

V. GOLLKOWSKI, Oelsnitz i.V.

Gedanken über eine entomologische Datenbank

Summary This publication proposes a database structure for entomological purposes based on the relational model. Hints for stepwise development of a data bank program are provided. The most important information for the format of entomological data is also provided.

Résumé Cette publication propose une structure pour une banque de données pour l'entomologie. Cette structure base sur un modèle relationnel. On donne des informations concernant les démarches nécessaires pour établir un programme d'une banque de données. Au plus, on présente une liste des informations plus importantes pour la préparation des données entomologiques.

Aufgrund hoher Leistungsfähigkeit und niedriger Anschaffungskosten der Personal-Computer gibt es immer mehr Entomologen, die von der klassischen Kartei auf eine rechnergestützte Fundortverwaltung umsteigen. Dafür sind allerdings zur Zeit nur sehr wenige Programme verfügbar, die speziell auf die Anforderungen der Insektenkunde zugeschnitten sind. Ein bereits existierendes Datenbank- und Grafikpaket, SoftCol von Herrn STRAUSS aus Biberach, wird in einem der letzten Hefte der „Mitteilungen der Arbeitsgemeinschaft Rheinischer Koleopterologen“, Bonn, vorgestellt. Wichtige Aussagen zur Computerfaunistik finden sich bei MÜLLER-MOTZFELD (1990). Besonders bedeutungsvoll für die mitteleuropäische Käferfaunistik ist der Katalog von LUCHT (1987), eine edv-gerechte Auflistung aller vorhandenen Arten in systematischer Reihenfolge.

Wer sich mit Hilfe eines der vorhandenen Standarddatenbank-Systeme sein eigenes Faunistikprogramm erstellen will, kann zum Beispiel zwischen Paradox, dBASE, FoxPro oder Superbase wählen. Entscheidend für die Auswahl sollten Punkte wie hohe Arbeitsgeschwindigkeit sowie gute Austausch- und vielfältige Bearbeitungsmöglichkeiten der Funddaten sein.

Zuerst ist zu klären, welche Informationen in der Datenbank abgespeichert werden sollen. Die wichtigsten Angaben zur genauen Fund- und Artbeschreibung sind nachstehend aufgelistet:

- Gattungsname
- Artname
- Familienzugehörigkeit
- Autor der Art
- Einklammerung des Autorennamens (ursprüngliche Beschreibung in einer anderen als der jetzigen Gattung)

- Jahresangabe der Artbeschreibung
- Staat
- , Provinz (z.B. Bundesland, Kanton)
- Kreis
- Landschaftsgebiet
- Bezugsort
- genauer Fundort
- Meereshöhe des Fundortes
- Koordinaten des Fundortes
- Fundumstände
- Funddatum
- Anzahl
- Sammler
- Sammlung
- Aufbewahrungsort
- Determination
- Bestimmungsjahr
- Kommentar

Diesen Datensatz muß der einzelne Sammler gegebenenfalls noch seinen speziellen Bedürfnissen anpassen. So ist zum Beispiel eine getrennte Abspeicherung der Anzahl der festgestellten Männchen und Weibchen, der Fraßpflanze oder der Sammelmethode denkbar. Wer nur seine eigenen Aufsammlungen erfassen will, kann eventuell auf Angaben zu Sammler und Sammlung verzichten. Die Angaben zum Fundort wurden in einen Bezugsort, der eine größere, in der Nähe gelegene Stadt sein sollte, und den genauen Fundort aufgeteilt. Letzterer muß nicht immer ein Ort sein, es kann sich dabei auch um eine Talsperre, einen Berg oder ähnliches handeln – daher die Zweiteilung. Bei den Koordinaten des Fundortes ist die Angabe einer der verschiedenen Rastereinteilungen (Meßtischblatt-Quadranten) oder auch der geografischen Längen- und Breitenangabe möglich. Durch das Abspeichern der Koordinaten kann später sehr einfach das Darstellen von Verbreitungskarten erfolgen.

Jetzt geht es an den Entwurf der Datenbankstruktur. Man könnte natürlich eine Datei einrichten, die gleich alle Angaben direkt in Form entsprechender Felder enthält. Das wäre allerdings nicht nur eine große Platzverschwendung, sondern auch ein Verstoß gegen das relationale Datenbankmodell, das hier Anwendung finden soll. Ein Hauptmerkmal dieses Modelles ist das Vermeiden der Wiederholung gleicher Informationen. Insbesondere müssen drei Regeln (Normalformen) gelten, damit keine Verstöße gegen das Relationenmodell vorliegen. Um das zu erreichen, muß die oben angeführte Informationsmenge einer Normalisierung (Dekomposition) unterzogen werden, was letztendlich bedeutet, daß die Gesamtmenge der Angaben auf jeweils zusammengehörige Teilmengen und damit auf mehrere Dateien aufgeteilt wird.

So fällt sofort auf, daß von einer Art Exemplare von verschiedenen Fundorten vorliegen können und andererseits von einem Fundort mehrere Arten, ein Sammler kann an vielen Stellen gesammelt haben usw. Ohne Dekompensation müßten jetzt jedesmal alle Eingaben zu Fundort, Art und den anderen Angaben vollständig wiederholt werden. Dabei können natürlich auch Eingabefehler entstehen, die eine spätere exakte Auswertung der Daten verhindern. Es ist also leicht ersichtlich, daß getrennte Dateien für Fundort, Art, Sammler, Sammlung, Determination und Fundumstände angelegt werden müssen. Zur Aufteilung der Gesamtinformation ist ein Fremdschlüsseleinschluß erforderlich. Das bedeutet, daß zusätzliche Felder zur Verknüpfung der zugeordneten Dateien zur Hauptdatei einzufügen sind. Jeder Fundort, jede Art usw. bekommt also ein eigenes, eindeutiges Zuordnungsfeld, eine Nummer. In der Hauptdatei ist nicht mehr die Ortsangabe selbst, sondern die auf die Ortsangabe verweisende Nummer abgelegt. Nun kann man denken, daß durch das Einfügen des Zuordnungsfeldes die Datenbank noch vergrößert wird. Wenn zum Beispiel ein Fundort nur einmal vorkommt, ist das auch der Fall. Sobald aber der gleiche Fundort mehrfach auftritt, kann wesentlich an Platz gespart werden, da in der Hauptdatei immer nur die zugehörige Nummer vorliegt, die wesentlich kürzer ist als die vollständige Ortsangabe. Ein Name kann sowohl für die Determination als auch als Sammler in Frage kommen, oftmals ist es sogar so, daß beide identisch sind. Deshalb reicht hier eine Namensdatei aus, auf die über zwei Zuordnungsfelder zugegriffen wird. Bei der Programmierung sind die Regeln des Systems für die doppelte Nutzung einer Datei zu beachten.

Denkt man in der aufgezeigten Richtung weiter, erkennt man, daß ähnliche Verhältnisse auf Staat, Provinz, Kreis und Landschaft bei dem Fundort sowie auf Familie, Gattung und Autor bei der Art zutreffen. Hier ist eine

weitere Aufteilung in einzelne, den jeweiligen Kategorien zugeordnete Dateien erforderlich. Eine entsprechende Verknüpfung ermöglicht es, daß die Zugehörigkeit einer Provinz zu einem Staat, eines Kreises zu einer Provinz und einer Gattung zu einer Familie nur einmal eingegeben werden muß und später immer wieder verfügbar ist. Ändert sich etwa der Name einer Gattung, braucht die Namensänderung lediglich an einer Stelle erfolgen. Ein wichtiges Problem der Datenaufteilung muß an dieser Stelle noch genannt werden, das Problem der referentiellen Integrität. In der Datenbank dürfen keine leeren Verweise auftreten. Wenn zum Beispiel für eine Käferart die Synonymie zu einer anderen festgestellt und die Art damit aus der Artdatei gelöscht wird, sind anschließend die bisher zu dieser Art gehörenden Nachweise ohne Artzugehörigkeit. Um so etwas zu vermeiden, sind konsistenzhaltende Nachfolge-Updates erforderlich, die leider bei den meisten Datenbanksystemen nicht automatisch erfolgen oder angemahnt werden, wodurch der Nutzer selbst auf dieses Problem achten muß.

Aus den bisher gemachten Ausführungen ergibt sich das in Bild 1 dargestellte Datenbankschema. Ein möglicher Kommentar wurde aus der Hauptdatei ausgegliedert, da er zumeist nur bei wenigen Fundangaben erforderlich ist, und somit keine unnötige Platzverschwendung erfolgt. Generell würde sich für den Kommentar ein sogenanntes Memo-Feld anbieten, soweit im Datenbanksystem verfügbar, allerdings hat dieser Feldtyp zumeist eine beträchtliche Größe. Jetzt sind noch der Datentyp und die Größe der einzelnen Felder festzulegen. An Datentypen sind zumeist Zahlen, Zeichen, Datumsangaben und logische Werte verfügbar. Es gilt zu überlegen, ob immer der Typ günstig ist, der für ein Feld zunächst als zutreffend erscheint. Für das Funddatum muß nicht immer die exakte Angabe vorhanden sein, aber vielleicht weiß man, daß der Käfer vom Ende des Monats, Anfang des Jahrzehnts oder Ende des Jahrhunderts stammt. Diese Information läßt sich nicht in einem Datumsfeld ablegen, hier sind nur gültige Datumsangaben zulässig. Deswegen wird für das Funddatum der Zeichentyp gewählt, mit dem sich die verschiedenen Kombinationen realisieren lassen. Im Voraus ist exakt festzulegen, was welche Kombination bedeutet. Ein gutes Beispiel dafür gibt MÜLLER-MOTZFELD (1990). Entsprechendes gilt für das Abspeichern der Anzahl der nachgewiesenen Exemplare.

Für die Verbindung der Haupt- mit der Artdatei ist eine Artnummer erforderlich. Für die mitteleuropäischen Käfer könnte hierfür direkt die Schlüsselzahl aus LUCHT (1987) eingesetzt werden. Kommen allerdings Fundangaben von nichtmitteleuropäischen Arten hinzu, ist die Schlüsselzahl nicht mehr als Verbindungsglied

nutzbar, ja für fast alle anderen Insektenordnungen existiert so ein einheitliches Schlüsselsystem überhaupt noch nicht. Weiterhin ist der Platzbedarf des Katalog-Codes mit elf Stellen wesentlich größer als der einer fortlaufenden Nummerierung, bei der für eine Millionen Arten sechs Stellen ausreichen. Dabei ist die Familiennummer des Kataloges bereits jetzt durch die moderne Systematik eingeholt, in HARDE & SEVERA (1988) werden 139 Käferfamilien aufgeführt, eine weitere Aufteilung ist abzusehen. Steigt die Zahl der Arten in den von UHLIG (1990) genannten Ausmaßen an, ist auch die Möglichkeit des Hinzufügens neuer Arten zum Katalogverzeichnis ohne zusätzliche Erweiterungen bald am Ende. Trotzdem hat der Katalog-Code einen wesentlichen Vorteil gegenüber einer einfachen Nummerierung, die Anordnung der Arten in systematischer Reihenfolge, was für die Durchsicht der Datenbank am Bildschirm im Vergleich mit der Literatur oder Katalogen sowie bei Nutzung der faunistischen Daten für Publikationen unerlässlich ist. Bei der durchgehenden Nummerierung wäre maximal eine Ordnung in alphabetischer Reihenfolge der Familien-, Gattungs- und Artnamen möglich. Um die Vorteile beider Nummerierungssysteme nutzen zu können, erfolgt das Abspeichern der Katalog-Schlüsselzahl innerhalb der Artdatei als eigenes Feld. Wenn sich jetzt die Zahl ändert, ist auch nur eine einzige Korrektur erforderlich. Für die anderen Ordnungen sollte baldigst ein gleichartiges System geschaffen werden, eventuell muß eine Erweiterung des Artenspektrums überdacht werden, so daß später nicht Entomologen verschiedener europäischer Gebiete ungleiche Nummern für die gleiche Art nutzen.

Sind die Aufgabenanalyse und der Entwurf des Datenbankaufbaus abgeschlossen, kann mit der eigentlichen Programmierung begonnen werden. Besondere Sorgfalt muß der Eingabe der Daten zukommen. Hier gemachte Fehler können später unter Umständen nicht wieder korrigiert werden, falsche Auswertungsergebnisse entstehen. Deswegen sind alle Eingaben auf deren Gültigkeit zu prüfen, für das Datum bedeutet das zum Beispiel die Kontrolle auf den Zahlenbereich oder eines der festgelegten Ersatzsymbole. Bei der Eingabe ist eine Übernahme bereits vorhandener Daten sinnvoll, wenn für eine Art eine Fundortliste oder für einen Ort mehrere Arten vorliegen. Durch die oben angeführte Verkettung der Fundort- und Artdaten wird die Fehlerwahrscheinlichkeit wesentlich verringert, das erfolgt auch durch den Einsatz von Auswahllisten. Sollte doch ein Eingabefehler passiert sein oder eine spätere Änderung erforderlich werden, etwa bei Namensänderung oder Fehlbestimmungen, müssen umfangreiche Korrekturmöglichkeiten unter Beachtung der genannten referentiellen Integrität vorhanden sein.

Der eigentliche Vorteil der Rechentechnik kommt bei der Datenauswertung, bei dem Erstellen von Sichten nach bestimmten Kriterien zum Vorschein. Während das Heraussuchen von speziellen Angaben aus einer Kartei Stunden oder Tage sowie oft eine Nachbearbeitung erfordert, liefert der Computer das Ergebnis in fertiger Form zumeist nach wenigen Sekunden. Es sind so viele Auswertungsmöglichkeiten denkbar, daß man sich bei der Programmierung auf häufig erforderliche Standardfälle beschränken sollte und spezielle Sichten dann erstellt, wenn sie benötigt werden. Oft genutzte Datenbankabfragen sind etwa die Angabe aller Arten für einen Ort, Kreis Provinz, Staat oder Landschaft, alle Fundorte für eine Art, die Datumsangaben für eine Art, alle Fundumstände einer Art oder die Zuordnung der Meereshöhe.

Damit ist die entomologische Datenbank einsatzbereit, eventuell ist noch eine Möglichkeit für den leichten Datenaustausch mit anderen Faunisten vorzusehen. Und vielleicht finden sich im Laufe der Nutzung noch Verbesserungen bei der Dateneingabe, um diese schneller und sicherer zu gestalten. Erfolgt die Realisierung unter Paradox 4.0, kann das dargestellte Datenbankschema direkt genutzt werden, weil hier eine Sicht aus bis zu 24 verknüpften Dateien bestehen kann, bei dBASE IV, 1.1, sind maximal acht möglich, so daß Einschränkungen im Aufbau hinzunehmen sind. Handelt es sich um sehr umfangreiche Datenmengen, ist es sinnvoll, den Übergang vom PC zu leistungsfähigerer Rechentechnik zu prüfen. Abschließend noch ein rechnerischer Größenvergleich: bei 10 000 Funddaten für 1 000 Arten (20 Familien, 200 Gattungen, 50 Autoren) von 100 Fundorten (5 Staaten, 10 Provinzen, 30 Kreise, 25 Landschaften, 100 Fundumstände, 15 Sammler, 20 Kommentare, 10 Sammlungen) wird für die Datenspeicherung mittels obiger Aufteilung eine Kapazität von 788 kByte in Anspruch genommen, bei Speicherung aller Daten in einer Datei sind es 6 836 kByte (Verhältnis 1:8,6). Bei der Arbeit mit der Datenbank kommen in beiden Fällen noch Indexdateien hinzu, die einen schnellen Zugriff und das Sortieren ermöglichen.

Literatur

- GRAFIK, W. (1990): dBASE III. – Berlin, München, 240 S.
 GRAFIK, W. (1992): Was kommt nach den PC-Datenbanken? – Mikroprozessortechnik, 6, H1.: 8 – 14.
 HARDE, K. W. & F. SEVERA (1988): Der Kosmos-Käferführer. – Stuttgart, 352 S.
 KLEIN, G. (1992): PC-Datenbanksysteme. – Mikroprozessortechnik, 6, H1.: 6–8.
 LUCHT, W. H. (1987): Die Käfer Mitteleuropas – Katalog. – Krefeld, 342 S.
 MÜLLER-MOTZFELD, G. (1992): Aufgaben und Möglichkeiten der Computerfaunistik. – Ent. Nachr. Ber., 34: 215–222.

ken. – Berlin, 264 S.

UHLIG, M. (1990): Erforschungsstand und Forschungstrends in der Kurzflüglergattung *Erichsonius* FAUVEL, 1874, – Überlegungen eines Entomologen zur Zahl der Tierarten der Weltfauna (Coleoptera, Staphylinidae, Philonthini). – Verh. Westd. Entom. Tag.: 121–146.

Anschrift des Verfassers:

Volker Gollkowski
Schmidtstraße 7a
08606 Oelsnitz i.V.

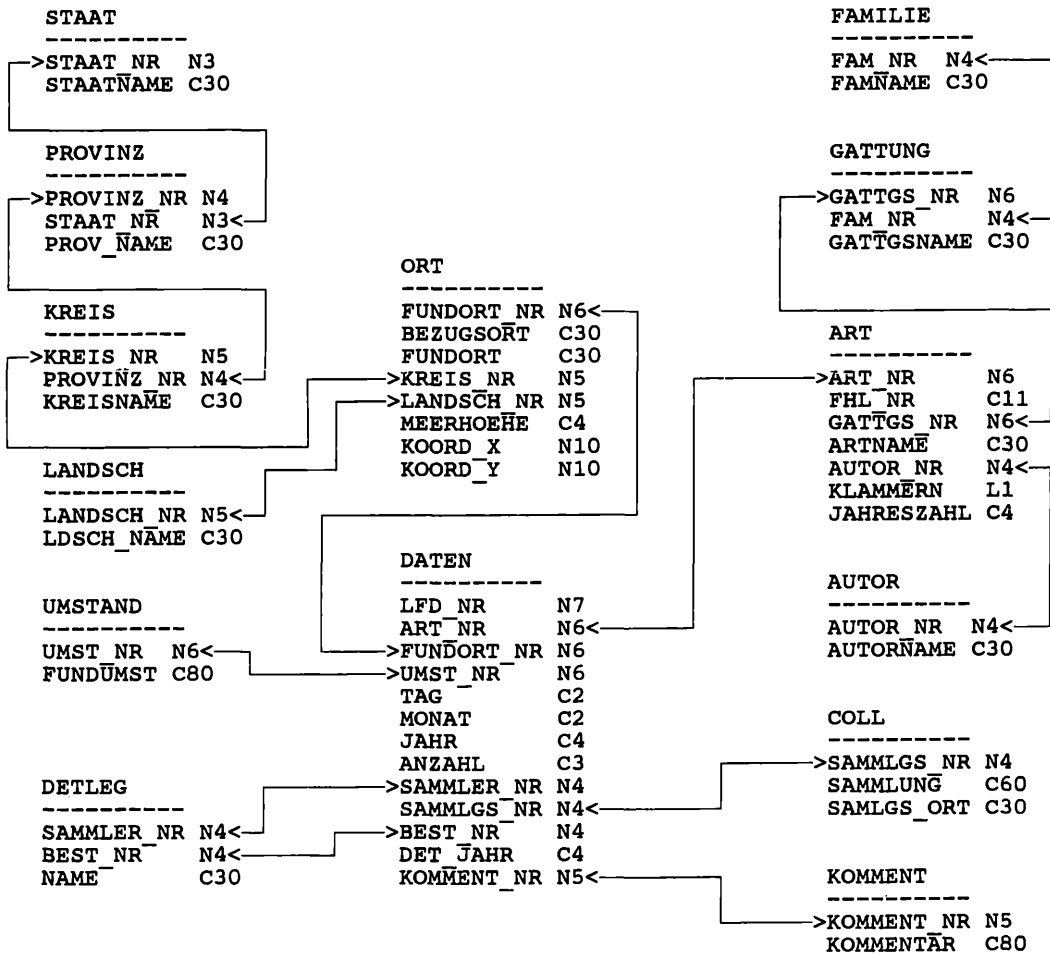


Bild 1: Schema für eine entomologische Datenbank. (Für die Feldnamen wurden datenbankgerechte Bezeichner gewählt, dahinter sind der Typ und die Feldgröße angegeben. N: Numerisch, C: Zeichen, L: Logisch)

ZOBODAT - www.zobodat.at

Zoologisch-Botanische Datenbank/Zoological-Botanical Database

Digitale Literatur/Digital Literature

Zeitschrift/Journal: [Entomologische Nachrichten und Berichte](#)

Jahr/Year: 1994

Band/Volume: [38](#)

Autor(en)/Author(s): Gollkowski Volker

Artikel/Article: [Gedanken über eine entomologische Datenbank. 121-124](#)