© Biologiezentrum Linz/Austria; download unter www.biologiezentrum.at

Linzer biol. Beitr. = Verh. XVIII. SIEEC, Linz 2003	37/1	49-69	25.7.2005

Die entscheidenden Faktoren für den langfristigen Erfolg einer tiergeografischen Datenbank - am Beispiel der ZOBODAT (früher: ZOODAT)

E. GEISER

A b s t r a c t : The critical factors for the permanent success of a zoogeographical database – explained by the Austrian biogeographical database ZOBODAT (former ZOODAT). Biogeographical databases are important tools für research purposes and a necessary base of decisions for protected areas, at national and international level. In the last decades many biogeographical databases of different range were installed in Europe. Some of them failed even at the start, many of them have a low data output despite of large money input.

In 1972, the zoogeographical database ZOODAT was installed by Ernst Reichl, a computer scientist in Linz. Soon it became to one of the most comprehensive zoogegraphical databases. Now named ZOBODAT, it still belongs to the top 3 of the most successful biogeographical databases of Europe.

This success can be explained by the following factors:

1 Use only the fields in the database, which are really neccessary.

2 Use a relational database management system with an optimized table design.

3 The user interface for the data input must be very ergonomic and optimized fot the special data structure of the organism group you intend to store.

4 Cooperation with the specialists of the different groups of organisms.

5 A pragmatical approach to the problems of systematics, based on the comprehensive and frequently used standard literature for the different groups of organisms.

6 Use the data frequently for research purposes and for different analysing programs. This factor causes

- permanent correction of data errors
- new data input sources of high quality
- new and interesting analysing programs

All these factors are discussed and examples are demonstrated on ZOBODAT.

K e y w o r d s : biogeographical database, biodiversity, biogeography, ergonomic software, problems of systematics

© Biologiezentrum Linz/Austria; download unter www.biologiezentrum.at

Inhalt

1 Einleitung 50
2 Zu welchem Zweck betreibt man eine tiergeografische Datenbank? 52
3 Der Mensch und seine begrenzten Fähigkeiten – ein häufig unterschätzter Faktor bei tiergeografischen Datenbanken 52
4 Welche Daten sollen und dürfen gespeichert werden? 53
4.1 Die Kunst der weisen Beschränkung auf das Wesentliche 53
4.2 Die Auswahl der wichtigen und richtigen Daten 54
5 Relationale Datenbanken, die adäquate Form der Datenspeicherung für tiergeografische Fragestellungen, und ihre historische Entwicklung 55
6 Wie speichert man Fundortangaben am besten in einer tiergeografischen Datenbank? 57
6.1 Die alte und überflüssige Debatte UTM-Gitter versus geografische Koordinaten 58
6.2 Die Erstellung von Fundortnummern als Schlüsselfelder in einer Datenbank: Vergleich eines üblichen Informatik-Konzeptes mit der optimalen Lösung in ZOODAT 60
7 Die Optimierung der Ergonomie der Dateneingabe: ein wesentlicher Kostenfaktor für den Betrieb einer Datenbank 62
8 Die Fallstricke der zoologischen Systematik 63
9 Die sozio-biologischen Faktoren: die Zusammenarbeit mit den Spezialisten bei der Dateneingabe und Auswertung 66
10 Zusammenstellung der wirklich entscheidenden Faktoren 67
11 Budgeteinsparung an der richtigen Stelle 67
12 Danksagung 68
13 Zusammenfassung 68
14 Literatur 68

1. Einleitung

Beim 18. SIEEC in Linz habe ich einen Vortrag mit dem Titel "Die entscheidenden Faktoren für den langfristigen Erfolg einer entomofaunistischen Datenbank – am Beispiel der ZOBODAT (früher: ZOODAT)" gehalten. Dabei konnte ich aber nur einige wenige Aspekte erläutern. Die anderen Faktoren habe ich im Rahmen verschiedener Vorträge auf nationalen und internationalen Tagungen behandelt, unter anderem bei einer Fachtagung aus Anlass des 30-jährigen Bestehens der ZOBODAT im September 2002 am Biozentrum in Linz, bei der International Senckenberg Conference zum Thema "Global Biodiversity Research in Europe" im Dezember 1996 in Frankfurt am Main und beim Colloquium des European Invertebrate Survey im September 2003 in Cardiff (Wales, United Kingdom).

Bei einem Vortrag kann man aus Zeitgründen immer nur einzelne Teilbereiche dieser umfangreichen Problematik erörtern. Daher halte ich es für sinnvoll, den Inhalt dieser Vorträge in einer einzigen Publikation zusammenzufassen. Jeder der hier erläuterten

© Biologiezentrum Linz/Austria; download unter www.biologiezentrum.at

Faktoren trägt ganz wesentlich zum Erfolg einer tiergeografischen Datenbank bei, jeder einzelne kann auch einen Misserfolg bewirken. Ich kenne seit mehr als 20 Jahren zahlreiche Datenbanken in Europa, tiergeografische und andere. Viele sind schon in der Konzeptphase gescheitert. Manche unternahmen schließlich einen zweiten Versuch, nachdem in der ersten Phase sehr viel Geld – meist aus Forschungsbudgets – verbraucht worden ist, bis man die wesentlichen Konzeptfehler erkannte. Anderen gelang der Start, aber seither vegetieren sie dahin anstatt zu florieren. Auch in diese Datenbanken wurden und werden viel Geld und – mindestens genauso tragisch – viel Idealismus und persönliches Engagement von einzelnen Zoologen gesteckt, die dann trotz jahrelanger Anstrengung nur magere Ergebnisse ernten.

Die ZOODAT hingegen entwickelte sich in kürzester Zeit zu einer der umfangreichsten tiergeografischen Datenbanken Europas. Aber nicht nur quantitativ, auch qualitativ ist sie bis heute im Spitzenfeld anzutreffen. Dabei hat sich in den letzten 3 Jahrzehnten nicht nur die Hardware, sondern auch das Informatik-Wissen rasant weiterentwickelt. Heute ist nicht nur der Computer ein ganz normales Arbeits- und sogar Haushaltsgerät, es gibt auch sehr viele und vor allem hervorragend ausgebildete Informatiker. Man könnte daher annehmen, dass es in den letzten Jahren wesentlich leichter geworden sei, eine biogeografische Datenbank zu konzipieren, zu implementieren und erfolgreich zu betreiben.

Die Informatik bzw. die Anwendung von Informatikkenntnissen macht aber nur einen Teil der Erfolgsfaktoren einer tiergeografischen Datenbank aus. Ebenso sind biologische und menschliche Faktoren für den Erfolg – oder das Scheitern – mit ausschlaggebend.

Der Gründer und langjährige Leiter der ZOODAT, Univ.-Prof. Dr. Ernst Rudolf Reichl (1926-1996), hat alle wesentlichen Faktoren in ausreichendem Maße berücksichtigt. Diese Publikation soll daher auch eine Würdigung dieser speziellen Leistungen von E. Reichl sein, die bis heute unübertroffen sind.

Es ist auch bezeichnend, dass diejenigen Datenbankbetreiber, die schon in der Konzeptphase Kontakt mit Ernst Reichl aufnahmen und sich von ihm beraten ließen, ebenfalls sehr rasch ein funktionierendes System aufbauen konnten, während viele von denjenigen, die glaubten, dass sie es ohnehin besser wüssten, trotz großer Investitionen nur vergleichsweise magere Ergebnisse zustande brachten.

Ich verwende häufig den Ausdruck tiergeografische Datenbank, auch wenn viele der hier erörterten Aspekte ebenso auf floristische bzw. auf biogeografische Datenbanken im allgemeinen zutreffen. Ebenso schreibe ich meist ZOODAT, aber auch fallweise ZOBODAT. Die heutige ZOBODAT wurde 1972 als rein tiergeografische Datenbank mit entomologischem Schwerpunkt gegründet. Seit 1999 werden auch botanische Daten mit aufgenommen und aus diesem Anlass wurde sie in ZOBODAT umbenannt. Die meisten der hier angeführten Aspekte sind aber bereits zur ZOODAT-Zeit eingeführt und angewandt worden. Daher ist es korrekter, in diesem Zusammenhang die Bezeichnung ZOODAT zu verwenden. Den Namen ZOBODAT verwende ich, wenn es sich um einen aktuellen Aspekt dieser Datenbank handelt.

Da die entscheidenden Faktoren für den Erfolg einer tiergeografischen Datenbank in vielfältigster Weise ineinandergreifen, werde ich zunächst die verschiedenen Aspekte, die für den Erfolg oder Misserfolg maßgebend sind, erörtern und diese entscheidenden Faktoren dann am Ende extrahieren.

© Biologiezentrum Linz/Austria; download unter www.biologiezentrum.at

2. Zu welchem Zweck betreibt man eine tiergeografische Datenbank?

In einer tiergeografischen Datenbank werden, vereinfacht ausgedrückt, genaue Fundmeldungen von einzelnen Tierarten gespeichert.

Die Hardware und Software, besonders aber die Betreuung derselben und die konkrete Dateneingabe sind nicht billig. Man betreibt diesen finanziellen und organisatorischen Aufwand, damit man schließlich die Daten in geeigneter zusammengefasster Form abrufen, also auswerten kann, z.B. alle Arten eines bestimmten geografischen Gebietes, die Verbreitung einzelner Arten als Punkt- oder Rasterkarte oder in Form komplexer Berechnungen wie Schutzwürdigkeits-Indizes (REICHL 1993) oder Meldungsdichte-Karten. Verschiedene konkrete Möglichkeiten zur Nutzung dieser Auswertungen für die zoologische Grundlagenforschung, für die angewandte Forschung und für den Naturschutz wurden bereits an anderer Stelle (GEISER 1998a) erörtert.

Bereits nach dieser kurzen und unvollständigen Aufzählung wird klar, dass Auswertungen nur dann sinnvolle Ergebnisse liefern, wenn die Datendichte genügend groß ist. Andernfalls enthält eine Rasterkarte nur einige Zufallspunkte, die keine Aussage über die wahre Verbreitung dieser Art ermöglichen. Eine Faunenliste eines prospektiven Naturschutzgebietes enthält nur dann Arten, wenn vorher welche eingegeben wurden. Daraus folgt unmittelbar: **Eine tiergeografische Datenbank ist umso wertvoller, je mehr Fundmeldungen sie enthält.**

Das klingt scheinbar trivial, aber ich habe oft genug erlebt, dass dieser Faktor im Zuge der Begeisterung für das neue Computersystem, das man zu diesem Zweck eingerichtet hat, völlig unterschätzt wurde. Die Betreiber (vor allem die Geldgeber) wollten nach wenigen Monaten bereits aussagekräftige Kartendarstellungen und komplexe Auswertungen sehen!

Dieser Faktor ist so wichtig, dass ich ihn als "Fundamentalsatz der Datenbanken" formulieren möchte: **Das wichtigste und wertvollste einer Datenbank sind die Daten, und nicht die Software und Hardware!**

Daraus folgt als "1. Ableitung": **Ein wesentlicher Erfolgsfaktor für eine tiergeografische Datenbank ist ein Konzept, das eine möglichst rasche und fehlertolerante Eingabe vieler Fundmeldungen ermöglicht.**

3. Der Mensch und seine begrenzten Fähigkeiten – ein häufig unterschätzter Faktor bei tiergeografischen Datenbanken

Datenbanken werden von Menschen gemacht und betrieben. Biogeografische Daten werden wohl noch für lange Zeit von Menschen in die Datenbank eingegeben werden. Daten aus Insektensammlungen, aus Alkoholgläsern und aus Herbarien lassen sich nicht einfach einscannen. Selbst der Transfer von eingescannten Literaturdaten in eine tiergeografische Datenbank erfordert eine so aufwändige Nachbearbeitung und ist so fehleranfällig, dass es bis auf wenige spezielle Ausnahmen auch nicht gemacht wird.

Bei großen und komplexen Softwareprojekten, an denen viele Leute mitwirken, ist eine klare Definition der Schnittstellen der einzelnen Systemteile eine Voraussetzung für das spätere Funktionieren des gesamtem Systems. Beim Konzept und beim Betrieb einer

© Biologiezentrum Linz/Austria; download unter www.biologiezentrum.at

tiergeografischen Datenbank muss man daher auch die speziellen Eigenschaften dieser menschlichen Schnittstellen mit berücksichtigen und zwar:

- beschränkte Zeit
- beschränkte Ressourcen
- beschränkte Geduld
- beschränktes Hirn

Für die Konzeption, die Implementierung und den Betrieb einer tiergeografischen Datenbank sind natürlich sehr spezielle Kenntnisse der Informatik und der Biologie notwendig. Diese speziellen Kenntnisse sind bei einem professionellen Informatiker und einem professionellen Biologen keineswegs automatisch vorhanden. (Die Biologen wissen das natürlich von den Biologen, nehmen aber an, dass jeder Informatiker das nötige "Know how" hat, und vice versa!). Selbst ein Team von an der Fragestellung interessierten Informatikern und computererfahrenen Biologen entwickelt nicht automatisch ein tragfähiges Konzept für eine tiergeografische Datenbank. Nun ist aber gerade das Konzept das wichtigste bei einer Datenbank und einem Programm. Ein guter Entwurf ist hundertmal mehr wert als eine noch so raffinierte Programmierung!

4. Welche Daten sollen und dürfen gespeichert werden?

4.1. Die Kunst der weisen Beschränkung auf das Wesentliche

Einer der häufigsten Fehler in der Konzeptphase tiergeografischer Datenbanken ist der Irrglaube vieler Biologen, dass man mit einer Datenbank "alles" machen kann und soll. Berauscht von den Möglichkeiten, die die digitale Erfassung und Auswertung bieten, wird bei der schlichten Frage, was denn nun genau erfasst werden sollte, gerne übers Ziel hinausgeschossen.

Natürlich muss man die konkreten Artnamen, den konkreten Fundort, das Funddatum und die Angabe der Datenquelle speichern, aber diese Grunddatenfelder allein sind ja viel zu banal. Zu den Artnamen braucht man natürlich sämtliche (!) Synonyme (siehe Kapitel 8). Bei jeder Fundmeldung sollten am besten auch "ökologische Angaben" wie die konkrete Pflanzenart, auf der der Käfer oder die Blattwespe gefunden wurde, die vegetationskundliche Einheit, diverse Biotopparameter, die Windstärke, die Lufttemperatur, die Sonneneinstrahlung und noch zahlreiche andere "ökologische Parameter" mit abgespeichert werden.

Bevor ich dieses Ansinnen als kontraproduktiv und illusorisch brandmarke, ist noch eine wichtige Unterscheidung zu treffen. Wenn jemand eine private Datenbank für seine Spezialtiergruppe anlegt, so kann es durchaus sinnvoll sein, zahlreiche Parameter bei jeder konkreten Fundmeldung mit einzuspeichern. Bei jeder Tiergruppe gibt es andere nützliche Zusatzfelder, die für spezielle Zwecke gebraucht werden. Auch für verschiedene Forschungsprojekte wird man, je nach Fragestellung, manchen der vorhin aufgezählten Parameter benötigen.

Möchte man allerdings eine umfassende tiergeografische Datenbank einrichten, die von verschiedenen Tiergruppen Funddaten aus verschiedenen Jahrzehnten und sehr verschiedenen Quellen enthält, so ist eine weise Beschränkung viel zielführender. Die abzuspei-

© Biologiezentrum Linz/Austria; download unter www.biologiezentrum.at

chernden Datenfelder dürfen in diesem Fall nicht die "Vereinigungsmenge" aller möglichen Parameter sein, sondern die "Durchschnittsmenge" aller notwendigen Informationen. Das sind die vorhin erwähnten Grunddatenfelder. Der Kern der ZOBODAT besteht bis heute aus nur 6 Tabellen, von denen jede eine sehr überschaubare Anzahl von Attributen hat (genauer Aufbau der ZOBODAT und weitere Details siehe MALICKY 2001).

Zu viele Felder widersprechen nämlich dem Fundamentalsatz der Datenbanken und seiner ersten Ableitung. Wenn man pro Datensatz 20 Parameter eingeben muss anstelle von 5 oder 8, so dauert die Dateneingabe entsprechend länger und es gelangen dadurch pro Zeit (also pro Geld) weniger Fundmeldungen in die Datenbank.

4.2. Die Auswahl der wichtigen und richtigen Daten

Angenommen, man hat nun eine tiergeografische Datenbank mit einem sinnvollen Datenbankentwurf und Tabellendesign. Bevor man die Daten in eine Datenbank einspeichern kann, muss man diese Daten erst akquirieren und einer ersten Qualitätskontrolle unterziehen.

Als Datenquellen für tiergeografische Datenbanken kommen öffentliche und private Sammlungen, Fachliteratur, Naturschutz-Gutachten, Biotopkartierungen, Exkursionsprotokolle, Karteien, Diplomarbeiten und Dissertationen und inzwischen auch private Dateien in Frage. Während die Kenntnis über öffentliche Sammlungen und die Standardliteratur noch leicht zu eruieren ist, erfährt man von Privatsammlungen und von Grauliteratur nicht so ohne weiteres. Dieses Wissen ist in den Köpfen örtlicher Spezialisten gespeichert. Diese Spezialisten wissen auch gut über die Qualität dieser Datenquellen Bescheid.

Einer der wesentlichsten Faktoren für den Erfolg einer tiergeografischen Datenbank ist daher eine gute und langjährige Zusammenarbeit mit entsprechend qualifizierten Spezialisten. Gerade Privatsammlungen und Grauliteratur können oft besonders wertvolle Datenquellen sein, andererseits ist gerade in diesem Bereich eine Qualitätsabschätzung vor der Einspeicherung enorm wichtig. Letzteres empfiehlt sich auch bei öffentlichen Sammlungen, hier kann aber der Sammlungsbetreuer im allgemeinen kompetent Auskunft geben.

Nach der Abschätzung der Datenqualität kann man entscheiden, ob man die vorliegende Datenquelle in die Datenbank einspeichern will. Es ist völlig unrealistisch zu fordern, dass ausschließlich revidierte Sammlungen aufgenommen werden dürfen. Jede Sammlung, auch eine von Spezialisten revidierte, enthält Fehler. Diese Fehler werden später bei der Auswertung erkannt. Andererseits sollten natürlich nicht allzu viele fehlerhafte Daten aufgenommen werden, denn dadurch wird die Auswertung zweifelhaft bis unbrauchbar (siehe dazu auch GEISER 1996).

Es gibt keine allgemeine Anleitung, nach welchen Kriterien oder nach welcher "Metrik" man die Qualität einer Datenquelle einstufen kann, denn das hängt ganz wesentlich von der Tiergruppe ab. Aber generell gilt, dass renommierte Spezialisten zuverlässige Beurteilungen abgeben. Mit diesen sollte man daher intensiv zusammenarbeiten und vor allem sollte man sie auf keinen Fall vergrämen. Meist arbeiten sie ohnehin gerne freiwillig und mit viel Idealismus mit und freuen sich, wenn ihr Urteil gefragt ist. Wenn man sie dann aber ausnützt und/oder mit schikanösen Sammelbeschränkungen verärgert oder ihr Urteil ignoriert und sich der Meinung von jemandem eindeutig minderqualifizierten

© Biologiezentrum Linz/Austria; download unter www.biologiezentrum.at

anschließt, verliert man eine wertvolle Informationsquelle. Nur die Datenbanken, die in beiderseitig erfreulicher Weise mit den Spezialisten zusammenarbeiten, erfahren überhaupt von genügend vielen guten Daten, die einzuspeichern auch wert sind!

5. Relationale Datenbanken, die adäquate Form der Datenspeicherung für tiergeografische Fragestellungen und ihre historische Entwicklung

Wenn man eine tiergeografische Datenbank plant, so wird man zu diesem Zweck eine relationale Datenbank verwenden. Das gilt heute als selbstverständlich. Als Ernst Reichl die ZOODAT 1972 entwarf und implementierte, entwarf er die Tabellenstruktur genau nach den Prinzipien einer relationalen Datenbank, aber zu einer Zeit, in der das keineswegs selbstverständlich war. Um die herausragende Leistung von Ernst Reichl beim Entwurf der Datenbank für ZOODAT würdigen zu können, betrachte man folgende Zusammenstellung zur Informatikgeschichte (rDBMS bedeutet: relationales Datenbank-Management-System).

vor 1970 wurden Daten in Listen gespeichert. Das ist sehr unpraktisch und führte zu verschiedenen Versuchen, ein besseres Konzept zu finden (hierarchisches Modell, Netzwerk-Modell, u.a.).

1970 erschien die bahnbrechende Publikation von E. Codd: "A Relational Model of Data for Large Storage Databanks"

Hierin beschreibt Codd die mathematischen Grundlagen dessen, was wir heute unter einem rDBMS verstehen. Diese beruhen auf der Mengenlehre[1], Relationen-Algebra[2] und Prädikatenlogik[3].

Nachdem nun die wesentlichen theoretischen Grundlagen veröffentlicht waren, begannen verschiedene Versuche, dieses rDBMS in die Praxis umzusetzen.

1974-1975 gelang IBM die Herstellung eines Prototypen eines rDBMS.

Nun konnte man die Daten in sinnvoller Struktur speichern. Aber gespeicherte Daten möchte man ja auswerten.

Das führte in den Jahren **1974-1977** zur Entwicklung der Datenbankabfragesprache SQL. Diese ist von allen frühen Versuchen und Evolutionslinien nicht nur bis heute in Gebrauch, sondern die Grundlage aller Auswertungsprogramme von rDBMS geblieben.

1976 erschien wieder eine wesentliche Publikation: P. CHEN: "Das Entity-Relationship-Modell". Sie enthält die Theorie und daraus abgeleitet die praktische Anleitung, wie man bei der Konzeption der Tabellen eines rDBMS vorgehen muss.

---

[1] Die Mengenlehre, die hier zur Anwendung kommt, ist eine anspruchsvolle und grundlegende mathematische Theorie. Nicht zu verwechseln mit der Schul-Mengenlehre, mit der jahrezehntelang Schüler im Unterricht sinnlos verwirrt wurden und werden!

[2] Daher kommt der Name "relationales" Datenbanksystem und nicht daher, weil die Tabellen zueinander in einer "Relation" stehen.

[3] Das ist – grob gesagt – eine Erweiterung der bekannten Wahr-Falsch-Tafeln, um weitere Regeln, mit denen man "ist größer", "ist kleiner" , "ist gleich", "es gibt ein", "für alle gilt" und ähnliches ebenso formal berechnen kann.

© Biologiezentrum Linz/Austria; download unter www.biologiezentrum.at

1979 bringt die Firma Oracle ein rDBMS mit SQL als kommerzielles Produkt heraus. Nun werden rDBMS auch für Firmen und Behörden nutzbar! Die ganze Entwicklung vor 1979 hat sich nur in Universitäten und im Forschungsbereich von Informatikfirmen abgespielt.

In den **1980iger Jahren** setzen sich rDBMS als Standard durch.

1985 kommt dBASE auf den Markt. Damit werden rDBMS auch für den Privatanwender nutzbar.

Wenn man die ZOODAT in dieses historische Schema einordnen möchte, muss man ziemlich weit zurückgehen. Denn Ernst Reichl hat die ZOODAT bereits 1972 entworfen und implementiert und einen funktionsfähigen Prototypen Ende Mai 1972 in Saarbrücken bei der Gründungstagung des European Invertebrate Survey vorgestellt. Diese Ur-ZOODAT war bereits ein astreines, vorbildliches rDBMS, dessen Tabellen genau nach dem Entity-Relationship-Modell aufgebaut sind! Ernst Reichl war auch hier Pionier und hat bereits 1972 jenes System für seine ZOODAT ausgewählt, das sich später als das optimale Datenbanksystem zum allgemeinen Standard etabliert hat.

Dabei muss man noch bedenken, dass 1972 weder so viel Speicherplatz noch bereits so ausgereifte Werkzeuge zur Datenbankentwicklung wie heute zur Verfügung standen. Gerade der begrenzte Speicherplatz war eine zusätzliche Beschränkung, die man schon ab dem Entwurf ständig berücksichtigen musste.

In Diskussionen über tiergeografische Datenbanken konnte man immer wieder hören, dass die ZOODAT ja deshalb so erfolgreich sei, weil Ernst Reichl sowohl Entomologe als auch Informatiker war. Das stimmt aber nur zum Teil. Ernst Reichl hatte ein solch gutes Konzept einer tiergeografischen Datenbank deshalb entworfen und implementiert, weil er ein guter Entomologe und ein hervorragender Informatiker war!

Die Entomologen kennen seine Informatik-Aktivitäten nur im Zusammenhang mit ZOODAT. Ernst Reichl hat sich aber auch mit vielen anderen Informatik-Themen befasst. Er war auch ein Pionier auf dem Gebiet der "neuronennetz-analogen Assoziativspeicher"[4] und der "Lernenden Systeme". Ein weiteres Forschungsgebiet von ihm war "Methoden der automatischen Zeichen- und Bilderkennung". Ich erwähne diese Forschungsbereiche von Ernst Reichl hier nicht nur um zu illustrieren, was für ein vielseitiger Informatiker er war, sondern auch deshalb, weil seine Kenntnisse auf diesen Gebieten in die Konzeption der ZOODAT unmittelbar eingeflossen sind. Ein ebenfalls hervorragender, aber mehr technisch orientierter Informatiker hätte sicher ein weniger optimales Konzept entworfen.

Es gibt aber noch einen weiteren Grund, warum ich diese Leistungen von Ernst Reichl hier betone: Personen mit einer Doppelbegabung werden immer nur zur Hälfte gewürdigt. Ernst Reichl war unter Informatikern hoch geschätzt und anerkannt, aber nicht wegen der ZOODAT. Die galt für seine Informatik-Fachkollegen als seltsames Hobby, wie das ältere Professoren manchmal so betreiben, und auch die meisten seiner Studenten interessierte die ZOODAT nicht. Von Informatikerseite wurde er für die ZOODAT nicht gewürdigt, weil ein Informatiker (der sich nicht zufällig mit solchen Fragestellungen intensiv befasst) weder die Bedeutung einer tiergeografischen Daten-

[4] Dieses Fachgebiet ist inzwischen unter dem griffigeren Namen "Neuronale Netze" bekannt geworden.

© Biologiezentrum Linz/Austria; download unter www.biologiezentrum.at

bank ermessen, noch den Überblick und Vergleich mit den anderen derartigen Datenbanken Europas haben kann! Und seinen Entomologie-Kollegen bleiben seine Leistungen auf dem Gebiet der Informatik ebenso naturgemäß verschlossen.

Dass die Kenntnisse von Ernst Reichl auf dem Gebiet der Neuronalen Netze und Lernenden Systeme beim Konzept der ZOODAT Eingang gefunden habe, möchte ich nun an Hand eines Beispiels erläutern, das jedem Entomologen und ganz besonders jedem Entomofaunisten besonders gut vertraut ist, und das ist der Fundort!

6. Wie speichert man Fundortangaben am besten in einer tiergeografischen Datenbank?

Die Einspeicherung des Fundortes ist ein essentieller Bestandteil jeder tiergeografische Datenbank. Ein Fundort muss bei jedem Datensatz angegeben werden. Daher ist er eine kritische Größe bei jeder Dateneingabe. Andererseits wird er auch bei jeder Auswertung benötigt, von der einfachen Faunenliste eines Ortes über verschiedene Kartendarstellungen bis hin zu komplexen Berechnungen. Weil eben der Fundort so wichtig ist, gibt es unter den Betreibern tiergeografischer Datenbanken heftige Debatten, in welcher Form er am besten zu speichern sei.

Die ersten biogeografischen Datenbanken wurden häufig von Botanikern betrieben, aber auch die Ornithologen verfügten bereits in den 60iger Jahren über große Datenmengen auf den Karteikarten der zahlreichen lokalen und überregionalen Arbeitgemeinschaften und begannen Datenbanken einzurichten. Bei Gefäßpflanzen und bei Vogelfunddaten ist es bis heute üblich, den Fundort als Raster in einem zuvor definierten Kartennetz anzugeben. Immerhin ist ja selbst ein standortstreuer Vogel noch wesentlich mobiler als die meisten Evertebraten. In der Vegetationskunde erfolgt die Kartierung meist mit Strichlisten pro Quadrant, z.B. in Österreich auf Bundesländerebene in einem 3 × 5-Minuten Raster und auf Staatenebene in 6 × 10 Minuten-Rastern bezogen auf die geografische Länge und Breite.

Das hat dazu geführt, dass diese Methode der Rasterkartierung anfangs auch dann noch angewendet wurde, als man auch Funddaten anderer Tiergruppen einzuspeichern begann. Das macht zwar heute niemand mehr, aber in den 70iger Jahren was es eine übliche Methode. Nun werden aber Evertebraten in Sammlungen und bei Faunenlisten praktisch nie mit einen Raster verortet, sondern mit einer konkreten Fundortangabe. Diese kann man zwar einem Raster zuordnen, aber dabei geht dann die genaue Fundortinformation verloren und auch die beste Datenbank der Welt kann beim Einspeichern verlorene Information nie wieder vervollständigen.

Warum man bei tiergeografischen Datenbanken nicht gerade Vogeldaten als Maß aller Dinge hernehmen sollte, zeigt ein Blick auf die Verteilung der Tierarten in Österreich. Die Prozentzahlen entsprechen der Verteilung der nicht-marinen Tierarten in Europa.
Aus Abbildung 1 geht hervor, dass 81 % aller Tierarten zu den Insekten gehören. Die Schmetterlinge und Käfer machen allein ein Viertel aller Arten aus. Von diesen beiden Tiergruppen sind in Form von Sammlungen und Literaturangaben besonders viele tiergeografische Daten vorhanden! Alle Evertebraten zusammen stellen 99 % aller Tierarten (auch weltweit!). Es ist daher sinnvoll, sich bei der Fundortspeicherung an den

© Biologiezentrum Linz/Austria; download unter www.biologiezentrum.at

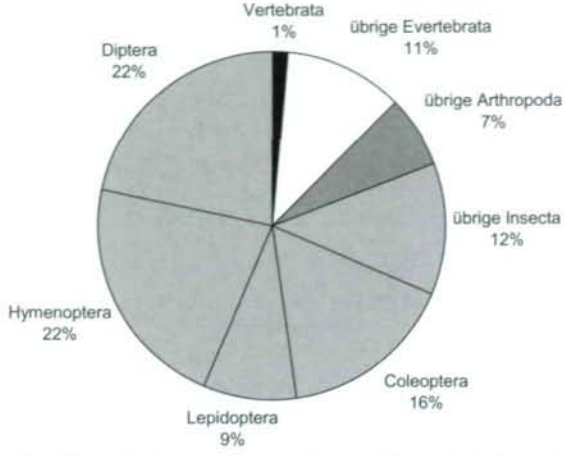

Abb. 1: Prozentuale Verteilung der Artenzahlen der Tierwelt Österreichs (nach GEISER 1998b).

Insektendaten zu orientieren, deren Kriterien auch für die Fundmeldungen der anderen Evertebraten gelten.

Eine punktgenaue Abspeicherung des Fundortes ermöglicht anschließend eine Kartendarstellung in einem Raster, aber aus einer Rasterspeicherung kann man nie wieder den ursprünglichen Fundort rekonstruieren!

6.1. Die alte und überflüssige Debatte UTM-Gitter versus geografische Koordinaten

Noch heftiger waren – und sind zum Teil – aber die Debatten um die "richtige" Rasterdarstellung. Während die Rasterspeicherung bei Insektenfunddaten heute kein Thema mehr ist, erlebt man immer wieder heftige Befürworter der Darstellung im UTM-Gitter oder in einem der nationalen Netze. Für die Darstellung von Verbreitungskarten, die mehrere europäische Länder umfassen, wird das UTM-Gitter forciert. Ernst Reichl hatte jahrezehntelang die Mehrheit der Betreiber tiergeografischer Datenbanken in Europa gegen sich, weil er seine Fundorte in einem geografischen Gradnetz speicherte. Erst seit dem Fall des Eisernen Vorhanges ist die Debatte deutlich abgeebbt (aber keineswegs verstummt), als sich herausstellte, dass es für Länder des ehemaligen Ostblockes äußerst schwierig ist, Karten mit UTM-Gitter aufzutreiben. Das geografische Gradnetz mit dem Nullmeridian durch Greenwich ist schon seit Jahrhunderten ein weltweit verwendeter Standard und auf (fast) jeder Karte abgedruckt.[5]

Die Debatte ist vor allem deshalb überflüssig, weil man punktgenau abgespeicherte Funddaten anschließend in jede gewünschte Rasterdarstellung umrechnen kann, also für eine Österreichkarte in ein 6 × 10 min geografisches Raster mit praktischen Planquadraten oder bei Karten, die ein Gebiet der Größe Mitteleuropas umfassen, eben in ein UTM-Gitter, wenn man das möchte.

Ernst Reichl entschied sich von vorn herein für eine punktgenaue Fundortabspeicherung.

---

[5] Vorsicht ist nur bei älteren Karten aus Frankreich geboten, da hier der Nullmeridian durch Paris gehen kann!

© Biologiezentrum Linz/Austria; download unter www.biologiezentrum.at

Nun können Fundortangeben aber sehr verschieden große Gebiete bezeichnen. So ist der Fundort Samer Mösl ein sehr kleines Gebiet (ein Moorrest im Stadtgebiet von Salzburg im Stadtteil Gnigl), die Fundortangabe Leopoldskron umfasst eine wesentlich größere Fläche und Fundortangaben wie "Salzburg Stadt" ohne genauere Einschränkungen gibt es ja auch zur Genüge. Ebenso unterscheiden sich die Fundorte Hackelsberg (bei Jois im Burgenland) und Lainzer Tiergarten (im 13. Bezirk in Wien) beträchtlich in ihrer Größe.

Manche tiergeografische Datenbanken verlangen bei der Fundorteingabe Flächenschätzungen, Kreisradien und Unschärfebereiche. Solche Angaben sind zwar bei aktuellen Sammelerhebungen feststellbar, bei Sammlungsbelegen und Literaturangaben aber nur selten und wenn überhaupt, dann nur sehr aufwendig eruierbar.

Von den vielen Möglichkeiten der Fundortabspeicherung wählte Ernst Reichl eine verblüffend einfache, aber äußerst zielführende und praktikable Methode:

Als Fundort wird in ZOODAT genau das abgespeichert, was auf dem Etikett steht!

Die überwiegende Mehrheit der Daten in ZOODAT stammt aus öffentlichen und privaten Sammlungen. Wenn der Fundort genau vom Etikett übernommen wird, erleidet man keinen Informationsverlust. Daher wird als Fundortangabe eben Samer Mösl, Leopoldskron oder Salzburg Stadt direkt abgespeichert.

Für die Einspeicherung in den Computer und für die Auswertung und Kartendarstellung ist es notwendig, jeden Fundort mit einer Nummer eindeutig zu kennzeichnen. Jedem Fundort wird also eine Bezeichnung zugeordnet (das, was auf den Etikett steht), bei Bedarf wird diese Bezeichnung noch ergänzt (manchmal steht nur St. Johann dort, weil der Sammler Förster in St. Johann im Pongau war, aber es gibt ja auch ein St. Johann in Tirol) und dann wird noch jeder Bezeichnung eine eindeutige Nummer zugeordnet. Jede so mit einer Nummer versehene Lokalität wird dann mit der Angabe von geografischer Länge und Breite auf 1/100 Grad genau ergänzt und mit der Nummer des Planquadrates[6] im 6x10 Minuten-Raster und der Höhenangabe bzw. den Höhenbereich (min und max Höhenmeter) als eigener Datensatz in der Fundorttabelle abgespeichert.

Die eindeutige Nummer (der Primärschlüssel) des Fundortes ermöglich dann die Verknüpfung mit anderen Tabellen der Datenbank. In der Fundmeldungstabelle müssen dann nur die Nummern abgespeichert werden und alle andern Zusatzinformation (wie der Ort heißt, wo er genau liegt usw.) holt man sich bei Bedarf für die Auswertungen aus der Fundorttabelle. Diese Tabellenstruktur ermöglicht es auch, auf sehr einfache Weise die Fundortinformation um zusätzliche Felder zu erweitern. So könnten in Zukunft bei jedem Fundort die Jahresmitteltemperatur, die durchschnittliche Niederschlagsmenge und weitere örtliche Parameter mit abgespeichert und dann entsprechend ausgewertet werden, wenn Bedarf dafür besteht.

[6] Das ist zwar eine redundante Information, weil man das Planquadrat aus der geografischen Länge und Breite berechnen kann und redundante Speicherung sollte man in Datenbanktabellen tunlichst vermeiden (Normalisierung). In besonderen Fällen sind aber manche Redundanzen gerechtfertigt und das ist hier der Fall, denn die Speicherung der Planquadratnummern erleichtert die Auswertung enorm!

© Biologiezentrum Linz/Austria; download unter www.biologiezentrum.at

6.2. Die Erstellung von Fundortnummern als Schlüsselfelder in einer Datenbank: Vergleich eines üblichen Informatik-Konzeptes mit der optimalen Lösung in ZOODAT

Diese eindeutige Fundortnummer ist also sehr wichtig und man kann sie in informationstheoretisch-mathematisch-kryptographisch einwandfreier Weise folgendermaßen zusammenstellen:

Man weist jeder Fundortbezeichnung eine 10-stellige Zahl zu, in der die Ortinformation verschlüsselt ist:

a_i (i = 1, ... 10) seien jeweils als ganze Zahlen definiert

Dann verwende man diese 10 Ziffern in folgender Weise:

a_1 a_2	a_3	a_4 a_5	a_6 a_7 a_8 a_9	a_{10}
Staat	Bundesland[7]	Bezirk	Fundort innerhalb Bezirk	Prüfziffer

also z.B.

11 7 02 1154 a_{10} für Zeilerberg

wenn 11: Österreich
7: Burgenland
02: Bezirk Neusiedl am See
1154: Zeilerberg am Südhang des Leithagebirges in der Gemeinde Jois

bedeuten würde.

Die Prüfziffer wird nun so berechnet, dass man einen fehlerkorrigierenden Code erhält:

a_{10} ist das Komplement der gewichteten Quersumme von a_1 bis a_9 modulo 11

Was hier in der zahlentheoretisch korrekten Bezeichnung so abschreckend kompliziert klingt, ist mit dem mathematischen Kenntnisstand eines Volksschul-Viertklasslers durchschaubar!

Gewichtete Quersumme bedeutet:

Jede Ziffer in dieser Fundortnummer wird mit einer unterschiedlichen Zahl multipliziert und anschließend werden die Ergebnisse zusammengezählt, z.B.:

$$10 \times a_1 + 9 \times a_2 + 8 \times a_3 + 7 \times a_4 + 6 \times a_5 + 5 \times a_6 + 4 \times a_7 + 3 \times a_8 + 2 \times a_9$$

Nun berechnet man die "gewichtete Quersumme" unserer Schlüsselzahl vom Zeilerberg:

$$10 \times 1 + 9 \times 1 + 8 \times 7 + 7 \times 0 + 6 \times 2 + 5 \times 1 + 4 \times 1 + 3 \times 5 + 2 \times 4 = 119$$

Modulo 11 bedeutet, dass man diese Zahl durch 11 dividiert und sich den Rest r merkt also

$$119 = 11 \times 10 + 9$$
$$r = 9$$

Das Komplement modulo 11 ist die Zahl, um die man den Rest ergänzen muss, damit man wieder 11 erhält, also hier: 9 + **2** = 11

a_{10} ist daher 2

---

[7] Hat man einen Staat mit mehr als 10 Bundesländern, so kann man kleinere Bundesländer mit benachbarten zusammenfassen, z. B. Rheinland-Pfalz mit dem Saarland oder Bremen mit Niedersachsen. Sehr große Staaten kann man in Gebiete mit mehreren Staatennummern aufteilen. In der Informatik sind solche Zuordnungen alle politisch korrekt!

© Biologiezentrum Linz/Austria; download unter www.biologiezentrum.at

Wenn man nun diese Zahl a_{10} zur gewichteten Quersumme dazuzählt, dann erhält man eine Zahl, die durch 11 teilbar ist.

Also in unserem Fall: $119 + 2 = 121 (= 11 \times 11)$

Auf diese Weise kann man den Computer so programmieren, dass er vor dem Speichern berechnet, ob die gewichtete Quersumme der eingetippten Zahl durch 11 teilbar ist. Wenn nicht, muss ein Tippfehler vorliegen und der Computer verweigert die Speicherung. Sobald man sich nämlich an einer Stelle vertippt, ist die gewichtete Quersumme nicht mehr durch 11 teilbar! Dasselbe passiert, wenn man 2 Zahlen vertauscht, was beim Eintippen ja häufig passiert. Das funktioniert deshalb so gut, weil 11 eine Primzahl ist (weitere Details für nun Interessierte in BEUTELSPACHER & ZSCHIEGNER 2002).

Wenn man sich die Rechnung mit dem Rest genauer überlegt, so bemerkt man, dass manchmal der Rest $r = 1$ bleiben kann und dann muss die Prüfziffer $a_{10} = 10$ sein. Wir haben aber nur eine Stelle dafür reserviert! In einem solchen Fall behilft man sich in der Informatik mit einem einfachen Trick: Man schreibt den Wert "10" einfach als römische Ziffer "X" und schon kommt man mit einer Stelle aus.[8] Man muss nur dem Computer vor der Berechnung mitteilen, das er mit "X" wie mit "10" rechnen soll. Da ein Computer ohnehin weder mit arabischen noch mit römischen Ziffern rechnet, sondern mit Binärzahlen (also nur mit Nullen und Einsen), ist ihm das egal.

Die Theorie der fehlerkorrigierenden Codes ist zwar faszinierend, aber dieses Verfahren ist nicht immer optimal. Bei der Datenaufnahme von biologischen Funddaten muss ja jemand – eine bezahlte Kraft oder ein wohlgesonnener Privatsammler – diese Fundortnummern eingeben. So viele Fehler kann ein fehlerkorrigierender Code gar nicht korrigieren, wie der Datenlieferant bei der Eingabe von zehnstelligen Zahlen macht! Hier kommen massiv die menschlichen Faktoren ins Spiel, die man bei einer tiergeografischen Datenbank berücksichtigen muss und zwar gleich alle vier: beschränkte Zeit, beschränkte Ressourcen, beschränkte Geduld und beschränktes Hirn.

Ernst Reichl, der zwar die Kryptographie in Theorie und Praxis beherrschte, aber auch ein Experte für Lernende Systeme war, verwendete hier ein ganz anderes, verblüffend einfaches, aber äußerst effektives Verfahren:

Die Fundortnummern in ZOODAT sind vierstellig, denn das kann man sich merken!

Ab 5 Ziffern sinkt bereits die Merkfähigkeit rapide ab, hier hat unser Hirn eine deutliche Schranke! Das ist auch der Grund, warum der Bankomat-Code und der PIN-Code beim Handy vierstellig sind. Aus sicherheitstechnischen Gründen wäre eine mehr als vierstellige Nummer viel geeigneter, denn einen vierstelligen Code kann man leichter knacken als einen mehrstelligen. Aber Nummern mit mehr als 4 Stellen merkt man sich nur sehr schwer, dann schreiben die Leute die Nummern auf Zettel und die Sicherheit ist erst recht dahin!

Mit 4 Ziffern kann man allerdings höchstens 10^4 Fundorte kennzeichnen. Die österreichischen Postleitzahlen, die ja auch "Fundorte" verschlüsseln, kommen zwar mit 4 Stellen

[8] Mit dem hier beschriebenen Verfahren sind übrigens die ISBN Nummern verschlüsselt. Viele seit 1970 erschienene Buch ist durch diese Nummer eindeutig gekennzeichnet. Bei einer Bestellung werden so die allermeisten Tippfehler gleich erkannt. Auch dort ist die letzte Ziffer die Prüfziffer und an dieser Stelle steht manchmal ein X, wie Sie an einer genügend großen Stichprobe Ihrer Privatbibliothek jederzeit nachprüfen können!

© Biologiezentrum Linz/Austria; download unter www.biologiezentrum.at

aus, diese Zuordnung ist aber lange nicht so genau wie die Fundorte, die in ZOODAT gespeichert werden. Hier werden wesentlich mehr als 10 000 Fundorte (innerhalb Österreichs und für einige angrenzende Gebiete, wie Südtirol oder die bayerischen Alpen) verwaltet. ZOODAT wendet dabei denselben Trick an, den ich schon bei der ISBN-Nummer erläutert habe: man kann bei Datenbanken statt Ziffern auch andere Zeichen, also z.B. Buchstaben, verwenden. Dann explodieren die zur Verfügung stehenden Schlüsselzahlen von 10^4 auf $(10 + 26)^4$, also von 10000 auf 1,65 Millionen. Damit kommt man für ein Gebiet wie Österreich auch bei punktgenauen Fundortangaben gut aus. Bisher werden in ZOBODAT Buchstaben nur an der Tausenderstelle verwendet, die gleichzeitig das Bundesland kennzeichnet.

Aus denselben Überlegungen sind in ZOODAT nicht nur die Fundortnummern, sondern auch die Artnummern vierstellig, allerdings sind sie nur innerhalb einer Tiergruppe (z.B. der Käfer oder der Schmetterlinge) eindeutig[9].

7. Die Optimierung der Ergonomie der Dateneingabe: ein wesentlicher Kostenfaktor für den Betrieb einer Datenbank

ZOODAT wurde lange vor der Erfindung und flächendeckenden Verbreitung des PCs konzipiert und die Hauptmenge der Dateneingaben (bis heute!) erfolgt mit Aufnahmeformularen, in das für jede Art und jeden Fundort eine vierstellige Nummer eingetragen werden

Wenn man aus einer Insektensammlung Daten aufnimmt, dann gibt es in jeder Sammlung zahlreiche Fundorte, die sich wiederholen. Man schreibt sich anfangs die Nummern heraus und die häufigen merkt man sich rasch, sodass man die Eintragung relativ schnell machen kann. Auch mit einem guten Computereingabeprogramm (mit Klartext, also jedesmal eintippen: St. Jakob am Thurn usw.) kann man nicht so schnell Daten eingeben wie mit diesem Nummernverfahren!!

ZOODAT wurde mit ihrer optimalen Anpassung an die Einspeicherung von Insektendaten und dem vierstelligen Nummernsystem, mit minimalen Personal und Budget und einer geringen Anzahl freiwilliger Mitarbeiter in kurzer Zeit zu einer der größten tiergeografischen Datenbanken Europas[10]. Ich kenne viele tiergeografische Datenbanken, aber keine hat mit so geringem Budget einen so großen Output zustande gebracht.

Wenn der Eingabemodus nicht optimiert ist, dann braucht man ganz leicht fünfmal so lang oder sogar zehnmal so lang, um einen einzigen Datensatz einzugeben. Das bedeutet, in der selben Zeit, mit demselben Aufwand (ob Geld oder Motivation von Datenlieferanten, denn Sammler geben ihre eigenen Daten meist umsonst ein), bekommt man nur ein Zehntel der Datenmenge! ZOBODAT hätte statt 2,5 Millionen nur 250.000 Datensätze nach über 30 Jahren bei gleichem finanziellen, personellem und idealistischen Aufwand eingespeichert, und die Auswertungen und Karten wären wesentlich weniger brauchbar.

---

[9] Auch bei ZOODAT gibt es natürlich eine Kontrolle auf Eingabefehler, aber nicht über die Nummer mit Prüfziffer, sondern über Klartextverifikation durch Spezialisten. Dieses "Verfahren" erkennt im übrigen wesentlich mehr Fehler als ein noch so raffinierter Algorithmus!

[10] Bis vor wenigen Jahren hatte nur die Datenbank am Institute for Terrestrial Ecology in Monkswood, Huntingdon (United Kingdom) mehr faunistische Datensätze gespeichert als ZOODAT. Vor kurzem wurden beide Datenbanken vom nationalen Erfassungssystem der finnischen Tierwelt überholt.

© Biologiezentrum Linz/Austria; download unter www.biologiezentrum.at

Ich betone die Ergonomie der Dateneingabe deshalb so sehr, weil dieser Bereich – auch und gerade von Informatikern – oft sehr vernachlässigt wird. Dabei hat dieser Faktor den größten finanziellen Einfluss! In der Privatwirtschaft wäre der Auftraggeber gerne bereit, etwas mehr für die Softwarelösung auszugeben, wenn sich dadurch seine Investition schneller amortisiert. Aber oft denken weder Auftraggeber noch der Projektleiter noch die Software-Entwickler daran. Sie alle geben ja normalerweise keine Daten in größeren Mengen ein. Der Programmieraufwand für eine Eingabemaske, die maximale Eingabegeschwindigkeit erlaubt, ist oft nur genauso hoch oder nur geringfügig aufwendiger, man muss nur rechtzeitig daran denken.

An diesem wesentlichen Faktor Eingabeergonomie hoffe ich gezeigt zu haben, dass bei tiergeografschen Datenbanken der Teufel im Detail steckt. Auch bei einem sehr guten Gesamtkonzept und einem optimalen Datenbankdesign kann eine zu geringe Aufmerksamkeit beim Dateneingabemodus dessen Effektivität um den Faktor 10 herunterdrücken und damit die Kosten vervielfachen!

8. Die Fallstricke der zoologischen Systematik

Ein Faktor, der sogar oft von Biologen im Bezug auf Datenbanken unterschätzt wird und von dem der Informatiker zunächst nichts ahnt, sind die Fallstricke der biologischen Systematik. Dieser (Stör)Faktor erschwert den Betrieb einer tiergeografschen Datenbank ganz außerordentlich. Die beteiligten Informatiker merken meist erst relativ spät, welche Abgründe sich da auftun, die mit keinem vorhandenen informationstheoretischen Konzept in den Griff zu kriegen sind!

Neben dem Fundort ist die genaue Angabe der Tierart ein essentieller Feldwert innerhalb eines tiergeografischen Datensatzes. Während man den Fundort mit geografischen Koordinaten eindeutig festlegen kann, ist das bei den Artnamen deutlich komplizierter. Bei vielen Tier- und Pflanzenarten stellt sich nämlich die berechtigte Frage, welchen Namen man abspeichern soll.[11]

Seit LINNAEUS 1758 die binäre Nomenklatur in die Zoologie eingeführt hat, werden Tierarten mit jeweils zwei lateinischen bzw. latinisierten Namen gekennzeichnet. Gültig ist dabei der Name der Erstbeschreibung. Soweit der Idealfall. Viele dieser Erstbeschreibungen waren (und sind?) nicht eindeutig, formal ungültig oder einfach jahrelang verschollen gewesen, sodass manche Arten zwei- bis fünfmal und öfter beschrieben wurden. Da manche dieser Erstbeschreibungen erst in den letzten Jahrzehnten (wieder)entdeckt wurden oder für mehrdeutige Beschreibungen erst in jüngerer Zeit die genaue Typusfestlegung erfolgte, haben sich viele Tier- und Pflanzennamen (aus eigentlich formaljuridischen Gründen!) in den letzten Jahrzehnten geändert, zum Teil sogar mehrfach. Das betrifft vor allem mitteleuropäische Arten, da diese schon sehr früh beschrieben wurden. Für tiergeografische Datenbanken, die Sammlungs- und Literaturdaten aus einem Zeitraum von mehr als 100 Jahren verarbeiten, sind diese Mehrfachnamen (Synonyme) ein gravierendes Problem.

---

[11] Ein Mitarbeiter einer Naturschutzbehörde antwortete mir in einer entsprechenden Diskussion: " Na, den wissenschaftlichen Namen natürlich!"

© Biologiezentrum Linz/Austria; download unter www.biologiezentrum.at

So findet man für einen häufigen violetten Blattkäfer der derzeit(!) *Chrysolina sturmii* WESTHOFF heißt, in der Literatur der letzten Jahrzehnte auch die Namen *Chrysomela violacea* MÜLLER, *Chrysomela goettingensis* L., *Chrysomela diversipes* BEDEL, *Chrysomela fuscipes* GMELIN und noch weitere Variationen mit diesen Gattungsnamen und mit anderen Gattungsnamen. PANZER beschrieb 1797 eine andere häufige violette Art derselben Gattung als *Chrysomela violacea*, die aber schon 1791 von SCRIBA als *Chrysomela coerulans* beschrieben wurde und heute *Chrysolina coerulans* (SCRIBA) heißt.

Zusätzlich erschwert wird die Sache durch die häufige Änderung von Gattungsnamen. Auch bei korrekten und kontinuierlich verwendeten Erstbeschreibungen haben viele Arten Gattungsnamen erhalten, die heute nicht mehr gültig sind. Gattungen sollten nah verwandte Arten – im Idealfall eine monophyletische Artgruppe – zusammenfassen. Im Realfall unterliegt aber die Gattungseinteilung der speziellen Kenntnis (oder Willkür?) des jeweiligen Spezialisten. Daher werden Gattungen immer wieder umgeordnet. Gültig ist diese Einteilung solange, bis sich wieder ein Spezialist dieser Gruppe annimmt. Auch wenn diese Rochaden nur einen Teil der Arten betreffen, bei diesem Prozess ist kein Ende absehbar.

Solange es sich nur um rein formale Namensänderungen handelt, die genau dieselben Arten betreffen, kann man das Problem mit Methoden der Informatik noch in den Griff bekommen. Man speichert zu der eindeutigen Artnummer zusätzlich verschiedene Synonyme ab. Im Detail kann das allerdings auch sehr heikel sein. So ist größte Vorsicht geboten, wenn altbekannte Gattungsnamen auf einmal für Arten anderer Gattungen verwendet werden wie z.B. bei den folgenden Bockkäfern.

"offizielle" Namensänderungen von 1987 bis 1998:

Strangalia maculata => *Leptura maculata*
Strangalia arcuata => *Leptura arcuata*
Strangalia melanura => *Stenurella melanura*
Leptura sexguttata => *Anoplodera sexguttata*
Leptura livida => *Pseudovadonia livida*
Leptura rubra => *Corymbia rubra*
Leptura sanguinolenta =>*Anastrangalia sanguinolenta*

usw.

Auch hier ist es unabdingbar, mit kompetenten Spezialisten der jeweiligen Gruppe zusammenzuarbeiten um die Artschlüsselnummern und die Synonymielisten festzulegen.

Viel schlimmer und mit Mitteln der Informatik prinzipiell unlösbar ist aber der echte Fortschritt in der Systematik, wenn neue Untersuchungen und Revisionen zu neuen Artabgrenzungen führen. Durch neue Erkenntnisse werden manche Arten in zwei oder mehrere Arten aufgespalten, aus 2 Arten werden 3 Arten identifiziert. Manchmal werden auch zwei Arten zusammenlegt und manchmal werden sie Jahrzehnte später wieder getrennt! (Für Unerschrockene gibt es weiter Beispiele in GEISER 2001.)

Nomenklatur bei einigen Arten der Bergblattkäfergattung *Oreina* bzw. *Chrysochloa*, Synonyme in eckigen Klammern:

Oreina bidentata BONTEMS

[*Oreina luctuosa* auct. nec OLIVIER = *Oreina bidentata tenebrosa* WEISE
Chrysochloa tristis sensu WEISE]

© Biologiezentrum Linz/Austria; download unter www.biologiezentrum.at

Oreina caerulea (OLIVIER)
[*Chysochloa rugulosa* (SUFFRIAN)
Chrysochloa regulosa (MOHR)
Oreina luctuosa (OLIVIER) sensu BONTEMS)]

Oreina canavesi BONTEMS wird von manchen Autoren als eigene Art, von anderen als Unterart vom *Oreina caerulea* aufgefasst. *Oreina collucens* DANIEL wird von manchen Autoren als eigene Art, von einigen als Unterart von *Oreina caerulea* und wieder von anderen als Unterart von *Oreina bidentata* angesehen.

Die Sammlungen und die Literatur, die aus einem Zeitraum von über 100 Jahren in tiergeografische Datenbanken eingespeichert werden, beruhen daher auf ganz unterschiedlicher Systematik. Für manche Tiergruppen gibt es mehrere konkurrierende Systeme, die von verschiedenen Autoren zur selben Zeit und in unterschiedlicher Weise verwendet werden.

Für welches System soll sich nun eine tiergeografsche Datenbank entscheiden? Ernst Reichl hat dieses Problem in der ZOODAT in einer sehr sinnvollen und zweckmäßigen Art und Weise gehandhabt, für die er von Beginn an und noch posthum kritisiert wurde:

Als Bezugssystem für die Artnummern verwende man nicht die neueste Nomenklatur, sondern die der letzten, großen zusammenfassenden Standardwerke einer Tiergruppe.

Eine tiergeografische Datenbank braucht ein Bezugssystem für die Artnummern. Das wurde von vielen Kritikern nicht verstanden, die dieses Bezugssystem der ZOODAT als "falsche Systematik" beanstandet haben[12]. Den Biologen unter den Kritikern war die Bedeutung eines Primärschlüsselfeldes für eine Datenbanktabelle nicht klar und für die Informatiker unter den Kritikern waren die Abgründe der zoologischen Systematik weitgehend unverständlich. (Letzteres gilt übrigens auch für die Mehrheit der Biologen!) Genaugenommen geht es hier aber gar nicht um Systematik, sondern um Nomenklatur, also darum, wie wir dieses real vorhandene abgestufte hierarchische System innerhalb der Organismen durchschauen oder auch nicht, bzw. wie wir daher die von uns erkannten Entitäten benennen.

Für dieses komplexe Problem gibt es – wie so oft in der Biologie und Informatik – keine einfache allgemeingültige Lösung. Hier benötigt man eine umfassende Kenntnis der jeweiligen Tiergruppe und Augenmaß, übrigens eine wesentliche Eigenschaft jedes guten Informatikers!

Empfehlenswert ist es, sich bei jeder Tiergruppe an das letzte große, möglichst umfassende Standardwerk zu halten. Das löst zwar auch nicht alle Probleme. Bei den Käfern Mitteleuropas ist das letzte umfassende Standardwerk der Katalog von LUCHT 1987, dessen Artnamen sich aber in den 3 Nachtragsbänden zum 11-bändigen Werk "Die Käfer Mitteleuropas" und mit dem Nachtragsband zu den Nachtragsbänden in großer Zahl

[12] Es gibt übrigens weder eine "richtige" noch eine "falsche" Systematik, auch wenn ein Autor einer systematischen Revision seine Ergebnisse auf seinem Spezialgebiet naturgemäß als "richtige" Systematik auffasst. Es gibt systematische Einteilungen, die relativ plausibel begründet sind und solche, bei denen man sich nur wundern kann, weshalb sie von einer Vielzahl von Fachkollegen akzeptiert werden. Aber ich selbst verwende auch nicht die "Systematik" oder die Nomenklatur, von deren Überlegenheit gegenüber anderen Systemen ich überzeugt bin, sondern meistens diejenige, die der jeweilige Schriftleiter fordert.

© Biologiezentrum Linz/Austria; download unter www.biologiezentrum.at

geändert haben. Für die Chrysomeliden Mitteleuropas liegt mit WARCHALOWSKI 2003 ein Standardwerk vor, das sogar auf den systematischen Stand von vor 1987 zurückgreift, weil viele der darin behandelten Käferarten noch nicht revidiert sind und sie sich sonst nicht auf sinnvolle Weise zu den mitteleuropäischen, besser bearbeiteten Arten zuordnen lassen.

Diese Situation in der zoologischen Systematik macht nicht nur den Betreibern einer tiergeografischen Datenbank bei der Dateneingabe und bei der Auswertung das Leben schwer. Sie ist auch ein wesentliches Hindernis für eine Vernetzung mehrerer tiergeografischer Datenbanken zur Erstellung von europaweiten Verbreitungskarten. Allerdings ist die Systematik dabei keineswegs das größte Hindernis, wie man auf Grund der obigen Ausführung annehmen möchte. Viel gravierender ist derzeit die Tatsache, dass die anderen tiergeografischen Datenbanken noch lange nicht auf einem so fortgeschrittenen Stand wie ZOBODAT sind!

9. Die sozio-biologischen Faktoren: die Zusammenarbeit mit den Spezialisten bei der Dateneingabe und Auswertung

Ich habe schon mehrmals betont, wie wichtig für eine tiergeografische Datenbank die Zusammenarbeit mit Spezialisten ist. Gerade im Bereich Systematik und Art-Grunddaten ist eine intensive Zusammenarbeit notwendig, aber gleichzeitig riskant. Ein Spezialist neigt dazu, immer die "neueste" Systematik verwenden zu wollen, natürlich nur auf seinem Spezialgebiet! Hier einen pragmatischen Kompromiss durchzusetzen, ohne dabei eine kritische Masse (im doppelten Sinn!) eines für eine tiergeografische Datenbank so wichtigen Personenkreises dauerhaft zu verärgern, gehört zu den wertvollsten und wichtigsten "menschlichen Faktoren" eines Datenbankbetreibers.

Die Zusammenarbeit mit Spezialisten bzw. Personen, die die tiergeografische Datenbank für ihre Forschungsprojekte nutzen, hat einen starken Einfluss auf die Qualität der Daten. Auch bei sorgfältiger Qualitätskontrolle der Daten vor und nach der Eingabe kommt es immer wieder vor, dass fehlerhafte Daten eingespeichert werden. Fundortverwechslungen oder einzelne Fehldeterminationen, die auch in einer sehr zuverlässigen Datenquelle vorhanden sind, lassen sich nicht immer ausschließen. Solche fehlerhaften Daten bemerkt man aber bei der Auswertung: ein abweichender Rasterpunkt auf einer Verbreitungskarte oder ein Artname in einer Faunenliste eines Gebietes, in dem diese Art eigentlich nicht vorkommen "dürfte" . Der auswertende Spezialist (und nur dieser!) erkennt solche fehlerhaften Daten, die nun endlich korrigiert werden können. Daher erhöht die Zusammenarbeit mit Spezialisten nicht nur die Quantität, sondern auch die Qualität der Daten einer tiergeografischen Datenbank.

Ebenso sind es die Spezialisten, die für bestimmte Fragestellungen ihrer Forschungsvorhaben neue Auswertungsmöglichkeiten benötigen. Hier ergibt sich ein sinnvolle Symbiose mit einer tiergeografischen Datenbank: Die Datenbanken bekommen von den Spezialisten die "Ideen" für neue Auswertungsmöglichkeiten geliefert. Dann programmiert man die neue Auswertungsroutine und stellt die damit erstellten Auswertungen dem Spezialisten zur Verfügung. Andererseits wird die Datenbank durch die neue Auswertungsroutine, die sie nun implementiert hat, selbst ebenfalls qualitativ aufgewertet.

© Biologiezentrum Linz/Austria; download unter www.biologiezentrum.at

10. Zusammenstellung der wirklich entscheidenden Faktoren

Die Berücksichtigung folgender Faktoren ist für den erfolgreichen Dauerbetrieb einer biogeografischen Datenbank essentiell:

1. Beschränkung auf die Datenfelder, die wirklich notwendig und sinnvoll sind
2. Verwendung eines relationalen Datenbanksystems mit optimaler Tabellenstruktur
3. Ergonomischer Eingabemodus, der für die jeweilige Tiergruppe optimiert ist
4. Intensive Zusammenarbeit mit Spezialisten der einzelnen Tiergruppen
5. Pragmatisches Vorgehen bei der Auswahl der Nomenklatur:
   Bezug auf umfassende Standardwerke für die jeweiligen Tiergruppen
6. Intensive Verwendung der Daten für verschiedene Auswertungen
   => Fehlerbehebung
   => Erschließung neuer, qualitativ hochwertiger Datenquellen
   => Erstellung neuer Auswertungsmöglichkeiten

11. Budgeteinsparung an der richtigen Stelle

Zum Schluss möchte ich noch auf eine sehr häufige und leider sehr aktuelle Fehlerquelle beim Betrieb von tiergeografischen Datenbanken hinweisen. Wie generell in Universitäten und im Forschungsbereich der öffentlichen Hand üblich, werden auch bei den dem öffentlichen Sektor zugehörigen tiergeografischen Datenbanken zahlreiche Routinetätigkeiten von hochqualifizierten Personen durchgeführt, die durchaus von niedriger qualifiziertem Personal durchführbar wären. Der gegenwärtige Informatiker der ZOBODAT ist ein besonders hoch qualifizierter Datenbankadministrator und einer der wenigen Spezialisten für die Kartendarstellungen vernetzter Datenbanken und deshalb als Mitarbeiter bei internationalen Projekten sehr begehrt. Zu seinen Dienstverpflichtungen gehört es aber auch, das Netzwerk der oberösterreichischen Museen zu betreuen, eine Tätigkeit, die ein engagierter HTL-Absolvent ebenso durchführen könnte[13].

Während in der Privatwirtschaft aus Kostengründen(!) jeder Mitarbeiter dazu angehalten wird, sämtliche Arbeiten, die ein Kollege in einer niedrigeren Gehaltsstufe durchführen kann, an diesen zu delegieren, werden im öffentlichen Wissenschaftsbetrieb nach wie vor Techniker und Laboranten eingespart und promovierte und habilitierte Mitarbeiter mit Aufgaben der Lagerverwaltung oder mit technischen und administrativen Routinearbeiten in erheblichem Ausmaß betraut. Auf diese Weise werden finanzielle Ressourcen verschwendet und der Betreiber einer tiergeografischen Datenbank, der ja auf öffentliche Mittel angewiesen ist, steht vor der nächsten Budgetkürzung, die ihn womöglich zwingt, weitere niedrig qualifizierte Tätigkeiten "nach oben" zu verlagern.

Auch in einem umfangreichen Artikel kann man nicht alle Aspekte der tiergeografischen Datenbanken erschöpfend behandeln. Ich hoffe, dass meine Formulierungen meinen Artikel sowohl für datenbankinteressierte Biologen als auch für an der Fragestellung

---

[13] Momentan wird zwar eine solche Technikerstelle von EU-Mitteln zu Entlastung des Informatikers eingerichtet, allerdings nur befristet auf 3 Jahre. Das Problem bleibt also weiterhin virulent!

© Biologiezentrum Linz/Austria; download unter www.biologiezentrum.at

interessierte Informatiker verständlich gemacht haben. Vielleicht können meine Ausführungen in Zukunft etwas dazu beitragen, Konzepte und Betrieb von tiergeografischen Datenbanken im deutschen Sprachraum zu verbessern, damit für die Tiergeografen endlich länderübergreifende Kartendarstellungen und Auswertungen ermöglicht werden.

12. Danksagung

Dem Informatiker Dipl.Ing. Michael Malicky danke ich für wertvolle fachliche Hinweise. Dem Nicht-Informatiker, aber Zoologen Dipl.biol. Remigius Geiser danke ich für zahlreiche Diskussionen zu den angeschnittenen Themen.

13. Zusammenfassung

Biogeografische Datenbanken sind zu unverzichtbaren Werkzeugen für die Biodiversitätsforschung und für naturschutzrelevante Entscheidungen geworden. Europaweit sind in den letzten Jahrzehnten zahlreiche biogeografische Datenbanken von regionalen und überregionalen Institutionen gegründet worden. Manche scheiterten bereits in der Gründungphase, andere bringen auch nach jahrelangem und kostenintensivem Betrieb nur magere Ergebnisse zustande. Die (lange Zeit rein tiergeografische) Datenbank ZOODAT, die der Linzer Informatiker Ernst Reichl 1972 gründete, entwickelte sich rasch zu einer der erfolgreichsten tiergeografischen Datenbanken und hält bis heute (nun ZOBODAT genannt) einen Spitzenplatz in Europa. Dieser Erfolg der ZOBODAT beruht auf folgenden Faktoren:

- Beschränkung auf die wenigen, unbedingt notwendigen Datenfelder
- Verwendung eines relationalen Datenbankmanagementsystems mit optimaler Tabellenstruktur
- Ein ergonomischer, für die jeweilige Tiergruppe optimierter Eingabemodus
- Intensive Zusammenarbeit mit Spezialisten
- Pragmatisches Vorgehen bei Problemen der Systematik und Nomenklatur
- Intensive Verwendung der Daten, wodurch die Datenbank quantitativ und qualitativ aufgewertet wird

Diese Faktoren werden am Beispiel der ZOODAT ausführlich erläutert.

14. Literatur

BEUTELSPACHER A. & M.A. ZSCHIEGNER (2002): Diskrete Mathematik für Einsteiger. — Vieweg Verlag Braunschweig / Wiesbaden: 1-216.

FREUDE H. et al. (1964-1998): Die Käfer Mitteleuropas, Band 1 bis 15. — Goecke & Evers, Krefeld.

GEISER E. (1996): Probleme der Datenerfassung, Determination und Nomenklatur bei umfangreichen faunistischen Datenbanken. — Verhandlungen des 14. Internationalen Symposiums über Entomofaunistik in Mitteleuropa (SIEEC), 4.-9. September 1994, München: 344-347.

GEISER E. (1998a): Die tiergeographische Datenbank ZOODAT, eine wertvolle Datenbasis für die Biodiversitätsforschung. — Stapfia (Linz) **55**: 35-46.

© Biologiezentrum Linz/Austria; download unter www.biologiezentrum.at

GEISER E. (1998b): Wie viele Tierarten leben in Österreich? Erfassung, Hochrechnung und Abschätzung. — Verhandlungen der Zoologisch-Botanischen Gesellschaft in Österreich, Band **135**: 81-93.

GEISER E. (2001): Risiken und Nebenwirkungen einer umfangreichen entomologischen Datenbank. — Entomologica Austriaca **4**/2001: 15-18.

LUCHT W. (1987): Die Käfer Mitteleuropas. Katalog. — Goecke & Evers, Krefeld: 342 pp.

MALICKY M. (2001): Die biogeografische Datenbank ZOBODAT (ehemalige ZOODAT) am OÖ. Landesmuseum/Biologiezentrum Linz, Oberösterreich. – eine öffentliche Einrichtung für Wissenschaft und Naturschutz im Umfeld nationaler und internationaler Kooperationsmöglichkeiten. — Entomologica Austriaca **4**: 8-14.

REICHL E.R. (1993): Ein Maß für die Schutzwürdigkeit von Arten und Biotopen. – Proceedings of the 9th International Colloquium of the European Invertebrate Survey, Helsinki. — WWF Finland Report 7.

WARCHALOWKSI A. (2003): Chrysomelidae. The Leaf-Beetles of Europe and the Mediterranean Area. — Natura Optima Dux Foundation, Warszawa 2003: 600 pp.

Anschrift der Verfasserin: Dr. Elisabeth GEISER
Saint-Julien-Str. 2/314
A-5020 Salzburg, Austria
E-Mail: geiser@salzburg.co.at

ZOBODAT - www.zobodat.at

Zoologisch-Botanische Datenbank/Zoological-Botanical Database

Digitale Literatur/Digital Literature

Zeitschrift/Journal: Linzer biologische Beiträge

Jahr/Year: 2005

Band/Volume: 0037_1

Autor(en)/Author(s): Geiser Elisabeth

Artikel/Article: Die entscheidenden Faktoren für den langfristigen Erfolg einer tiergeografischen Datenbank - am Beispiel der ZOBODAT (früher: ZOODAT) 49-69