

INFORMATIONSSYSTEME UND IHRE ERSCHLIESSUNG

SCIFINDER SCHOLAR UND CROSSFIRE UND WEB OF SCIENCE UND ... - LUXUS ODER NOTWENDIGKEIT ?

ENGELBERT ZASS

ABSTRACT

Das Informationsangebot in der Chemie ist, im Unterschied etwa zur Biologie, fast völlig von kommerziellen Anbietern dominiert, deren elektronische Produkte vor allem einen zahlungskräftigen Industriemarkt bedienen. Das ist eine für Hochschulen problematische Situation, trotz zahlreicher „academic programs“ für solche Datenbanken. Die Produzenten haben in den letzten Jahren u.a. durch Erweiterung der inhaltlichen Erfassung, Erwerb weiterer Datenbanken und Optimierung ihrer Benutzeroberflächen für Endnutzer versucht, ihre Marktanteile zu vergrößern. Vor diesem Hintergrund stellt sich die Frage, ob nicht eine oder zwei dieser kostspieligen, umfangreichen Quellen für Endnutzer die wesentlichen Bedürfnisse im Bereich Sekundär/Tertiärliteratur abdecken kann. Aufgrund unserer langjährigen Rechercheerfahrung und von Datenbankvergleichen müssen wir diese Frage leider verneinen.

EINLEITUNG

Das wissenschaftlich-technische Informationsangebot, von Gratisquellen wie *Wikipedia* und *Google Scholar* bis zu den kommerziellen, kostenpflichtigen Datenbanken hat inzwischen einen Umfang angenommen, der selbst für Spezialisten nur noch schwer überschaubar ist. Vielfalt und Komplexität nehmen weiter zu, vor allem in der Chemie und verwandten Gebieten mit ihrem zahlungskräftigen und informationshungrigen Markt in der chemischen und pharmazeutischen Industrie. Dies ist ein Problem sowohl für Benutzer dieser Datenbanken als auch für Bibliothekare: letztere müssen in einer Zeit eingefrorener oder gekürzter Budgets eine den Bedürfnissen ihrer Klientel angemessene Auswahl bereitstellen (lizenzieren), und das nicht nur bei ständig neuen Quellen, sondern auch steigenden Preisen für bereits lizenzierte Datenbanken.

Um ein angemessenes Kosten/Nutzen-Verhältnis zu erreichen, sind Datenbanken aber nicht nur zu finanzieren und zu lizenzieren, sondern man muss auch entsprechende Benutzerunterstützung anbieten, d.h. Ausbildung, Schulung, Betreuung und Beratung zur Informationsbeschaffung und Datenbanknutzung. Die heutige Generation der Studierenden nutzt zur Informationsbeschaffung mit Selbstverständlichkeit das *World Wide Web*, Suchmaschinen wie *Google* und überall frei zugängliche Quellen wie *Wikipedia*. Eines der wesentlichen Probleme bei Studienanfängern ist daher die Vermittlung der Tatsache, dass dies im Studium und in der wissenschaftlichen Tätigkeit bei weitem nicht ausreicht, und dass alle Quellen nur mit entsprechender kritischer Bewertung zu nutzen sind. Weiter ist zu vermitteln, dass mit traditionellen Informationsquellen gleichmässige Qualität, dauernde Verfügbarkeit, und repräsentative (angemessene), nachvollziehbare Abdeckung von Themenbereichen eher gewährleistet sind. Mit einer solchen Erweiterung der Informationsbeschaffung ist das Problem aber nicht gelöst: Benutzer neigen nach unserer Beobachtung in einem problematischen Ausmass dazu, für ein Problem statt der fallweise jeweils geeignetsten Informationsquelle diejenige zu nehmen, welche sie am besten kennen, oder die am einfachsten zugänglich bzw. zu benutzen ist. Dafür verantwortlich sind sowohl die (abschreckende) Vielfalt des Datenbankangebots, als auch die Tatsache, dass in der Chemie grosse Datenbanken wie *Chemical Abstracts* (CA [1]) oder *Beilstein* [2] tatsächlich Antworten auf sehr viele Fragen bieten - aber eben nicht auf alle, und nicht immer die beste Antwort. Äusserungen von Benutzern wie z.B. «Warum *DiscoveryGate* testen, ich finde alles in *SciFinder Scholar*» erhellen die Problematik der Einführung neuer Informationsressourcen, selbst wenn Bibliothekare sie zuvor sorgfältig getestet und die lokalen Informationsbedürfnisse analysiert haben.

An der ETH Zürich haben wir uns immer schon bemüht, im Rahmen der Ziele dieser Hochschule mit den verfügbaren Erwerbungsmitteln eine möglichst optimale Informationsversorgung zu bieten. Für die vom Informationszentrum Chemie Biologie Pharmazie [3] abgedeckten Fachbereiche sind das zur Zeit über 40 kostenpflichtige Datenbanken wie *SciFinder Scholar* [4] (*Chemical Abstracts* [1]), *CrossFire* [5] *Beilstein* [2] und *Gmelin* [6], *Web of Knowledge* [7] (*Web of Science* [8]), und zahlreiche Spezialdatenbanken, z.B. *Bretherick's Reactive Chemical Hazards Database*, die elektronische Fassung eines wichtigen Handbuches mit Sicherheitsinformationen über chemische Reaktionen auf CD-ROM. Dieses breite Angebot für Endnutzer (Zugriff am Arbeitsplatz oder via WLAN in öffentlichen Räumen) wird ergänzt durch das Angebot vermittelter Recherchen durch Spezialisten in Datenbanken der Hosts *STN International* [9] und *Dialog* [10].

DATENBANKANGEBOT UND BEWERTUNG

Im Folgenden wird skizziert, aufgrund welcher Argumente und Kriterien wir ein solches, in der Lizenzierung kostspieliges und in der Betreuung aufwändiges Angebot aufrechterhalten und weiterentwickeln. Unsere Diskussion beschränkt sich dabei auf Datenbanken aus dem Bereich der Sekundärquellen, d.h. Information des Typs, der gedruckt in Handbüchern und Referateorganen angeboten wurde, unter Ausschluss von elektronischen Zeitschriften und Monographien.

Sorgfältige Evaluierung führte dazu, dass wir aufgrund von Bedarfsanalysen, Kosten/Nutzen-Abschätzungen einschliesslich Berücksichtigung alternativer elektronischer Quellen z.B. neben der gedruckten Ausgabe des *Bretherick* auch die erwähnte Datenbank anbieten, während wir uns etwa beim *Merck Index* [11] oder beim *Kleemann-Engel* [12], die beide auch elektronisch verfügbar sind, auf die gedruckte Version beschränken.

Zum Datenbankangebot gehört nicht nur die Lizenzierung, sondern auch technische Unterstützung, Propagierung auf Webseiten und im WebOPAC, denn bei einem umfangreichen Angebot ist die Information darüber an Benutzer eine zentraler Aufgabe. Während dies bei Monographien und Periodika unabhängig vom Medium (Papier oder elektronisch) durch Nachweis in unserem WebOPAC CLICAPS [13] geschieht, pflegen wir für Datenbanken eine «meta-Datenbank» [14], in der «Google-like» nach an der ETH Zürich verfügbaren Datenbanken zu Chemie, Biologie, Pharmazie gesucht werden kann. Aufgeführt sind sowohl kostenpflichtige, lizenzierte als auch gratis im Web verfügbare Quellen. Ausserdem kann man sich zu wichtigen Themenbereichen wie z.B. Kataloge, Patente, Toxikologie die jeweils verfügbaren Datenbanken anzeigen lassen. Hier wird auch organisatorisch-technische Information, Zugriffsberechtigungen, oder Installationsanleitungen für ggf. benötigte spezielle Client-Software bereitgestellt; für via Web-Browser abfragbare Datenbanken ist natürlich jeweils der direkte Link dazu verfügbar.

Die enorme Vielfalt der Quellen wird bei Datenbanken noch durch die Vielfalt der Benutzeroberflächen vergrössert, denn praktisch alle wichtigen Datenbanken sind unter mehreren Benutzeroberflächen verfügbar, im Falle von *Chemical Abstracts* - einem zugegebenermassen extremen Beispiel - sind es derzeit nicht weniger als neun: *SciFinder* [15], *SciFinder Scholar* [4], *CA on CD* [16], *CA Student Edition* [17], *STN Messenger* [18], *STN on the Web* [19], *STN Easy* [20] sowie die Retrievalsprachen der Hosts *Dialog* [10] und *DataStar* [21]. Typischerweise bedeuten unterschiedliche Benutzeroberflächen auch unterschiedliche Zugriffsmöglichkeiten zur Information, dahinter können sich aber auch bezüglich Inhalt oder zeitlicher Abdeckung

unterschiedliche Versionen einer Datenbank verbergen, z.B. *CAplus* [22] bei *SciFinder Scholar*, wahlweise *CA* [23] oder *CAplus* [24] bei STN International [9]; *Science Citation Index* [25] oder *Science Citation Index Expanded* [26], und oft auch unterschiedliche Preismodelle wie ein jährlicher Festpreis für *Web of Science* gegenüber «pay-per-use» für den *Science Citation Index* bei den Hosts STN [9], Dialog [10], DataStar [21].

Für wichtige Quellen kann es erforderlich sein, mehr als eine Benutzeroberfläche (Datenbankversion) vorhalten zu müssen - in der Chemie ist *Chemical Abstracts* [1] zweifellos die umfassendste und bedeutendste einzelne Informationsquelle, und *SciFinder Scholar* [4] ist die Benutzeroberfläche der Wahl für Endnutzerrecherchen im Hochschulbereich. Wir bieten aber als Alternative trotz der zusätzlichen Kosten auch noch vermittelte Recherchen in *Chemical Abstracts* via STN [9] an. Dies wegen Systemgrenzen bei der Substrukturrecherche in *SciFinder Scholar* («AutoFix»), und weil einige wichtige Typen von Fragestellungen - in der Biochemie Sequenzen und Subsequenzen von Biopolymeren, die Zusammensetzung anorganischer Verbindungen und Werkstoffe («alle Ti-N-Verbindungen ohne weitere Elemente»), Suche nach «therapeutic use» von Verbindungen, usw. - in *SciFinder Scholar* im Unterschied zu den *Chemical Abstracts* Datenbanken bei STN nicht suchbar sind. Bei thematischen Recherchen in *Chemical Abstracts* genügt *SciFinder Scholar* dann nicht, wenn diese Recherchen komplex sind, oder wenn hohe Präzision bei möglichst umfassender Abdeckung gefordert ist: eine Suche nach «nuclear overhauser effect difference spectroscopy» gab 56 Literaturzitate in *SciFinder Scholar* gegenüber 859 in STN. Solche Recherchen gehören in die Hand von Spezialisten, und dafür wurde das Endnutzersystem *SciFinder Scholar* auch nicht entwickelt.

Auch wenn sie als solche angepriesen werden, sind die derzeitigen Benutzeroberflächen nicht benutzerfreundlich, sie sind lediglich bedienungsfreundlich. Das ist ein Unterschied, der leider nicht nur von den Produzenten, sondern auch in Anwenderkreisen zu wenig differenziert wird. Bedienungsfreundliche Benutzeroberflächen wie etwa *SciFinder Scholar* oder *Web of Knowledge* erwecken leider beim Benutzer den Eindruck, alles sei so einfach. Dieser Eindruck wird durch das Marketing der Produzenten noch verstärkt. Dabei müssten diese am besten wissen, wie gefährlich z.B. eine einfach wirkende Benutzeroberfläche über einer so komplexen Datenbank wie *Chemical Abstracts* sein kann - diese Diskrepanz ist andererseits eine überzeugende *raison d'être* für Ausbildung, Schulung und Betreuung von Benutzern durch Bibliothekare und Informationsspezialisten! Nur so sind Chemiker in der Lage, zu entscheiden, ob sie ein Problem selber am Arbeitsplatz lösen können, oder ob sie die Hilfe eines Spezialisten, und damit ggf. auch andere Benutzeroberflächen oder Datenbanken als die gewohnten benötigen.

ERFASSUNG UND ERSCHLIESSUNG IN DATENBANKEN

Wichtige Kriterien für die Auswahl von Datenbanken und Benutzeroberflächen sind also Suchbarkeit (Zugriff), Bedienungsfreundlichkeit und Preismodelle; Festpreise etwa lohnen sich nur für intensiv genutzte Datenbanken. Noch wichtiger ist natürlich der Inhalt der Datenbanken - die beste Oberfläche nützt ja nichts, wenn die Datenbank die gewünschte Information nicht, nicht mehr oder noch nicht enthält, denn Erfassung und Erschliessung der Primärliteratur haben sich im Laufe der Zeit bei den meisten Datenbanken wesentlich geändert. Die entsprechende meta-Information ist viel zu wenig bekannt, und sie wird von den Produzenten allzu oft gar nicht zur Verfügung gestellt.

Wichtige Kriterien für die Bewertung von Datenbanken sind:

- Erfassungskriterien (Auswahl der einzelnen Publikationen)
 - o Typen der Quellen (Primärliteratur: Zeitschriften, Patente, usw.)
 - o Zahl der Titel (Zeitschriften, Patentbehörden, usw.)
 - o Zeitrahmen der Erfassung
- Erschliessungstiefe (Indexierung)
- Aktualität

Die aufgrund dieses Bewertungsrasters erhobenen qualitativen und vor allem die quantitativen Daten erfordern aber einen sorgfältigen Umgang. Im Vergleich von *Chemical Abstracts* [1] und *Beilstein* [2] ist die Tatsache, dass letzterer Patente nur bis 1980 erfasst hat, während *CA* dies von Anfang an getan hat (allerdings mit grossen Unterschieden in der Auswahl der Patentdokumente der einzelnen Länder, die im Laufe der Zeit immer wieder erweitert wurde) ein klares qualitatives Kriterium. Ein simpler quantitativer Vergleich der erfassten Zeitschriftentitel hingegen - *CA* ca. 9500, *Beilstein* z.Z. 174 - lässt nicht nur ausser Acht, dass der *Beilstein* vor 1980 mehr Zeitschriften erfasste, sondern diese Zahlen sind so gar nicht vergleichbar: betrachtet man bei *CA* nur die organischen Sektionen (ohne organometallische Chemie, d.h. den vom *Beilstein* erfassten Bereich der Chemie), dann findet man für das Jahr 2000 26151 Artikel aus insgesamt nur 1173 verschiedenen Zeitschriften; davon wurde aber aus nicht weniger als 387 Zeitschriften nur je ein Artikel für die Erfassung ausgewählt, während die 150 wichtigsten Zeitschriften 81 % aller Artikel lieferten. *Chemical Abstracts* ist die umfassendste Quelle für die Chemie, die Abdeckung des *Beilstein* für den Bereich der organischen Chemie ist aber durchaus vergleichbar. Diese Vorsicht gilt auch für einzelne vergleichende Recherchen: die detaillierte Analyse der Resultate einer Literaturrecherche zur Isolierung von Steroiden aus der chinesischen (!) Medizinalpflanze *Artemisia annua* gab zwei Literaturzitate im *CrossFire Beilstein* [2] gegenüber 10 in *Chemical Abstracts* [1]. Bei den neun exklusiv in *CA* gefundenen

Artikeln handelt es sich, mit Ausnahme der in Irland erscheinenden Zeitschrift *Plant Science*, um Artikel chinesischer Zeitschriften, also in einer Sprache, mit der ein durchschnittlicher Benutzer an einer Hochschule kaum etwas anfangen kann. Je ein Artikel aus der Zeitschrift *Phytochemistry* wurde aber exklusiv nur in *CA* bzw. *Beilstein* gefunden, obwohl die betreffenden beiden Artikel von beiden Datenbanken erfasst wurden - der Unterschied im Resultat ist hier durch die unterschiedliche Erschliessung (Indexierung) bedingt. Alle gefundenen Artikel ausser den chinesischen sind auch im *Science Citation Index* [25] erfasst, dort aber mit unserer Fragestellung nach Verbindungsklassen (Steroide) nicht auffindbar, denn es gibt in dieser Datenbank weder die Möglichkeit der (Sub)struktursuche (wie in *Beilstein* und *CA*), noch eine Indexierung von Verbindungsklassen mit Stichworten.

Dieses Beispiel illustriert, dass eben nicht nur die Erfassung der Primärquellen, sondern auch ihre Erschliessung in der Datenbank ein wichtiges Kriterium für die Auswahl von Datenbanken ist. Diese Bewertung muss sowohl aufgrund der Erfassungs- und Erschliessungskriterien, als auch aufgrund von Vergleichsrecherchen erfolgen [27].

Eine Suche nach Literatur zum Heterozyklus 6,7-Dimethoxy-1,2,3,4-tetrahydroisochinolin, insbesondere zu seiner Herstellung, gab in den Datenbanken von *Chemical Abstracts* [1] bzw. *Beilstein* [2] folgende Resultate:

	CAplus [22]	Beilstein [2]
Literatur total	189	121
Literatur zur Herstellung	26	36

Von den Literaturzitatzen zur Herstellung kommen 17 sowohl in *Chemical Abstracts* als auch im *Beilstein* vor. Von den 9 Zitaten exklusiv in *CAplus* war eines aus dem Jahr 2004 noch zu neu - Handbücher hatten wegen ihrer aufwändigen Erfassung traditionell einen grösseren Zeitrückstand auf die Originalliteratur als Referateorgane wie *Chemical Abstracts*, und das hat sich auf die entsprechenden Datenbanken übertragen: wichtige CAS-Datenbanken werden täglich aktualisiert, der *Beilstein* nur viermal pro Jahr. Vier Artikel stammen aus den Zeitschriften *Acta Poloniae Pharmaceutica*, *Journal of Heterocyclic Chemistry* und *Phytochemistry*, die vom *Beilstein* an sich erfasst werden, in diesem Fall aber offensichtlich nicht. Die übrigen stammen aus Informationsquellen, die im *Beilstein* nicht erfasst werden: ein japanisches Patent, *Journal of Chromatography*, *B: Biomedical Applications*, *Dokladi na Bulgarskata Akademiya na Naukite* (bulgarisch) und *Szegedi Tanarkepzo Foiskola Tudomanyos Kozlemenyei* (ungarisch). Die 19 Publikationen zur Herstellung exklusiv nur im *Beilstein* stammen alle aus Zeitschriften, die an sich für *CA* ausgewertet werden, und aus dem von *CA* abgedeckten Zeitraum. Auch hier liegen die Unterschiede wieder in der nach unterschiedlichen Kriterien

und Richtlinien erfolgenden inhaltlichen Erschliessung der Primärliteratur. Aus diesem und vielen ähnlichen von uns analysierten Beispielen geht klar hervor, dass für eine einigermaßen umfassende Literaturrecherche zur Herstellung organischer Verbindungen sowohl *Beilstein* als auch *Chemical Abstracts* erforderlich sind. Die Unterschiede zwischen *CA* und *Beilstein* sind oft noch ausgeprägter für die Zeit vor 1967, in der die CAS-Datenbanken erst durch nachträgliche Erfassung und maschinelle Informationsverarbeitung der gedruckten *Chemical Abstracts* erweitert worden sind; dies gilt zwar auch für die *Beilstein*-Datenbank, nur war hier die Erfassung aus dem Handbuch wegen der andersartigen Datenstruktur einfacher, und die Erschliessung von Verbindungen und Reaktionen war im *Beilstein*-Handbuch derjenigen in *CA* damals oft überlegen: aus der Publikation zur ersten Totalsynthese des Steroidhormons Östron (1942) wurden von den insgesamt 60 organischen und anorganischen Verbindungen und 38 Reaktionen in diesem Zeitschriftenartikel im *Beilstein* 35 Verbindungen und 36 Reaktionen erfasst, in *CA* hingegen nur 15 Verbindungen und 2 Reaktionen.

Gravierende Unterschiede in wichtigen Datenbanken existieren auch bei der scheinbar so einfachen Autorensuche: Der *Beilstein* als Verbindungsdatenbank reflektiert weiterhin Zielsetzung und Datenstruktur des gedruckten Handbuchs, aus dem die Datenbank entwickelt wurde. Autorennamen waren da nicht wichtig, und es wurden daher bis 1979 nur maximal zwei Autoren erfasst, ab 1980 dann maximal 6, jeweils mit et al., wenn die Originalpublikation mehr Autoren hatte. In Literaturdatenbanken wie dem *Science Citation Index* oder *Chemical Abstracts* sind Autorennamen ein wichtiger Teil der Datenbank; dennoch hat *CA* bis 1996 die Erfassung auf maximal 10 Autoren beschränkt, erst seit 1997 werden mehr Autoren erfasst: nach unseren Tests nicht alle, aber mindestens 150 - leider fehlt auch hier wieder einmal eine klare Angabe des Datenbankproduzenten. Im *Science Citation Index* hingegen sind alle Autoren und alle Adressen erfasst. Adressen fehlen im *Beilstein* völlig, und in *CA* ist immer noch nur die Adresse des ersten Autors angegeben; das ist aber oft nicht der Hauptautor oder der in der Publikation als Korrespondenzautor angegebene Ansprechpartner.

Diese Beispiele zur unterschiedlichen Erfassung und Erschliessung der Primärliteratur in wichtigen Datenbanken zeigen auch, dass für zuverlässige Rechercheresultate entsprechende Kenntnisse über die Datenbankinhalte (Erfassungsrahmen, Erschliessungskriterien) unverzichtbar sind. Diese Kenntnisse können aber derzeit weder aus den «help messages» noch aus der Dokumentation (Manuals, Webseiten) der Datenbankproduzenten in hinreichendem Masse gewonnen werden, gerade hier ist Schulung und Betreuung durch versierte Informationsspezialisten unverzichtbar - leider ist dies aber bei vielen Benutzern und Bibliotheken noch nicht ausreichend angekommen!

EINSATZ DER DATENBANKEN

Für Recherchen in der Chemie existierte lange eine relativ klare Aufgabenteilung für häufige Fragestellungen in den wichtigsten grossen Datenbanken, z.B. für:

- Autorenrecherchen: primär *Chemical Abstracts* [1] (nicht alle Autoren, aber mehr Chemie-Zeitschriften), sekundär *Science Citation Index* [25]
- Verbindungsrecherchen: *Beilstein* [2] (organische Verbindungen) / *Gmelin* [6] (anorganische und metallorganische Verbindungen) und *Chemical Abstracts*
- Eigenschaften und Daten von Verbindungen: *Beilstein/Gmelin*
- Zitationsrecherchen: *Science Citation Index*

Für spezielle Fragestellungen nach Reaktionen, Patenten, Spektren, usw. benutzte (und benutzt) man neben diesen «zentralen» Datenbanken entsprechende Spezialdatenbanken.

Diese relativ klare Aufgabenteilung und Rangordnung wurde durch eine ganze Reihe neuerer Entwicklungen verwischt und kompliziert. Im Kampf um grössere Anteile im attraktiven, zahlungskräftigen Markt der Chemieinformation begannen die Datenbankproduzenten den von den gedruckten Quellen her überlieferten Rahmen zu sprengen und gewissermassen in «fremden Revieren zu wildern»: ab 1985 produzierte der *Chemical Abstracts Service* eine eigene Reaktionsdatenbank CASREACT [28], *Beilstein* erfasste ab 1980 auch Abstracts und Titel der Publikationen, aus denen bis dahin nur Verbindungen und deren physikalische und chemische Eigenschaften (einschliesslich Reaktionen) erfasst wurden; *CA* brach mit den ab 1999 (zurück bis 1997) erfassten Zitationen das quasi-Monopol des *Science Citation Index* für Zitationsrecherchen, der *Science Citation Index* sah sich auf der anderen Seite genötigt, auch Literatur vor 1945 zurück bis 1900 zu erfassen (Projekt *Century of Science* [29]). *Chemical Abstracts* stärkte seine Position im wichtigen Feld der Eigenschaftssuche durch die umfassende Inkorporierung von berechneten (2001) und ab 2003 auch gemessenen Eigenschaften von Verbindungen, und bietet damit eine direkte Konkurrenz zu *Beilstein* und *Gmelin*, welche bis dahin diesen Bereich dominierten. Auch das bedarf einer genaueren Analyse: für die anorganische, auch technisch bedeutende Verbindung Siliziumtetrafluorid etwa fand man in *Chemical Abstracts* (Verbindungsdatenbank *CAS Registry* [30]) 10 gemessene physikalische Eigenschaften, in *CrossFire Gmelin* [6] dagegen nicht weniger als 64. Für das Mykotoxin Deoxynivalenol z.B. bot der *Beilstein* 6 physikalische und 26 biologische/pharmakologische Eigenschaften, gegenüber 7 bzw. 1 in *Chemical Abstracts*. Literatur zur Toxikologie dieser problematischen Verunreinigung in Getreideprodukten fanden wir wie folgt: *Chemical Abstracts*, 262 Zitate; *CrossFire Beilstein*, 27; *MDL Toxicity Database* [31], 34; dabei fand man unter den 27 bzw. 34 Zitaten in *Beilstein* bzw.

Toxicity Database lediglich vier Zitate in beiden Datenbanken. Dies zeigt wieder den umfassenden Charakter von *CA*, mancher Praktiker wird aber die beiden anderen Quellen vorziehen, denn weniger ist of mehr bei zu lesender Literatur.

SCHLUSS: EIGENSCHAFTEN-QUELLEN-INDEX

Die Beispiele sollen veranschaulichen, warum wir an der ETH Zürich im Bereich Chemie/Biologie/Pharmazie ein derart grosses und aufwändiges Angebot an Datenbanken vorhalten und in ständigem Kontakt mit den Benutzern und Produzenten weiter pflegen und optimieren. In diesem Prozess ist es auch erforderlich, Monopolisierungsansprüchen der grossen Anbieter («bei uns finden Sie alles was Sie brauchen») entgegenzutreten, indem eine vernünftige Vielfalt an Quellen lizenziert und vor allem auch unterstützt wird. Bei einem solchen grossen Datenbankangebot genügt es nämlich nicht, den Benutzern eine Übersicht über die lokal verfügbaren (lizenzieren) Datenbanken zu bieten, wie wir das mit unserer bereits erwähnten meta-Datenbank [14] tun, es ist etwa im komplexen Bereich der chemischen Verbindungen, für die mehrere hundert physikalische und chemische Eigenschaften existieren können, eine entsprechende Benutzerunterstützung zum gezielten Suchen nach solchen Eigenschaften zu bieten. Dabei sollten nicht nur die bekannten Datenbanken, sondern auch für den Benutzer weniger offensichtliche Werke berücksichtigt werden. Für diese Thematik wurde von M. Brändle (InfoZentrum) ein dreisprachiger Eigenschaften-Quellen-Index (EQI) geplant und unter Mitarbeit von Jana Sonnenstuhl (HU Berlin), Arun Kumar (InfoZentrum), Francine Dreier, Cédric Noir, Rachel Bays (alle Univ. Genève, französische Fassung) realisiert. Der seit April 2007 öffentlich zugängliche EQI [32] verknüpft derzeit insgesamt 1.042 verschiedene Eigenschaften mit den jeweiligen Quellen (821, elektronisch oder gedruckt) und chemischen Systemen, wobei die Quellen nach gemessenen bzw. berechneten Daten, Definition, Messmethode und Vorhersage zu den Eigenschaften differenziert sind. Ein zunehmender Anteil der Quellen ist hinsichtlich ihrer Nützlichkeit (Qualität, Quantität und Auffindbarkeit der Daten, Zugänglichkeit und Bedienungsfreundlichkeit der Quellen) bewertet. Zu den Eigenschaften sind nicht nur die korrekten Einheiten und Symbole sowie Synonyme und Übersetzungen angegeben, es wird auch ein Link auf die entsprechende Beschreibung einer Eigenschaft im *Römpp Online* [33] angeboten. Gegenwärtig führt der EQI die an der ETH Zürich verfügbaren und die frei zugänglichen Quellen auf. Beabsichtigt ist, den EQI mit Reference Linking zu erweitern, damit weitere Bibliotheken und Verbände mitwirken können und deren lokaler Bestand nachgewiesen wird. Um die Sichtbarkeit des EQI im WWW zu erhöhen, haben wir die Eigenschaften und verknüpften Quellenangaben durch Google indizieren lassen. Der EQI wurde als System polyhierarchischer Thesauri realisiert, die für die Sprachen Deutsch und Englisch sowie Französisch ausgelegt und auch für zusätzliche Sprachen erweiterbar sind. Mit

dem EQI manifestieren wir unsere Politik, Datenbanken nicht nur sorgfältig vor und nach Lizenzierung zu evaluieren, sondern im Hinblick auf die knappen Mittel mit einer ganzen Palette von Massnahmen - von elektronischen Hilfsmittel wie WebOPAC [13] und EQI [33] bis hin zu Kursen und elektronischem Unterrichtsmaterial [34] - dafür Sorge zu tragen, dass sie angemessen genutzt werden.

Der Autor dankt dem Team des Informationszentrums Chemie Biologie Pharmazie für die Unterstützung, insbesondere Dr. M. Brändle für Ergänzungen zum Manuskript.

ANMERKUNGEN

(Links zuletzt geprüft am 16.07.2008)

- 1 CAS (Chemical Abstracts Service) Chemical Abstracts (CA): <http://www.cas.org/expertise/cascontent/index.html>; unter «Chemical Abstracts» verstehen wir hier der Einfachheit halber eine oder mehrere dieser Datenbanken, wenn nichts anderes erwähnt ist.
- 2 Beilstein-Datenbank: <http://www.crossfirebeilstein.com/> (CrossFire); <http://www.stn-international.de/stndatabases/databases/beilstei.html> (STN).
- 3 Informationszentrum Chemie Biologie Pharmazie (ETH Zürich): <http://www.infochembio.ethz.ch>.
- 4 CAS (Chemical Abstracts Service) SciFinder Scholar: <http://www.cas.org/products/sfacad/index.html>.
- 5 CrossFire: <http://www.beilstein.com/>; <http://www.infochembio.ethz.ch/Xfire.html>.
- 6 Gmelin-Datenbank: <http://www.gdch.de/taetigkeiten/gmelin.htm>; <http://info.crossfiregmelin.com/>.
- 7 Thomson Reuters Web of Knowledge: <http://www.isiwebofknowledge.com/>.
- 8 Thomson Reuters Web of Science: <http://scientific.thomsonreuters.com/products/wos/>.
- 9 Host STN International: <http://www.stn-international.de/>.
- 10 Host Dialog: <http://www.dialog.com/>.
- 11 Merck Index: <http://www.merckbooks.com/mindex/>.
- 12 Kleemann-Engel, Pharmaceutical Substances: <http://www.thieme.com/SID2394652840960/productssubpages/pubid-368040547.html>.
- 13 CLICAPS (Chemistry Library Information Control and Presentation System): integriertes Bibliothekssystem zur EDV-Unterstützung aller Arbeitsprozesse und der Informationsvermittlung an die Benutzer; Eigenentwicklung in der ETH Chemiebibliothek (jetzt Informationszentrum [3]) auf der Basis der kommerziellen Datenbanksoftware FileMaker Pro; öffentlicher WebOPAC: <http://www.clicaps.ethz.ch/>.
- 14 Datenbanken zur Chemie/Biologie/Pharmazie an der ETH Zürich ("meta-Datenbank"): <http://www.infochembio.ethz.ch/db.html>.

- 15 CAS SciFinder: <http://www.cas.org/products/scifindr/index.html>.
- 16 CA on CD wurde für 2008 noch angeboten, erscheint aber nicht mehr auf den CAS-Webseiten.
- 17 CA Student Edition: <http://www.cas.org/support/academic/pricing.html>.
- 18 STN Messenger (Retrievalsprache des Hosts STN International [9]):
http://www.stn-international.de/training_center/rl/retrieval_short.pdf;
http://www.stn-international.de/training_center/mat_sea_stn.html.
- 19 STN on the Web (Web-Oberfläche für STN Messenger [18]):
<http://www.cas.org/products/stnweb/index.html>.
- 20 STN Easy (Web-Interface für Datenbanken beim Host STN International [9]):
<http://stneasy.cas.org/easy5/index.html>.
- 21 Host DataStar: <http://www.dialog.com/products/datastar/>.
- 22 CAplus: <http://www.cas.org/expertise/cascontent/caplus/index.html>.
- 23 CA (STN): <http://www.stn-international.de/stndatabases/databases/ca.html>.
- 24 CAplus (STN): <http://www.stn-international.de/stndatabases/databases/caplus.html>.
- 25 Science Citation Index: <http://scientific.thomson.com/products/sci/>.
- 26 Science Citation Index Expanded: <http://scientific.thomson.com/products/scie/>.
- 27 Nach unserer Erfahrung ist die veröffentlichte Information der Produzenten für eine angemessene Beurteilung von Datenbanken völlig unzureichend, man benötigt persönliche Kontakte zu den Produzenten, Erfahrungsaustausch mit Kollegen (im Falle der Chemie auch aus der Industrie!), und muss Zeit in aufwändige Vergleichsrecherchen oder beta-Tests neuer Informationsprodukte investieren.
- 28 CASREACT: <http://www.cas.org/expertise/cascontent/casreact.html>.
- 29 Projekt Century of Science: http://webofknowledge.com/currentuser_wokhome/backfiles/centsci/.
- 30 CAS Registry: <http://www.cas.org/expertise/cascontent/registry/index.html>.
- 31 MDL Toxicity Database: <http://www.mdl.com/products/predictive/toxicity/index.jsp>.
- 32 EQI: <http://www.eqi.ethz.ch/>.
- 33 Römpp Online: <http://www.roempp.com/index.shtml>.
- 34 Kurse des InfoZentrums: <http://www.infochembio.ethz.ch/kurse.html>.

ADRESSE DES AUTORS

Dr. Engelbert Zass

ETH Zürich

Informationszentrum Chemie Biologie Pharmazie

Wolfgang-Pauli-Str. 10, CH-8093 Zürich

E-Mail: zass@chem.ethz.ch

<http://www.infochembio.ethz.ch>