

# Aufbau einer „Datenbank der Datenquellen“

## – Toxikologische Informationen zu Chemikalien

Kristina Voigt

Gesellschaft für Strahlen- und Umweltforschung mbH, Ingolstädter Landstraße 1, D-8042 Neuherberg

W. Mücke

Bayerisches Staatsministerium für Landesentwicklung und Umweltfragen, Rosenkavalierplatz 2, D-8000 München 81

**Zusammenfassung.** Zur toxikologischen Bewertung von Altstoffen im Sinne des Chemikalienrechts ist ein Minimalsatz an gesicherten Daten unabdingbar notwendig. Da national und international eine Fülle von Datenquellen auch toxikologische Daten beinhalten, geht es unseres Erachtens weniger darum, eine neue Datenbank auf dem Gebiet der Toxikologie zu erstellen, sondern rasche Zugriffsmöglichkeiten auf schon vorhandene toxikologische Informationen zu schaffen.

Für den Aufbau einer solchen „Datenbank der Datenquellen“ für toxikologische Informationen (DaToDa) wird zunächst eine Auswahl der toxikologisch-relevanten Datenquellen aus der Gesamtheit an Datenquellen vorgenommen. Dazu werden die Datenquellen gemäß der Anwendung der Chemikalien sowie der toxikologischen Parameter unterschieden. Das Kernstück der Arbeit bilden die aufzubauenen *Datenfelder*, wobei neben bibliographischen und inhaltserschließenden Datenfeldern den Informationsretrieval-Hinweisen besondere Bedeutung zukommt; darunter sind zu verstehen: Suchbegriffe, Recherche-Anweisungen, Querverweise auf andere relevante Datenquellen sowie die Verfügbarkeit der Quellen. Besonders wichtig wird die Pflege und Aktualisierung dieser Datenbank sein. Zur Überprüfung der Arbeiten wird ein anwendungsbezogenes Chemikalienset mit ca. 100 Chemikalien herangezogen.

- 1 000 bis 10 000 Jahrestonnen ca. 700 Stoffe
- 10 000 Jahrestonnen ca. 400 Stoffe [1].

In der Bundesrepublik Deutschland ist das Chemikaliengesetz am 1. 1. 1982 in Kraft getreten [2]. Es schreibt vor, daß Stoffe vor dem Inverkehrbringen angemeldet werden müssen (sog. „Neue Stoffe“), wobei ein Satz an Substanzdaten vorgelegt werden muß. *Prüfnachweise*, wie sie für Neue Stoffe gefordert werden, sind für die Gesamtzahl der bereits im Verkehr befindlichen Stoffe (sog. „Alte Stoffe“) angesichts des hohen Kosten- und Zeitaufwandes im Gesetz nicht vorgesehen. Allerdings können Alte Stoffe einer Prüfpflicht unterworfen werden, wenn sich tatsächliche Anhaltspunkte dafür ergeben, daß sie „gefährlich“ im Sinne des Gesetzes sind. Mit der *Prioritätensetzung* ist das bei der Gesellschaft Deutscher Chemiker eingerichtete Beratergremium für umweltrelevante Altstoffe (BUA) befaßt. Voraussetzung für die Bewältigung des Altstoffproblems ist in jedem Fall, mit einer zwar unzureichenden Datenbasis zu beginnen, die vorhandenen Daten jedoch so gut wie möglich zu nutzen. Die verbleibenden Wissenslücken müssen dann entsprechend den festgelegten Prioritäten durch weitere Untersuchungen geschlossen werden.

Die entscheidende Voraussetzung für ein ökonomisches Vorgehen ist eine möglichst umfassende Kenntnis von Aufbau und Inhalt der schon bestehenden Datenquellen über Chemikalien. Aus diesem Grund ist ein systematischer Aufbau eines EDV-gestützten Systems, d.h. einer Datenbank der Datenquellen, sinnvoll. In dieser Arbeit werden toxikologische Fragestellungen angesprochen, die im Rahmen einer „Datenbank für toxikologisch-relevante Datenquellen“ (DaToDa) bearbeitet werden (die hier abgedeckten toxikologischen Parameter → *Abschnitt 4.1 und Tabelle*). Der Aufbau einer eigenen Datenbank mit toxikologischen Daten und Informationen, sei es in Zitatform oder in Form von Fakten, erscheint uns angesichts der Vielzahl der bereits existierenden Datenbanken auf diesem Gebiet derzeit weder realistisch noch sinnvoll.

### 1 Problemstellung

Unter den acht Millionen Einträgen des „Chemical Abstract Service“ werden nach dem derzeitigen Stand des „European Inventory of Existing Commercial Substances“ (EINECS) ca. 100 000 verschiedene Chemikalien kommerziell hergestellt. Davon wird jedoch nur ein verhältnismäßig geringer Anteil in größerer Menge produziert. So hat eine 1986 vom Verband der Chemischen Industrie unter den Mitgliedsfirmen in der Bundesrepublik Deutschland durchgeführte Umfrage ergeben: Im Jahr 1985 wurden insgesamt 4 600 Stoffe in Mengen von über 10 Jahrestonnen hergestellt:

- 10 bis 1 000 Jahrestonnen ca. 2 200 Stoffe
- 100 bis 1 000 Jahrestonnen ca. 1 300 Stoffe

## 2 Sachstand

Als erster Schritt zur Aufhellung des Problems der *Datenverfügbarkeit* wurde unter Berücksichtigung von Vorarbeiten für eine OECD-Expertengruppe [3] eine Aufstellung von Datenquellen für Umweltchemikalien erarbeitet [4]. Dazu wurden 714 Datenquellen für Chemikalien (Handbücher, Enzyklopädien, Firmenverzeichnisse, Reports, Reviewbände etc.) und Datenbanken kritisch auf Datentypen durchgesehen und mit Registern aufbereitet. Darüberhinaus wird, darauf aufbauend, ein weitgehend EDV-gestütztes System für einen Zugriff auf Datenquellen für Umweltchemikalien entwickelt [5, 6, 7].

Beide Ansätze beziehen sich auf eine Vielzahl von *Datentypen* – Identifizierungsmerkmale, ökonomische Daten und Anwendungsformen, physikalisch-chemische Parameter, Abbau- und Akkumulationsverhalten – und schließen auch toxikologische Parameter ein. Als vorläufiges Ergebnis dieser Arbeiten kann festgestellt werden, daß sich „toxikologische Informationen“ für ca. 30 % der Alten Stoffe auffinden lassen. Hinsichtlich der Struktur der Datenquellen unterscheiden wir gedruckte Dokumente und computer-gestützte Datenbanken.

### 2.1 Gedruckte Dokumente

Für den Bereich der Toxikologie existieren ca. zehn Standardwerke\*. Der Aufbau dieser Datenquellen ist unterschiedlich; die meisten gedruckten Dokumente verfügen über Indices, die das Auffinden der gewünschten Chemikalie erleichtern. Beispiele für Indices sind: Chemikalien-Index (chemische Bezeichnungen oder Synonyma und Handelsnamen), Synonyma-Index, CAS-Nummern-Index, Strukturformel-Index, Summenformel-Index, Molekulargewicht-Index, Schmelzpunkt/Siedepunkt-Index, Chemikaliengruppen-Index und Anwendungs-Index.

Bedauerlicherweise sind die meisten gedruckten Dokumente nach den Chemikalienbezeichnungen aufgebaut und weisen nicht die eindeutigere Einteilung nach den CAS-Nummern auf.

### 2.2 Datenbanken

Datenbanken lassen sich in drei Kategorien unterteilen: Bibliographische Datenbanken, numerische oder Faktendatenbanken und Volltextdatenbanken.

In den **bibliographischen Datenbanken** können die Zitate (mit und ohne Kurzreferat), in denen sich die gewünschten Informationen befinden, abgefragt werden. Die relevanten Artikel müssen dann entweder „online“ bestellt oder über Bibliotheken beschafft werden. **Faktendatenbanken** beinhalten Daten mit Zitatangabe; in Zweifelsfällen ist ebenfalls die Beschaffung der Originalliteratur angezeigt. In **Volltextdatenbanken** sind die vollständigen Artikel abgespeichert und können dementsprechend abgefragt werden (diese Datenbanken sind die teuersten).

### Toxikologisch relevante Datenbanken [8]:

- BIOSIS Previews, CA (Chemical Abstract Service), CAB (Commonwealth Agricultural Bureaux), CHI (Chemical Hazards in Industry), HSELINE (Health and Safety Executive Line), MEDLARS (Index Medicus), PHYTO-MED, TOXLINE (**Bibliographische Datenbanken**);
- AGRC (Agrochemicals Databank), AQUIRE (Aquatic Information Retrieval System), CTCP (Clinical Toxicology of Commercial Chemical Products), ECDIN (European Communities Data Information Network), EDAP (European Database of Agricultural Products) GENOTOX (Genetic Toxicity Databank), RTECS (Registry on Toxic Effects of Chemical Substances) (**Faktendatenbanken**);
- KIRK (Kirk-Othmer Encyclopedia of Chemical Technology) (**Volltextdatenbank**).

### 2.3 Bisherige Auswertungsansätze

Die aufgeführten Datenbanken sind – ebenso wie die gedruckten Dokumente – unterschiedlich aufgebaut; daher wird für jede Datenbank eine eigene *Suchstrategie* benötigt, um die gewünschten Informationen zu erhalten. Darüberhinaus liegen die angesprochenen Quellen bei verschiedenen Datenbank-Anbietern (Hosts) auf. Jeder einzelne Host besitzt seine eigene Retrieval-, d.h. Abfragesprache.

Daraus folgt, daß man nicht nur über das toxikologische Fachwissen, sondern auch über den Datenbankaufbau und über Kenntnisse mehrerer Retrievalsprachen verfügen muß, um die Vielzahl der Datenquellen effektiv nutzen zu können. Erleichtert wird die Suche dadurch, daß heute bereits Softwarepakete angeboten werden, die durch Einführung eines Menu-gesteuerten Systems die Kenntnis der betreffenden Retrievalsprachen überflüssig werden lassen. Ein solches Softwarepaket ist beispielsweise Micro-CSIN, mit dessen Hilfe in einigen Faktendatenbanken des Datenbank-anbieters CIS recherchiert werden kann [9]. Mit dem Software-Paket SCI-Mate (Universal Online Searcher) können ebenfalls Recherchen bei den amerikanischen Hosts DIALOG, BRS, NLM und SDC mit einer Menu-Technik durchgeführt werden [10]. Beide Softwarepakete sind z.Zt. jedoch noch nicht auf die großen europäischen Hosts anwendbar wie z.B.

- DIMDI (Deutsches Institut für Medizinische Dokumentation und Information),
- ESA (European Space Administration),
- Datastar oder
- STN (Scientific and Technical Information Network).

Eine Studie zur Bewertung von Informationsquellen in der chemischen Toxikologie hat ergeben, daß für gewisse toxikologische Datentypen „Chemical Abstracts“ und „BIOSIS“ die wertvollsten Datenquellen sind [11]. Für diese Untersuchung wurden nur bibliographische, jedoch keine Faktendatenbanken herangezogen. Außerdem wurde diese Arbeit bereits Anfang der achtziger Jahre durchgeführt. Da sich das Datenbankangebot seither sehr stark erweitert hat, ist die Aussage u.U. nicht mehr zutreffend.

Ein weiteres Hilfsmittel sind sog. **Datenbankführer**, entweder in gedruckter Form und/oder als „Online“-Datenbank.

\* Nachweis bei den Autoren

Der größte Katalog der „online“ angebotenen Datenbanken, „CUADRA Directory of Databases“, ist sowohl in gedruckter Form als auch als Datenbank erhältlich [12]. Dieser Datenbankführer umfaßt Datenbanken, die weltweit „online“ von den verschiedenen Hosts angeboten werden. Er schließt selbstverständlich auch Datenbanken mit toxikologischen Inhalten ein. Allerdings ist er in bezug auf die inhaltliche Erschließung der Datenbanken sehr allgemein gehalten, so daß man sich zwar einen Überblick der Datenbanken verschaffen kann, jedoch keine Hinweise auf die zu nutzenden Schlagwörter, die eigentliche Recherche und Querverweise zu anderen Quellen erhält. Auch sind hier keine gedruckten Dokumente gespeichert, die für manche Recherchen bereits ausreichend sind.

### 3 Auswahl toxikologisch-relevanter Datenquellen

Vordringlich erscheint es, in Zukunft eine geeignete *Auswahl* an Daten- und Informationsquellen für toxikologische Fragestellungen zu treffen und diese Quellen in einer Datenbank toxikologisch-relevanter Datenquellen (DaToDa) so aufzubereiten, daß auch nicht ständig mit Datensuche befaßte Wissenschaftler einen schnellen Datenzugang gewinnen. Aus praktischen Gründen wird die Auswahl einerseits aus für geeignet erachteten Quellen, andererseits aus leicht zugänglichen Quellen erfolgen müssen. Eine erste Unterscheidung läßt sich nach anwendungs- und parameterspezifischen Gesichtspunkten treffen.

#### 3.1 Anwendungsspezifische Unterscheidung der Datenquellen

Es erweist sich nur in wenigen Fällen als sinnvoll, sämtliche Quellen für eine bestimmte Fragestellung nacheinander durchzusuchen. Bei der Auswahl der Datenquellen kommt darüberhinaus dem *Verwendungszweck* einer Chemikalie eine Schlüsselgröße zu [13]. Der Grund liegt darin, daß es verschiedene Datenquellen gibt, die sich auf bestimmte Chemikalien gemäß ihrer Anwendung spezialisieren. Dazu gehören gedruckte Dokumente und Datenbanken.

##### 3.1.1 Gedruckte Dokumente

Umfangreiche Dokumente liegen für Pestizide, Pharmazeutika, Polymere, Lösemittel und Farbstoffe vor\*.

##### 3.1.2 Datenbanken

###### PESTIZIDE

- AGRC (Agrochemicals Databank), Faktendatenbank;  
Host: DATASTAR
- EDAP (European Database of Agrochemical Products), Faktendatenbank;  
Host: DATASTAR

###### PHARMAZEUTIKA

- MART (Martindale Online), Faktendatenbank;  
Host: DATASTAR

- ABDA (ABDA-Fertigarzneimittel), Faktendatenbank;  
Host: DIMDI
- LINE (Pharmline), bibliographische Datenbank;  
Host: DATASTAR
- DELT (Drug Effects on Laboratory Tests), Faktendatenbank;  
Host: DATASTAR

###### POLYMERE UND DEREN MONOMERE

- KKF (Kunststoffe Kautschuk Fasern), bibliographische Datenbank;  
Host: STN
- RAPRA Abstracts (Rubber and Plastics Research Association), bibliographische Datenbank;  
Host: STN, Pergamon Infoline

###### FARBSTOFFE UND LÖSEMittel

- WSCA (World Surface Coatings Abstracts), bibliographische Datenbank;  
Host: ORBIT, Pergamon Infoline

Man wird also beispielsweise nach Pestiziden zunächst in den o.g. Datenquellen recherchieren. Kosten und Zeit, die für die Informationsgewinnung benötigt werden, sind bei diesem Vorgehen geringer als bei der herkömmlichen Recherche in Datenbank-unspezifischen Quellen.

Bei der Auswahl der Quellen sollte auch das Kriterium „Zugriffsmöglichkeit auf die Quelle“ Berücksichtigung finden. Gedruckte Dokumente, die einer Arbeitsgruppe vorliegen, sind in der Regel immer greifbar, während Online-Datenbanken am Abend, am frühen Morgen sowie an Wochenenden meist nicht verfügbar sind. Auch sind Störungen im Hostrechner oder bei der eigenen Hardware zu berücksichtigen.

#### 3.2 Parameterspezifische Unterscheidung von Datenquellen

Neben den anwendungsspezifischen lassen sich parameterspezifische Datenquellen unterscheiden; sie weisen vorwiegend Informationen über eine oder mehrere spezielle Eigenschaften von Stoffen bzw. bestimmte Stoffmerkmale auf.

##### 3.2.1 Gedruckte Dokumente

Für toxikologische Parameter sind z.T. zusammenfassende, z.T. aber auch spezifische gedruckte Dokumente verfügbar\*.

##### 3.2.2 Datenbanken

Auch bei den Datenbanken lassen sich parameterspezifische Quellen unterscheiden:

- CANCERLIT (Cancer Literature), bibliographische Datenbank;  
Host: DIMDI (Parameter: Kanzerogenität)
- CTCP (Clinical Toxicology of Commercial Products), Faktendatenbank;  
Host: CIS (Parameter: akute Toxizität)

\* Nachweis bei den Autoren

- DERMAL (Dermal Absorption), Faktendatenbank;  
Host: CIS (Parameter: akute dermale Toxizität)
- GENETOX (Genetic Toxicity), Faktendatenbank;  
Host: CIS (Parameter: Mutagenität)
- GIABS (Gastrointestinal Absorption Database), Faktendatenbank;  
Host: CIS; (Parameter: Absorption, Verteilung und Metabolismus)
- IRCS (Medical Science Research), Volltextdatenbank;  
Host: DIMDI, DATASTAR (Parameter: Metabolismus, akute und chronische Toxizität).

Über anwendungsspezifische Datenquellen sowie parameterspezifische Datenquellen hinaus sollten auch Datenquellen übergreifender Art wie z.B. CAS-Online, TOXALL, BIOSIS Previews, ECDIN etc. Berücksichtigung finden, da die Datenbestände dieser Quellen mehrere Sachgebiete, Parameter und Anwendungsformen von Chemikalien enthalten.

#### 4 Festlegung der Datenfelder und Durchführung

Um aus der getroffenen Auswahl eine „Datenbank der Datenquellen für toxikologische Informationen“ herzustellen, sind zunächst die sog. **Datenfelder** zu definieren. Bei den meisten herkömmlichen Datenbanken unterscheidet man in diesem Zusammenhang zwischen den „bibliographischen“ Datenfeldern – sie enthalten bibliographische Angaben – und „inhaltserschließenden“ Feldern – sie dokumentieren den Inhalt der Datenbank mehr oder weniger genau. Diese beiden Arten von Feldern wurden bereits in unserem weiter gefaßten Ansatz realisiert [5, 6]. Sie müssen selbstverständlich auch in die Datenbank für toxikologische Informationen integriert werden. Als neuer Felder-Oberbegriff werden „pragmatische Informationsretrieval-Hinweise“ eingeführt; sie sollen hier vordringlich erarbeitet werden.

##### 4.1 Erarbeitung der geeigneten Suchbegriffe für toxikologisch-relevante Datenquellen

Für toxikologische Fragestellungen werden zumindest die in der folgenden Tabelle aufgeführten Parameter benötigt:

Tabelle: Parameter für toxikologische Fragestellungen

- 
- Allgemeiner Wirkungscharakter
  - Wirkungsmechanismus
  - Metabolismus, Toxikokinetik (Resorption, Verteilung, Ausscheidung)
  - Akute und subakute Toxizität
  - Haut- und Schleimhautverträglichkeit
  - Sensibilisierende Wirkung
  - Subchronische und chronische Toxizität
  - Mutagenität (in vitro, in vivo)
  - Kanzerogenität
  - Reproduktionstoxizität
  - Wirkungen auf das Immunsystem
  - Sonstige Wirkungen (z.B. Neurotoxizität, Kombinationswirkungen)
  - Erfahrungen beim Menschen (akute Vergiftungen, chronische Vergiftungen, epidemiologische Daten)
- 

Diese Parameter werden auch bei der Abfassung von Stoffberichten durch das Beratergremium für umweltrelevante Altstoffe (BUA) bei der Gesellschaft Deutscher Chemiker (GDCh) berücksichtigt [14]. Sicherlich wären weitere, darüber hinausreichende toxikologische Angaben wünschenswert; es müssen jedoch Kompromisse geschlossen werden zwischen den Anforderungen und den realistischen Möglichkeiten, toxikologische Informationen aus Datenquellen zu extrahieren. Insofern erfolgt hier eine Beschränkung auf die oben angegebenen Parameter.

Die Auswertung der Datenquellen sollte nicht nur zeigen, in welcher Quelle zu welchem Parameter Informationen vorliegen (vgl. 5), vielmehr sollen hier die konkreten *Suchbegriffe* und eventuelle Verknüpfungen dargestellt werden. Dazu müssen die entsprechenden *Thesauri* (Schlagwortkataloge) systematisch durchgearbeitet und relevante Schlagworte aufbereitet werden.

##### 4.2 Rechercheanweisungen

Da Datenquellen in aller Regel unterschiedlich aufgebaut sind, ist es notwendig, für die einzelnen Quellen eine Art „Leitfaden“ zu besitzen, nach dem in einer gegebenen Quelle eine Recherche durchgeführt werden kann.

Für ein **gedrucktes Dokument** wird eine solche Rechercheanweisung beispielsweise lauten: Aufbau der Quelle nach systematischen Namen der Chemikalien, Synonymregister ab Seite x, CAS-Nummern-Register ab Seite y. Hieraus könnte man entnehmen, daß zur eindeutigen Auffindung der gewünschten Chemikalien eine Registersuche angezeigt ist.

Bei **Datenbanken** werden sich umfangreichere Rechercheanweisungen ergeben: Neben den Abfragemöglichkeiten nach Chemikalien-Namen, CAS-Nummer, Summenformel und Molekulargewicht müssen die wichtigsten Besonderheiten der Retrievalsprache einbezogen werden:

- Angabe der Datenbank (BASE Datenbankname)
- Suche der Information (FIND Begriff)
- Suche der Chemikalien (FIND CR = xxxxx-xx-x; FIND; FT = Chemikalienname)
- Verknüpfung (FIND Profil-Nr. X AND/OR/NOT Profil-Nummer Y)
- Ausgabe der Ergebnisse (SHOW Profil-Nr.; BIB; Antwort 1-n)
- Ende des Dialogs (STOP).

Mit diesen Rechercheanweisungen in Zusammenhang mit den unter 4.1 beschriebenen Suchbegriffen ist nun in der Regel eine Recherche nach toxikologischen Informationen auch für den nicht mit den entsprechenden Datenquellen vertrauten Wissenschaftler möglich. Mit Hilfe der oben beschriebenen Anweisungen ist der Benutzer flexibel, sowohl in der Auswahl der *Schlagworte*, als auch in der Auswahl der *Chemikalienbezeichnung* (CAS-Nummer oder Chemikalienname).

##### 4.3 Querverweise

Neben gezielten Datenfeldern sollten auch andere Quellen als Ergänzung oder Alternative genutzt werden. Solche

Verweise basieren auf der praktischen Erfahrung im Umgang mit Datenquellen. Bei einer Datenbank wie Chemical Abstract Service könnte hier beispielsweise folgendes Vorgehen zweckmäßig sein:

- für den Parameter „dermale Toxizität“ zunächst in DERMAL recherchieren;
- bei Pestiziden zunächst im „Pestizide Manual“ suchen.

Für die Auffüllung dieser Rubrik „Querverweise“ sind die im Abschnitt 5 enthaltenen Hinweise zur Überprüfung der Datenquellen mittels eines Chemikaliensetsets zu berücksichtigen.

#### 4.4 Verfügbarkeit der Datenquellen

Bei Arbeiten mit Datenquellen spielt das Problem der Datenverfügbarkeit an sich und das der zugehörigen Hintergrundliteratur eine wesentliche Rolle. Praktische Hinweise sind die folgenden:

Bei den **gedruckten Dokumenten** sollten unter dem Feldoberbegriff „Verfügbarkeit“ Standort und Signatur sowie

bei Bibliotheken Hinweise auf Ausleihmöglichkeiten und Öffnungszeiten der Bibliothek vermerkt werden.

Bei den **Datenbanken** sollten Hinweise auf die Zugriffsmöglichkeit gegeben werden; in der Mehrzahl der „online“ verfügbaren Datenbanken sind Recherchen nur werktags etwa für 10 bis 12 Stunden möglich. Darüberhinaus ist bei den **bibliographischen Datenbanken** festzuhalten, wo die Hintergrundliteratur am schnellsten und kostengünstigsten beschafft werden kann. Da große toxikologische Datenbanken ca. 1 000 Fachzeitschriften abdecken, ist vorgesehen, das oben beschriebene Feld der Beschaffung von Hintergrundliteratur nur schwerpunktmäßig zu bearbeiten. Allerdings erscheint es wichtig, die Arbeit mittels eines *Testsets* zu überprüfen (s. unten).

Auch bei den **Faktendatenbanken** ist die Frage der Hintergrundliteratur wegen der eventuellen Überprüfung der Daten mittels Originalliteratur bedeutsam. Dasselbe gilt heute noch, wenn auch in geringerem Umfang, für **Volltextdatenbanken**: Hier kann zwar das vollständige Zitat „online“ abgefragt werden, jedoch bekommt man Tabellen, Formeln, Abbildungen etc. noch nicht oder nur unvollständig.

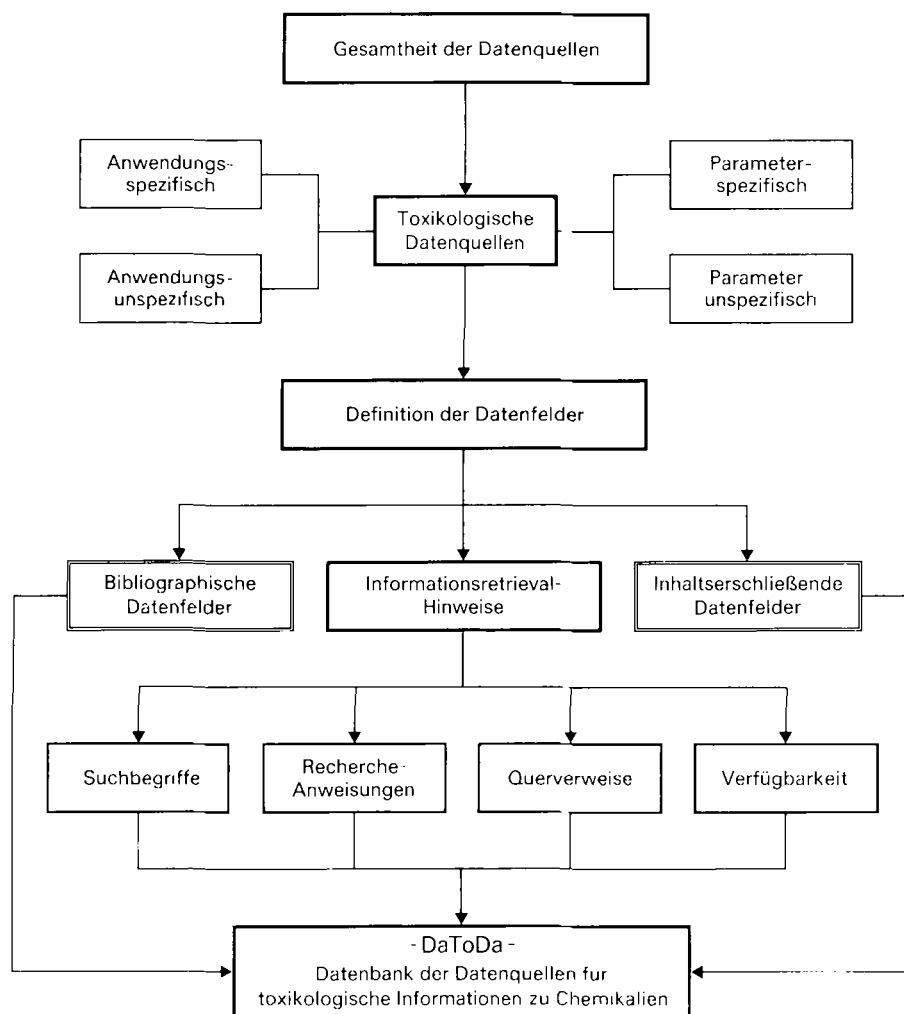


Abb.: Aufbau einer „Datenbank der Datenquellen“ für toxikologische Informationen zu Chemikalien

In der Abbildung ist der gesamte Aufbau der „Datenbank der Datenquellen“ für toxikologische Informationen (DaToDa) schematisch dargestellt:

## 5 Überprüfung, Pflege und Fortschreibung

Die oben beschriebenen Datenbankfelder und die dort vorgenommenen Eintragungen sollen mittels eines *Chemikaliensets* auf Verfügbarkeit, Vollständigkeit, Einheitlichkeit, eventuelle Fehler etc. überprüft werden. „Chemikalienset“ heißt in diesem Zusammenhang, daß eine Auswahl von ca. 100 Substanzen aus den verschiedenen Verwendungszwecken getroffen wird. Für diese Stoffe werden Recherchen in der oben angegebenen Weise durchgeführt, um die Datenbankfelder mit Informationen aufzufüllen. Es sollen dazu die schon oben erwähnten Verwendungszwecke (z.B. Pestizide, Pharmazeutika, Polymere, Farbstoffe und Lösemittel) herangezogen werden mit jeweils 10–20 Stoffen. Diese Auswahl erfaßt wesentliche Anwendungsgebiete von Chemikalien mit z.T. hohen Produktionsmengen. Soweit verfügbar, sollen diese als Auswahlkriterium für das Testset herangezogen werden.

Von besonderer Bedeutung ist die *Pflege einer Datenbank*. Ein gedrucktes Dokument ist in der Regel bei seinem Erscheinen schon veraltet. Eine Datenbank gleicht einem Dokument, wenn sie nicht ständig überprüft und auf den neuesten Stand gebracht wird. Da dies prinzipiell schnell und leicht durchführbar ist, können und werden sie meist auch aktualisiert und gepflegt, so daß „online“ angebotene Datenbanken in diesem Punkt gedruckten Dokumenten überlegen sind. Eine nicht gepflegte und aktualisierte Datenbank ist jedoch binnen zwei bis drei Jahren derart überholt und veraltet, daß sie nur noch sehr bedingt einsatzfähig ist. Innerhalb der aufzubauenden „Datenbank der Datenquellen“ für toxikologische Informationen zu Chemikalien (DaToDa) werden in diesem Zusammenhang *zwei Daueraufgaben* zu erfüllen sein: die Pflege der Datenbestände und die Aufnahme neuer Datenquellen.

### 5.1 Pflege und Aktualisierung vorhandener Datenbestände

Über die Eliminierung von Erfassungsfehlern hinaus sollen die Quellen in regelmäßigen Abständen, etwa vierteljährlich, überprüft und gegebenenfalls korrigiert oder ergänzt werden. Konkret würde das für *gedruckte Dokumente* bedeuten, daß neue Auflagen derjenigen Dokumente, die sich bereits in der Datenbank befinden, eingearbeitet werden. Die von den wissenschaftlichen Verlagen herausgegebenen Neuerscheinungslisten sowie die jährlich erscheinenden Standardnachsschlagewerke sollen Grundlage dieser Arbeit sein [15, 16, 17].

Auch bei den **Datenbanken** ist eine Pflege unerlässlich. Hier geben alle Datenbankanbieter bei der Veränderung ihres Datenbankangebotes „online“ und „offline“ Informationen sowie neue Datenbank-Beschreibungsblätter heraus. Die Inhalte müssen ausgewertet und in die Datenbank eingearbeitet werden.

### 5.2 Aufnahme neuer Datenquellen

Da ständig neue Datenquellen, in gedruckter Form oder „online“ auf den Markt gelangen, müssen diese Neuerscheinungen in die Datenbank inkorporiert werden. Bei den Büchern müssen hier wieder die Neuerscheinungskataloge der Verlage und die genannten Nachschlagewerke herangezogen werden. Bei den Datenbanken sollten in regelmäßigen Abständen die Datenbankführer, beispielsweise der „online“ angebotene CUADRA [12], abgefragt werden. Darüberhinaus weisen die bereits angemieteten Hosts jeweils auf neue von ihnen angebotene Datenbanken hin. Für die so vorgenommenen Aktualisierungen ist ein gesondertes Feld „Aktualisierungsstand“ vorgesehen.

## Literatur

- [1] Gesellschaft Deutscher Chemiker – Beratergremium für umweltrelevante Altstoffe: Altstoffbeurteilung, ein Beitrag zur Verbesserung der Umwelt, GDCh, Frankfurt am Main 1987
- [2] Gesetz zum Schutz vor gefährlichen Stoffen (Chemikaliengesetz – ChemG) vom 16. September 1980, (BGBl. I S. 1718, geändert durch Gesetz vom 15. 9. 1986, BGBl. I S. 1505)
- [3] OECD – Organization for Economic Cooperation and Development, OECD Existing Chemicals Program; Final Report of Expert Groups III and IV. Umweltbundesamt, Berlin 1985
- [4] K. VOIGT; H. ROHLEDER: Datenquellen für Umweltchemikalien. ecomed-Verlag, Landsberg/Lech 1986
- [5] W. MÜCKE; K. VOIGT; J. BENZ: System to Access Data Sources for Environmental Chemicals. Toxicol. Environ. Chem. 17, 237–247 (1988)
- [6] K. VOIGT; J. BENZ: Informationssystem für Umweltchemikalien, Materialien, Bayerisches Staatsministerium für Landesentwicklung und Umweltfragen (im Druck)
- [7] J. BENZ; K. VOIGT: Konzeption rechnergestützter Suchhilfen für die Beschaffung von Chemikaliendaten, in: A. Jaeschke, Page B., Informatikanwendungen im Umweltbereich, Informatik-Fachberichte 170. Springer Verlag, Berlin 1988
- [8] H. R. PICHLER: Online-Recherchen für Chemiker. VCH Verlagsgesellschaft mbH, Weinheim 1986
- [9] J. M. HUSHON: Strategy for Using Information Resources to Collect Chemicals Data, in: Environmental Modelling for Priority Setting among Existing Chemicals. ecomed-Verlag, Landsberg/Lech 1986
- [10] E. GARFIELD: Current Contents 8, 5–11 (1984)
- [11] D. M. SANDERSON: Methods of Data Retrieval-Manual, in: Richardson M., Toxic Hazard Assessment of Chemicals, The Royal Society of Chemistry, London 1986
- [12] CUADRA, CUADRA Directory of Online Databases, Host: DATASTAR, London 1988
- [13] H. ROHLEDER; M. MATTHIES; J. BENZ; R. BRÜGGEMANN; B. MÜNZER; R. TRENKLE; K. VOIGT: Umweltmodelle und rechnergestützte Entscheidungshilfen für die vergleichende Bewertung und Prioritätensetzung bei Umweltchemikalien. GSF-Bericht 42/86, Gesellschaft für Strahlen- und Umweltforschung mbH, Neuherberg 1987
- [14] Gesellschaft Deutscher Chemiker – Beratergremium für umweltrelevante Altstoffe: Stoffberichte über umweltrelevante Altstoffe (sämtlich erschienen bei VCH-Verlagsgesellschaft, Weinheim)
- [15] Books in Print, R. R. Bowker Company, New York & London 1987
- [16] Verlag der Buchhändler-Vereinigung GmbH, Verzeichnis lieferbarer Bücher, German Books in Print, Verlag der Buchhändler-Vereinigung GmbH, Frankfurt am Main 1987
- [17] Führer durch die technische Literatur, Katalog technischer Werke für Studium und Praxis, Universitätsbuchhandlung Lachner, München 1987