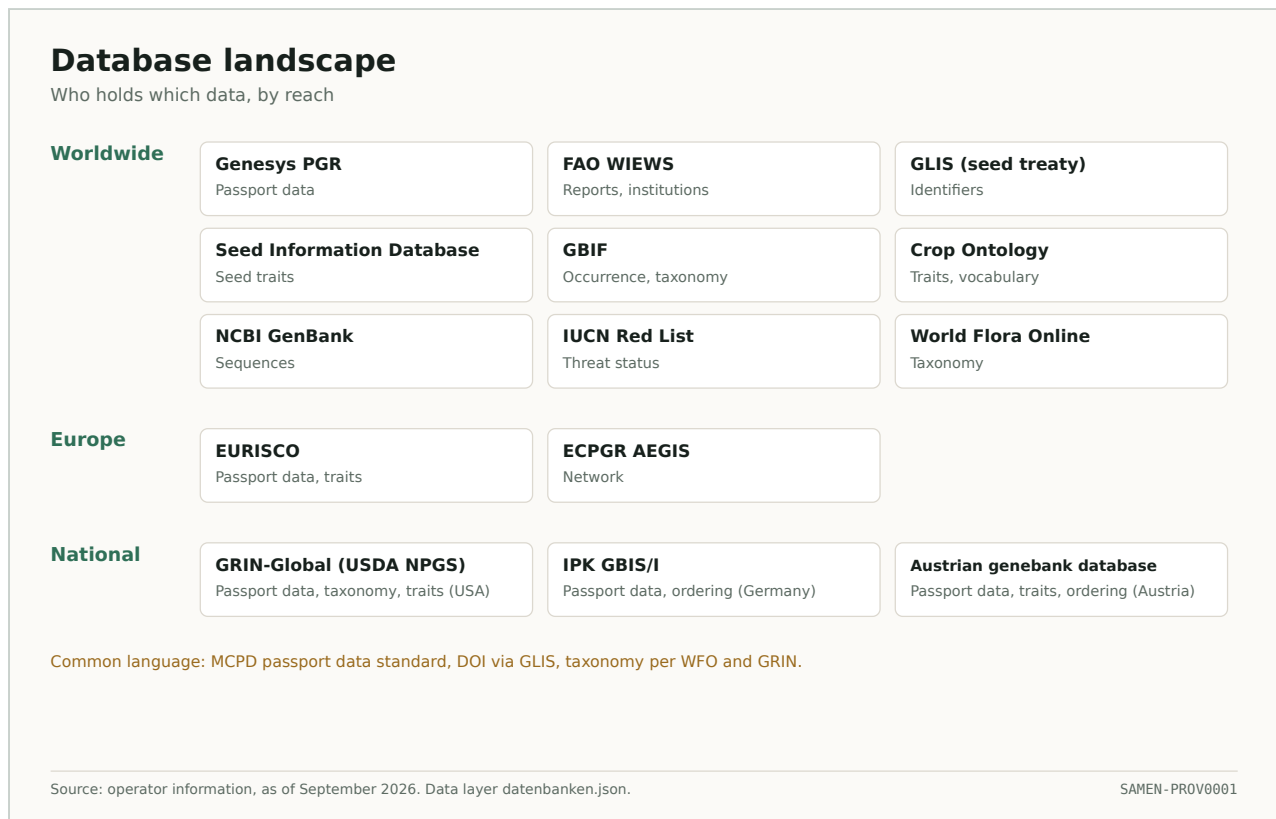


Figure 1. Database landscape of plant genetic resources. Source: data layer of the strand, see figure footer.

Limits of this paper

The paper reflects the state of the interfaces as of September 2026, compiled from the operators' documentation on 24 September 2026; all systems carry the status “partly verified”. Seven query patterns were additionally tested by retrieval: the patterns for Genesys (institute page), the GBIF programming interface and the ENA search returned valid responses; the World Flora Online pattern returned only the documentation, the Crop Ontology pattern a server error, and NCBI and the GBIF web interface do not allow automated retrieval. The result is shown in the Database Guide next to each pattern. Addresses, query patterns and terms of use change. Figures differ depending on source and date, as with Genesys. The selection of fourteen systems is not complete; further national inventories, herbarium databases and crop-specific systems are missing. The data quality of the catalogues was not checked.

References

1. Genesys PGR: *APIs and Terms and conditions of use*. <https://www.genesys-pgr.org/documentation/apis>; <https://www.genesys-pgr.org/content/legal/terms>
2. Crop Trust, Annual Report 2025: *Digitizing genebanks — Genesys and GGCE*. <https://report.croptrust.org/2025/our-work/digitizing-genebanks-genesys-and-ggce/>
3. IPK Gatersleben: *Genebank documentation*. <https://www.ipk-gatersleben.de/en/research/genebank/genebank-documentation>

4. Aguilar, C. H. M.; Oppermann, M.; Weise, S. (2026): FAIRness at all costs? Practical limitations in applying the FAIR principles to plant genetic resources data. *Frontiers in Plant Science* 17, 1888670. <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2026.1888670/pdf>
5. ECPGR: *AEGIS — European accessions*. <https://www.ecpgr.org/aegis/european-collection/european-accessions>
6. IPK Gatersleben: *Information and ordering; Breeding API*. <https://www.ipk-gatersleben.de/en/infrastructure/gene-bank/information-and-ordering>; <https://fair-ipk.ipk-gatersleben.de/public/breedingapi.html>
7. AGES: *Österreichische Genbank* (Austrian Genebank). <https://www.genbank.at/en/>
8. USDA Agricultural Research Service: *GRIN*. <https://www.ars-grin.gov/>
9. FAO: *WIEWS*. <https://www.fao.org/wiews/en/>
10. International Seed Treaty: *GLIS*, home page and FAQs. <https://glis.fao.org/glis/>; <https://glis.fao.org/glis/static/en/faqs.html>
11. Society for Ecological Restoration: *Seed Information Database*. <https://ser-insr.org/seed-information-database>
12. GBIF: *API documentation; Terms of use*. <https://techdocs.gbif.org/en/openapi/>; <https://www.gbif.org/terms>
13. Crop Ontology: *API help*. https://croponontology.org/api_help
14. EMBL-EBI: *ENA policies*. <https://ena-browser-docs.readthedocs.io/en/latest/about/policies.html>
15. NCBI: *Entrez Programming Utilities Help*. <https://www.ncbi.nlm.nih.gov/books/NBK25497/>
16. IUCN: *Red List API; The IUCN Red List of Threatened Species*. <https://api.iucnredlist.org/>; <https://www.iucnredlist.org/>
17. World Flora Online: *Background*. <https://wfoplantlist.org/background>
18. CBD: *Decision 16/2*. <https://www.cbd.int/doc/decisions/cop-16/cop-16-dec-02-en.pdf>
19. Papers P02, P03, P04, P07, P11, P13 and P18 of this strand.