

Data layers of conservation

Passport data, catalogues and interoperability

Identifier	SAMEN-PROV0001-P04
Author	Thomas Schüller, Syntagma Forschung Wien · ORCID 0009-0003-9799-6747
Contact	forschung@tschueller.com
Version and date	1.2 · 2026-09-22
Rights	Text all rights reserved; tools and data layers MIT

Abstract

A seed sample without passport data is not conserved, only stored. This paper describes the data architecture of plant genetic conservation — from the accession through the national inventories to the worldwide catalogues —, names the point at which the chain breaks, and derives from this the rules according to which the data layers of this strand are constructed.

1. The chain

The conservation data are organised in five levels, each of which builds on the one below it.

Level 1 — the accession. The smallest unit is the sample, which a genebank holds under an accession number of its own. The number is unique within the institution and usually consists of a prefix, a serial number and occasionally an addition distinguishing subsamples of the same original material.

Level 2 — the passport data. The description format is the Multi-Crop Passport Descriptors. They determine which fields a sample carries as a minimum. For an entry in a catalogue, three items suffice: the accession number, the genus and the institute code according to the FAO register. Everything further — species, variety name, biological status, country of origin, collecting site including coordinates, collecting date, source — considerably increases the value of the entry.

Level 3 — the institute identifier. The FAO institute code is the key through which all systems speak to one another. It holds the data model together: an accession number is unique only in combination with it.

Level 4 — the national inventories. Each country maintains a register of its holdings. In Austria this is genbank.at, in which, alongside the AGES genebank, holdings of civil society bodies are also documented.

Level 5 — the aggregating catalogues. EURISCO brings together the national inventories of Europe and documents more than two million accessions from more than 400 institutes. Genesys brings together EURISCO and other networks and contains more than four million accessions from over 500 collections. Alongside these, the FAO's World Information and Early Warning System WIEWS carries the reporting obligations, among them those for Sustainable Development Goal target 2.5.1, and maintains the institute codes.

Alongside these levels run two further reference systems: the taxonomic basis, for which Genesys uses the taxonomy of the US germplasm system, and the climatic conditions at the collecting site, which are supplied via WorldClim. Both are examples of the fact that the data layer of a genebank comprises more than its own records.

2. Where the chain breaks

According to the Crop Trust annual report 2025, Genesys lists over 4.46 million samples, about three quarters of global ex situ holdings. Around a quarter exists physically but cannot be found through Genesys. The Third Report explicitly names inadequate documentation as a bottleneck that hinders research and breeding.

The gap has several causes, which differ in their effect:

- **Unrecorded holdings.** Collections that were never fed into a national inventory, mostly for lack of staff.
- **Recorded but unpublished holdings.** Institutions that hold their data but do not pass them on. The ARCHE NOAH association gives a substantive reason for this: as long as the seed archive is being continuously reviewed for conservation priorities and duplicates, only part of the holdings is in the EURISCO inventory.
- **Recorded but unusable data.** Entries without a statement of origin or without taxonomic assignment. They appear in the count and contribute nothing to findability.
- **Recorded data on dead material.** The most serious case: the catalogue shows an accession that has become non-viable. The Bavarian wild plant genebank has disclosed this for its holdings — in 21 per cent of the accessions examined, more than half of the seeds were empty, infested or dead. A holdings figure without a statement on germination is a quantity figure, not a statement about conserved diversity.

From this follows the most important reading rule for all figures in this field: **accession count, safety-duplicated accessions and viable accessions are three different quantities.** Of the worldwide holdings of some 5.9 million accessions, about 41 per cent were safety-duplicated at the end of 2022, and 24 per cent are awaiting regeneration. The three figures refer to the same objects and say different things.

3. Rules of the data layers of this strand

From this finding follow the rules according to which the data files of this package are constructed. They are at the same time a proposal for comparable undertakings.

Separation of substantive data and presentation. The canonical data files (`lebensdauer.json`, `sammlungen.json`, `recht.json`) contain all figures and texts. The presentation layer contains none. The loader that connects the two is generated and never changed by hand. In this way every displayed figure can be traced back to a line in a single file.

Date on the figure, not on the document. Every indicator carries its own status date. The AGES figure is from 7 January 2026, that of the seed vault from 17 June 2026, that of the Millennium Seed Bank from 1 March 2023. A common document date would conceal this difference.

Evidence or marking. Every entry refers to an item of evidence. Where none is available, the status `ungeprueft` is given together with a note that the figure is to be obtained from the source before publication. The tool displays this status; a filter hides unverified figures.

No estimates in place of missing values. Where a figure is missing, `null` is given. In version 0.1 the Millennium Seed Bank was therefore included with its role only; the holdings figure was obtained from the source in version 1.0 and added — with its own date, not that of the file.

Parity of languages in the data layer. Role descriptions and statements are held in German and English versions side by side in the same data record, not in two separate files that can drift apart.

4. Transferability

The three data files are constructed so that they remain usable without the rest of the package. They are available under the MIT licence and can be taken over into evaluations of one's own. For such adoption, three points are to be observed:

1. Carry the status indication with the data. Passing on a figure from these holdings without its status reverses the purpose of the marking.
2. Carry the status date with the data. The holdings change with every deposit.
3. Carry the list of evidence with the data. It is part of the file, not an accessory.

Figures

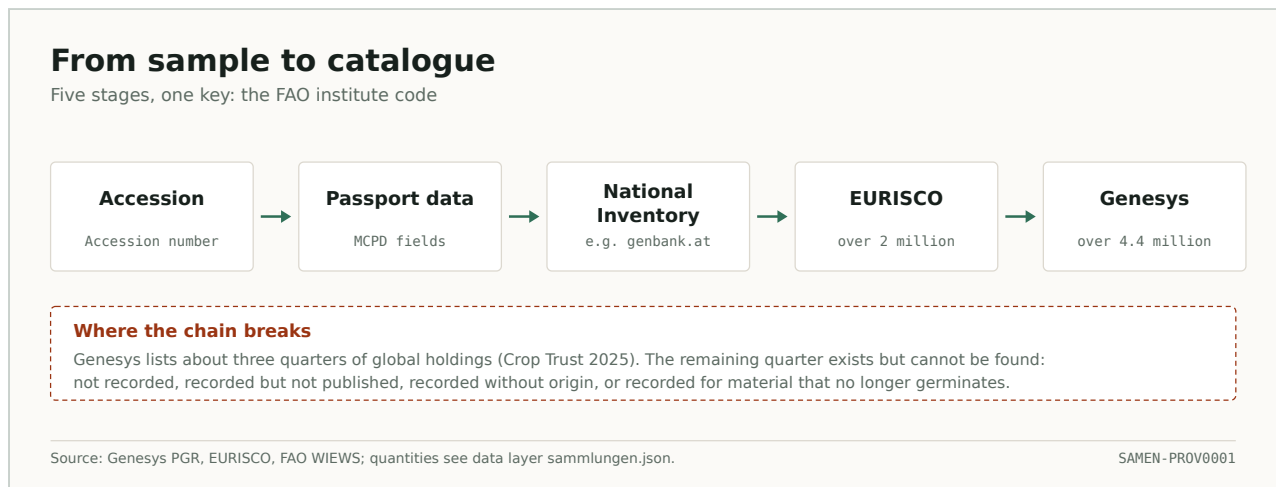


Figure 1. Data flow from accession to global catalogue. Source: data layer of the strand, see figure footer.

Limits of this paper

The paper describes the architecture from the perspective of the publicly accessible documentation of the systems. An examination of the actual data quality in the catalogues — for example of the share of entries with complete statements of origin — was not carried out and would be a separate work step. The statements on Austria's national inventory rest on the information provided by the participating institutions, not on a query of the register of our own.

References

1. Genesys PGR: *Accession passport data basics*. <https://www.genesys-pgr.org/documentation/basics>
2. Genesys PGR: *About Genesys and Data providers to Genesys*. <https://www.genesys-pgr.org/content/about/about>
3. Genesys PGR: *Frequently asked questions*. <https://www.genesys-pgr.org/content/about/frequently-asked-questions>
4. FAO: *WIEWS — World Information and Early Warning System on Plant Genetic Resources for Food and Agriculture*. <https://www.fao.org/wIEWS/resources/en>
5. FAO (2025): *The Third Report on the State of the World's Plant Genetic Resources for Food and Agriculture*. Rome.
6. Bioversity International / FAO: *Multi-Crop Passport Descriptors (MCPD)*.
7. ARCHE NOAH: *Sortenbestände und Sortendaten*. <https://www.arche-noah.at/sortenerhaltung/samenarchiv/sortendaten/>
8. Bayerisches Landesamt für Umwelt: *Genbank Bayern Arche*. https://www.lfu.bayern.de/natur/bayaz/artenschutz_pflanzen/artenhilfsprogramme_gefaesspflanzen/genbank_bayern_arche/index.htm
9. AGES: *Informationen zu Biodiversität*, as of 07.01.2026. <https://www.ages.at/pflanze/biodiversitaet/informationen-zu-biodiversitaet>