

Datenschichten der Erhaltung

Passdaten, Kataloge und Interoperabilität

Kennung	SAMEN-PROV0001-P04
Verfasser	Thomas Schüller, Syntagma Forschung Wien · ORCID 0009-0003-9799-6747
Kontakt	forschung@tschueller.com
Fassung und Stand	1.2 · 2026-09-22
Rechte	Text alle Rechte vorbehalten; Werkzeuge und Datenschichten MIT

Kurzfassung

Ein Saatgutmuster ohne Passdaten ist nicht erhalten, sondern nur aufbewahrt. Dieses Papier beschreibt die Datenarchitektur der pflanzen genetischen Erhaltung — von der Akzession über die nationalen Inventare bis zu den weltweiten Katalogen —, benennt die Stelle, an der die Kette reißt, und leitet daraus die Regeln ab, nach denen die Datenschichten dieses Strangs aufgebaut sind.

1. Die Kette

Die Erhaltungsdaten sind in fünf Stufen organisiert, von denen jede auf der darunterliegenden aufsetzt.

Stufe 1 — Die Akzession. Kleinste Einheit ist das Muster, das eine Genbank unter einer eigenen Akzessionsnummer führt. Die Nummer ist innerhalb der Einrichtung eindeutig und besteht üblicherweise aus einem Vorsatz, einer laufenden Nummer und gelegentlich einem Zusatz, der Teilproben desselben Ausgangsmaterials unterscheidet.

Stufe 2 — Die Passdaten. Das Beschreibungsformat sind die Multi-Crop Passport Descriptors. Sie legen fest, welche Felder ein Muster mindestens trägt. Für einen Eintrag in einen Katalog genügen drei Angaben: die Akzessionsnummer, die Gattung und der Einrichtungscode nach dem Verzeichnis der FAO. Alles Weitere — Art, Sortenname, biologischer Status, Herkunftsland, Sammelort samt Koordinaten, Sammeldatum, Quelle — erhöht den Wert des Eintrags erheblich.

Stufe 3 — Die Einrichtungskennung. Der Einrichtungscode der FAO ist der Schlüssel, über den alle Systeme miteinander sprechen. Er hält das Datenmodell zusammen: Eine Akzessionsnummer ist nur zusammen mit ihm eindeutig.

Stufe 4 — Die nationalen Inventare. Jedes Land führt ein Verzeichnis seiner Bestände. In Österreich ist das genbank.at, in dem neben der AGES-Genbank auch Bestände zivilgesellschaftlicher Träger nachgewiesen sind.

Stufe 5 — Die zusammenführenden Kataloge. EURISCO fasst die nationalen Inventare Europas zusammen und weist über zwei Millionen Akzessionen von mehr als 400 Instituten nach. Genesys führt EURISCO und andere Netze zusammen und enthält mehr als vier Millionen Akzessionen aus über 500 Sammlungen. Das Informations- und Frühwarnsystem WIEWS der FAO trägt daneben die Berichtspflichten, unter anderem zum Nachhaltigkeitsziel 2.5.1, und pflegt die Einrichtungscodes.

Neben diesen Stufen laufen zwei weitere Bezugssysteme: die taxonomische Grundlage, für die Genesys die Taxonomie des US-amerikanischen Germplasmsystems verwendet, und die klimatischen Bedingungen am Sammelort, die über WorldClim zugespielt werden. Beide sind Beispiele dafür, dass die Datenschicht einer

Genbank nicht nur ihre eigenen Aufzeichnungen umfasst.

2. Wo die Kette reißt

Genesys weist nach dem Jahresbericht 2025 des Crop Trust über 4,46 Millionen Muster nach, rund drei Viertel des weltweiten Ex-situ-Bestands. Rund ein Viertel existiert physisch, ist aber über Genesys nicht auffindbar. Der dritte Weltzustandsbericht nennt mangelhafte Dokumentation ausdrücklich als Engpass, der Forschung und Züchtung behindert.

Die Lücke hat mehrere Ursachen, die sich in ihrer Wirkung unterscheiden:

- **Nicht erfasste Bestände.** Sammlungen, die nie in ein nationales Inventar eingespeist wurden, meist aus Personalmangel.
- **Erfasste, aber nicht veröffentlichte Bestände.** Einrichtungen, die ihre Daten halten, aber nicht weitergeben. Der Verein ARCHE NOAH nennt dafür einen sachlichen Grund: Solange das Samenarchiv laufend auf Erhaltungsprioritäten und Duplikate durchgesehen wird, ist nur ein Teil des Bestands im EURISCO-Inventar.
- **Erfasste, aber unbrauchbare Daten.** Einträge ohne Herkunftsangabe oder ohne taxonomische Zuordnung. Sie erscheinen in der Zählung und tragen nichts zur Auffindbarkeit bei.
- **Erfasste Daten zu totem Material.** Der gravierendste Fall: Der Katalog weist eine Akzession aus, die keimunfähig geworden ist. Die bayerische Wildpflanzen-Genbank hat das für ihren Bestand offengelegt — bei 21 Prozent der untersuchten Akzessionen waren mehr als die Hälfte der Samen leer, befallen oder tot. Eine Bestandszahl ohne Angabe zur Keimfähigkeit ist eine Mengenangabe, keine Aussage über erhaltene Vielfalt.

Daraus folgt die wichtigste Leseregeln für alle Zahlen dieses Feldes: **Akzessionszahl, sicherheitsduplizierte Akzessionen und keimfähige Akzessionen sind drei verschiedene Größen.** Vom weltweiten Bestand von rund 5,9 Millionen Akzessionen waren Ende 2022 rund 41 Prozent sicherheitsdupliziert, und 24 Prozent warten auf Regeneration. Die drei Zahlen beziehen sich auf dieselben Objekte und sagen Verschiedenes.

3. Regeln der Datenschichten dieses Strangs

Aus dem Befund ergeben sich die Regeln, nach denen die Datendateien dieses Pakets aufgebaut sind. Sie sind zugleich ein Vorschlag für vergleichbare Vorhaben.

Trennung von Sachdaten und Darstellung. Die kanonischen Datendateien (`lebensdauer.json`, `sammlungen.json`, `recht.json`) enthalten alle Zahlen und Texte. Die Darstellungsschicht enthält keine. Der Loader, der beides verbindet, wird erzeugt und nie von Hand geändert. Damit lässt sich jede angezeigte Zahl auf eine Zeile in einer einzigen Datei zurückführen.

Datum an der Zahl, nicht am Dokument. Jede Kennzahl trägt ihren eigenen Stand. Die Angabe der AGES stammt vom 7. Jänner 2026, die des Saatguttresors vom 17. Juni 2026, die des Millennium Seed Bank vom 1. März 2023. Ein gemeinsames Dokumentdatum würde diesen Unterschied verdecken.

Beleg oder Kennzeichnung. Jeder Eintrag verweist auf einen Beleg. Wo keiner vorliegt, steht der Status `ungeprüft` und ein Hinweis, dass die Zahl vor Veröffentlichung an der Quelle zu erheben ist. Das Werkzeug zeigt diesen Status an; ein Filter blendet ungeprüfte Zahlen aus.

Keine Schätzungen anstelle fehlender Werte. Wo eine Zahl fehlt, steht `null`. In Fassung 0.1 war das Millennium Seed Bank deshalb nur mit seiner Rolle aufgenommen; die Bestandszahl wurde in Fassung 1.0 an der Quelle erhoben und nachgetragen — mit ihrem eigenen Stand, nicht mit dem der Datei.

Parität der Sprachen in der Datenschicht. Rollenbeschreibungen und Aussagen liegen in deutscher und englischer Fassung nebeneinander im selben Datensatz, nicht in zwei getrennten Dateien, die auseinanderlaufen können.

4. Übertragbarkeit

Die drei Datendateien sind so aufgebaut, dass sie ohne den Rest des Pakets brauchbar bleiben. Sie stehen unter der MIT-Lizenz und lassen sich in eigene Auswertungen übernehmen. Für eine Übernahme sind drei Punkte zu beachten:

1. Die Statusangabe mitführen. Eine Zahl aus diesem Bestand ohne ihren Status weiterzugeben, kehrt den Zweck der Kennzeichnung um.
2. Den Stand mitführen. Die Bestände ändern sich mit jeder Einlagerung.
3. Die Belegliste mitführen. Sie ist Teil der Datei, nicht Beiwerk.

Abbildungen



Abbildung 1. Datenfluss von der Akzession zum weltweiten Katalog. Quelle: Datenschicht des Strangs, siehe Fußzeile der Abbildung.

Grenzen dieses Papiers

Das Papier beschreibt die Architektur aus der Sicht der öffentlich zugänglichen Dokumentation der Systeme. Eine Prüfung der tatsächlichen Datenqualität in den Katalogen — etwa des Anteils der Einträge mit vollständigen Herkunftsangaben — wurde nicht vorgenommen und wäre ein eigener Arbeitsschritt. Die Aussagen zum nationalen Inventar Österreichs stützen sich auf die Angaben der beteiligten Einrichtungen, nicht auf eine eigene Abfrage des Verzeichnisses.

Literatur

1. Genesys PGR: *Accession passport data basics*. <https://www.genesys-pgr.org/documentation/basics>
2. Genesys PGR: *About Genesys und 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*. Rom.
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*, Stand 07.01.2026. <https://www.ages.at/pflanze/biodiversitaet/informationen-zu-biodiversitaet>