Comparison of some major information resources in pharmaceutical competitor tracking

Alexander Mullen, Martin Blunck and Klaus Eike Möller

Computerized information sources for competitor tracking provide invaluable intelligence resources in the environment of corporate R&D. The authors discuss needs and resources for tracking competitor development products, and compare the scope, strengths and weaknesses of some of the major commercially available databases. The vision of the ideal competitor information file is outlined together with the current, unfulfilled needs of industrial users.

Pharmaceutical R&D is a knowledge-based, high-tech activity that thrives on innovation and has a great thirst for current, value-added, reliable information. Different types of information, all properly packed and designed for the scientific 'consumer', are required at the various stages in the discovery process and development pipeline. According to a recent report, the average time period required to develop a new drug is 12 years and the development costs are about $330 million[1].

In this review, attention is focused on some selected computerized sources for competitor tracking that are currently used within the corporate environment. A paper discussing concepts behind competitor information systems of this type is due for publication shortly[2].

Because of the pressures on a development product during its passage through the R&D pipeline, scientists and management are constantly challenged to reassess their projects. Can the investment still be justified? What is the likelihood of success and market potential? How strong is the competition? Commercial database producers and publishers have made great strides to service this need for scientific and commercial information on competitor development products for in-house use within the pharmaceutical industry[3,4].

Databases are supplied for in-house use in various formats and on different media (for example, diskettes, CD-ROMs, tapes and electronic files) and the updates are issued at a range of intervals, from daily to monthly[2]. Most larger pharmaceutical companies, such as those represented by the 29 corporate members of the Pharma Documentation Ring (PDR)[4–7], operate such databases on the corporate computer network, thereby making the data available to R&D and marketing personnel as well as to staff in other divisions, such as strategic planning and licensing functions. The current trend[4] is to make these resources available via internal corporate internets – intranets – thereby providing easy-to-use browser front-ends.

Around 1990, a mere handful of commercial databases on competitor development products were available for in-house use[8], but this has since grown to more than 15 discrete files. Some of these files on competitor development and launched products are represented as 'fruits' on the

Alexander Mullen*, **Martin Blunck** and **Klaus Eike Möller**, Bayer AG, Pharmaceutical Research Centre, 42096 Wuppertal, Germany. tel: +49 202 368495, fax: +49 202 364200, e-mail: DEBAY679@IBMMAIL.COM

 1359-6446/97/$17.00. PII: S1359-6446(97)01002-7

competitor information 'tree of knowledge'[6,7] shown in Figure 1. For a more detailed discussion of the content of such files, see Ref. 2.

Information content of the 'ideal' file

The 'ideal' file on competitor development products should contain key information, presented in a comprehensive, coherently indexed manner (Box 1). No single file of those currently available covers all of the requirements and thus 'serious' industrial users are obliged to employ several of the files as well as consult the original literature for competitor tracking purposes. In fact, even the degree of overlap between the different files, whether based on discrete chemical structures or CAS registry numbers (see below) has been found to be relatively low[2] (CAS, 2540 Olentangy River Road, PO Box 3012, Columbus, OH 43210-0012, USA). No file provides either comprehensive literature coverage or adequate incidence and prevalence data, or even extensive turnover projections.

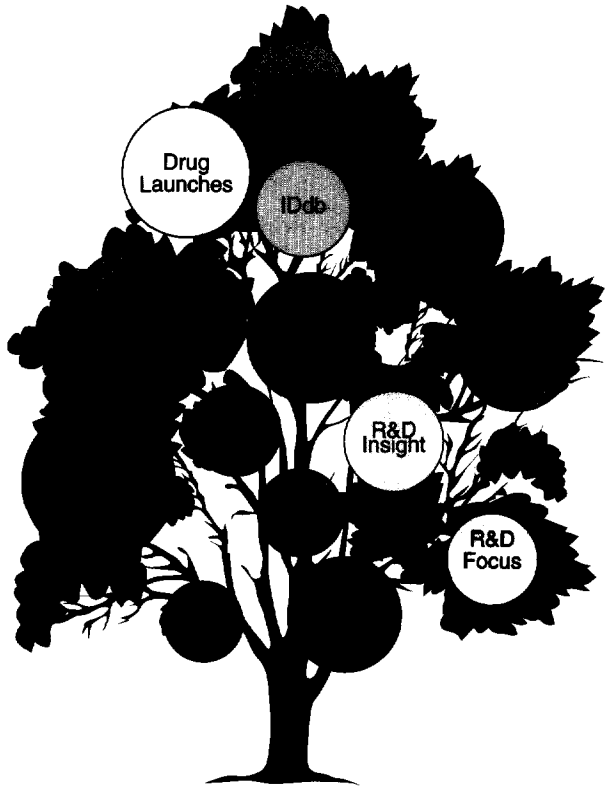

Figure 1. *'Knowledge tree' of drug information files on development and launched products.*

Box 1. Information content of the 'ideal' record from a file on competitor development products

- Chemical structure (substructure searching capability necessary)
- Compound names [test compound number, chemical nomenclature, trivial names, international non-proprietary names (INNs), trade names, etc.]
- CAS registry number (to facilitate the unequivocal identification of a substance in its different guises)
- Originating company/organization (allocated to a corporate group) and its location
- Licensees (comprehensive coverage)
- Therapeutic uses (in a standardized manner)
- Mechanisms of action (in a standardized manner)
- Development status in different countries in relation to developer (company/organization), therapeutic use and mechanism of action
- Detailed summary of scientific and marketing aspects of substance
- Likely synthetic route as revealed in (process) patent literature
- Patent number, priority date, patent expiry date in major countries
- Incidence and prevalence data of disease(s) to be treated by compound
- Scientific appraisal of compound: level of innovation, how does it compare with other compounds of its type with regard to activity, etc.
- Commercial potential (turnover estimates), i.e. 'blockbuster rating'
- Comprehensive list of current literature references [including (preferably upgraded) abstracts plus electronic full text copies plus images of major publications]

Relative content of different files

To give an impression of the type of data on competitor development products provided by these files, a recent introduction – the *R&D Insight* CD-ROM database from ADIS International (ADIS International Ltd, Chowley Oak Lane, Tattenhall, Chester, UK CH3 9GA) – will be discussed in the next section and comparison made with the more established files *Pharmaprojects* (PJB Publications, 18–20 Hill Rise, Richmond, Surrey, UK TW10 6UA) and *R&D Focus* (IMS Global Services, 7 Harewood Avenue, London, UK NW1 6JB). The three files are available online via commercial host computers and are also available for in-house use in CD-ROM versions. Tables 1 and 2 provide a comparison of the main features of the three selected files.

Table 1. Quantitative aspects of *Pharmaprojects*, *R&D Focus* and *R&D Insight* files[a]

	Pharmaprojects	*R&D Focus*	*R&D Insight*
Total records	19,300	9,300	6,100
Substances in 'active' phase[b]	5,700	7,200	4,500
Substances in 'no development reported/suspended' phases	5,700	200	400
Discontinued projects (substances) or withdrawn products	6,900	1,900	1,200
Established launched products	1,000	–	–
Updating frequency of CD-ROM	Monthly	Monthly	Monthly
Approximate cost of CD-ROM for single user	£3,500	£3,650[c]	£5,300
Producer	PJB Publications, 18–20 Hill Rise, Richmond, Surrey, UK TW10 6UA	IMS Global Services, 7 Harewood Avenue, London, UK NW1 6JB	ADIS International, 41 Centorian Drive, Mairangi Bay, 10, Auckland, NZ

[a]Values shown as rounded figures; status as at July 1996.
[b]Preclinical phase – recent launch.
[c]Basic price covers 1–5 networked users.

Qualitative aspects are inherently difficult to assess; however, Table 2 is an attempt to provide a general impression of the strengths of the different sources and stresses, once again, the need to use multiple files.

Pharmaprojects *file*

This database produced by PJB has been available since the early 1980s in hard copy form and later became available as an online file. In terms of coverage and currency – more than 19,300 substance-based records are present in the database – it is a highly regarded information source. The file is updated and the development status of the products is revised probably more frequently than is the case for the other sources. Of 19,300 records, about 5,700 compounds are in the 'active' development phase, about 6,900 records related to 'suspended' or 'no development reported' projects and the remainder represent mainly discontinued projects (see Table 1). The CD-ROM includes chemical structures and the main fields are presented in Table 2 together with those of *R&D Focus* and the *R&D Insight* files. At the beginning of 1996, some 18 PDR companies[4,6,7] (of 27 possible) had installed this file in-house.

Table 2. Qualitative comparison of field contents in *Pharmaprojects*, *R&D Focus* and *R&D Insight* files (rating scale range: from 0 to ++++).

	Pharmaprojects	*R&D Focus*	*R&D Insight*
Substance name	++++	+++	+++
CAS registry number	++	+	++
Therapeutic use	++++	++	+++
Mechanism of action	+++	++	++
Development status	++++	++	+++
Company	+++	++	++
Licensee	+++	++	++
Patent information	++	++(+)	0
Substance profile	+(+)	++	++++
Commercial information (potential turnover, etc.)	+	++	+++
Comprehensive list of relevant literature references	0	0	++
(Upgraded) summaries of main literature references	0	(+)	++
History of development	++	+++	+
Currency	++	+	+

The proportion of compounds undergoing active development corresponds, in general terms, to the distribution for *R&D Insight*, as presented in Figure 2.

Pharmaprojects is characterized by the systematic consistent manner in which the data are maintained and presented[2] (see Table 2). For example, companies are

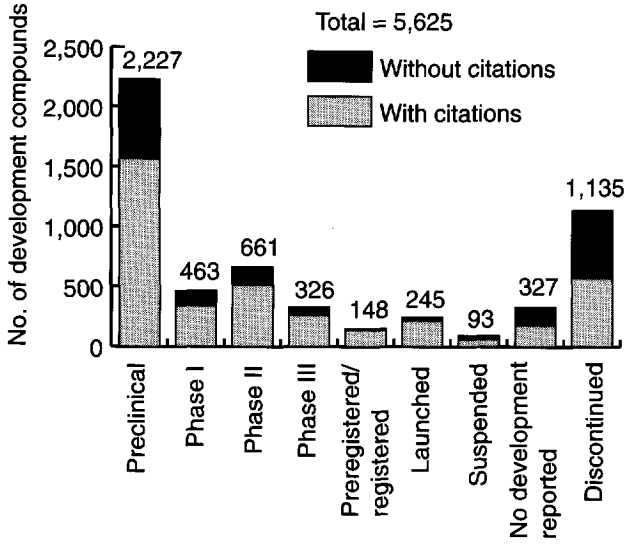

Figure 2. R&D Insight – *distribution of development compounds according to phase.*

allocated to corporate groups, and there is a clear differentiation made between mechanism-of-action and therapeutic-use terminology. On the other hand, there are few literature citations, and the descriptive summary is relatively brief. However, it is currently the only file that links company, therapeutic use, biological activity and development phase.

R&D Focus *file*

The *R&D Focus* file, produced by IMS, supplies reports on the latest scientific and commercial developments in international pharmaceutical R&D. It covers company and therapeutic area developments, latest phase changes, licensing opportunities, chemical structures and commercial activities. As can be appreciated from Table 1, the file currently contains some 9,300 records, each referring to a specific compound. The spread of the compounds in *R&D Focus* over the different development phases is also analogous to that in Figure 2. The apparently large number of compounds in the 'active' phase (7,200) must be treated with caution, because the most recent update[2] of these records must also be taken into consideration.

The text abstract encompasses a compound description, scientific and commercial summary. As with most files of this type, the reports are compiled from a range of information sources, including companies involved in R&D, official press releases, meetings and publications. A more detailed recent report on this file can be found in the literature[2]. A 1996 survey indicated that some 17 PDR companies had installed this file in-house.

R&D Insight *File*

Before attempting to compare the relative coverage and degree of overlap of the *Pharmaprojects*, *R&D Focus* and *R&D Insight* files, the latter – the latest arrival on the scene – will be discussed in more depth. At the beginning of 1996, the *R&D Insight* file from ADIS was already installed as a CD-ROM version in seven PDR companies.

The focal point of this file (Table 3), which is also comparable in its range of coverage to *Pharmaprojects* and *R&D Focus*, is enhanced, upgraded data. The coverage includes extensive, critical substance profiles with detailed sections on, for example, pharmacodynamics, pharmacokinetics, together with turnover estimates from Lehman Bros and an attempt to rate the innovative level of a development compound.

Furthermore, the compound profile data is backed up by literature references, about 45% of which are ADIS' value-added literature monitoring service (LMS) literature summaries, which present the main aspects of a cited publication in a brief, coherent form, often including tabular data. Figure 2 presents the distribution of the compounds in the *R&D Insight* file according to development phase.

As with all files of this type, most compounds are in the preclinical phase, the subsequent 'weeding out' is evident. About 70% of the substance profiles contain literature references, less than half (45%) of which are ADIS LMS literature summaries (Figure 2). Figure 3 indicates the distribution of the literature references according to publication year and

Table 3. *R&D Insight* – overview of upgrade and referencing features

Total records (substance profiles)	5,630
Upgrading	
Records with turnover estimates from Lehman Bros	500 (8.9%)
Records with ADIS rating of development compounds	762 (13.5%)
Literature references	
Records with literature references	3,882 (69%)
Total number of literature references	13,220
Number of literature references with LMS summaries	5,874 (44.4%)

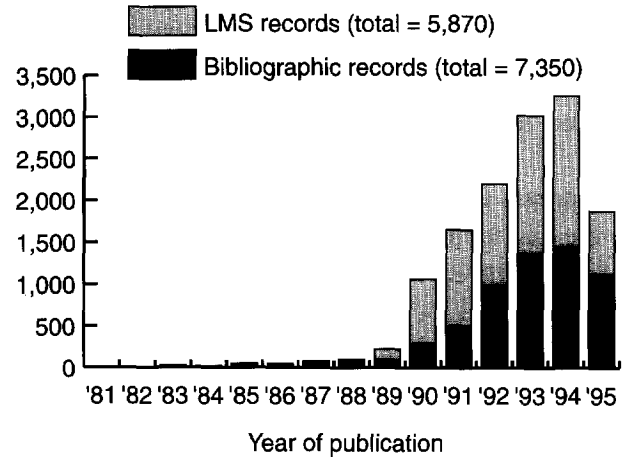

Figure 3. *Distribution of ADIS LMS literature references according to year of publication.*

Box 2. ADIS's rating system for compounds in the preclinical phase[a]

Preclinical rating

A New agent with innovative structure or mechanism of action and/or potentially provides major pharmacological advantages relative to existing compounds of the same class.

B New agent with similarities to existing compounds but some significant pharmacological differences potentially providing important advantages.

C New agent resembling existing compounds with only minor pharmacological differences and potentially providing only modest advantages.

D Uncertain potential or little or no pharmacological advantages evident relative to existing compounds of the same class.

O Insufficient data available to rate the compound at this time.

Supplementary classification

1 New chemical entry
2 New salt/ester or enantiomer
3 New formulation
4 New indication

[a]Source: *R&D Insight* (ADIS).

whether they contain only bibliographic data or the more extensive, and therefore much more valuable, ADIS LMS summaries. The *R&D Insight* file incorporates literature references mostly from 1990 onward. It should be stressed that by increasing the present share (about 45%) of ADIS LMS summaries in the *R&D Insight* file, this resource would become more of a competitor information 'one-stop shop' for users. It is also essential to know what kind of criteria ADIS use to select their LMS summaries for inclusion in *R&D Insight*. In a future publication, we intend to study this aspect in more detail.

An extension of the turnover forecasts from Lehman Bros (see Table 3) – currently only about 9% of compounds – would also be welcomed.

Part of the system used by ADIS to rate compounds is shown in Box 2, and this is a commendable approach. However, only about 14% of compounds are covered, and even then, many are categorized as 'O1' – new chemical entry but insufficient data to rate the compound. Nevertheless, ADIS has made steps to introduce new features, in terms of upgraded information, which have been requested by experienced, critical users of competitor information sources[8] (for example, PDR in 1991).

Record on zileuton. For the purpose of this review, it is useful to examine a typical substance profile for a compound in clinical development. The information on zileuton (Figure 4 shows the introduction) encompasses some eight DIN A4 pages of substance profiles plus about 20 pages of references, some as ADIS LMS summaries. The synonyms, therapeutic uses (as codes) and development status in different countries for the various therapeutic uses are displayed in the introduction. In subsequent pages, the chemical structure is presented plus a turnover forecast and the ADIS rating – again, more of this type of data should be included in other substance records.

After some current news, a detailed, critical review of the compound commences with its pharmacokinetics. Extensive therapeutic trials data are also embedded in the substance profile. The latter part of the substance profile contains 43 literature references. However, in the case of zileuton, a mere 12 of the 43 references are present as ADIS LMS summaries on the *R&D Insight* CD-ROM, although some 37 ADIS LMS summaries on zileuton are actually available online in ADIS' LMS *Drug Alerts* file. The 'missing' 25 summaries would be a welcome enhancement. However, it must also be considered that at least six times as many articles (about 250) on zileuton have been published, about half of which have the name (i.e. synonym, trial number, etc.) of the development compound in the title.

Synopsis

Name:
Zileuton
Chemical Name:
1-(1-Benzo[b]thien-2-ylethyl) 1 hydroxyurea
Molecular Formula:
C11 H12 N2 02 S
Synonyms:
ABT 077; Abbott 64077; A 64077
Brand Names:
Leutrol®
Mechanisms:
5-Lipoxygenase inhibitors
CAS Numbers:
111406-87-2
WHO ATC Codes:
R03; M01; A02X
EphMRA ATC Codes:
R3; M1; A2B9
Originator Companies:
Abbott

Phase Of Development

Indication	Phase	Country
Ulcerative colitis	Discontinued(III)	USA
Rheumatoid arthritis	II	United Kingdom
As thma	II	United Kingdom
Allergic rhinitis	II	United Kingdom
Rheumatoid arthritis	II	Denmark
Asthma	II	Denmark
Allergic rhinitis	II	Denmark
Rheumatoid arthritis	III	USA
Allergic rhinitis	III	USA
Rheumatoid arthritis	III	Canada
Asthma	III	
Allergic rhinitis		
Asthma		

Figure 4. *Introduction from* R&D Insight *record on zileuton.*

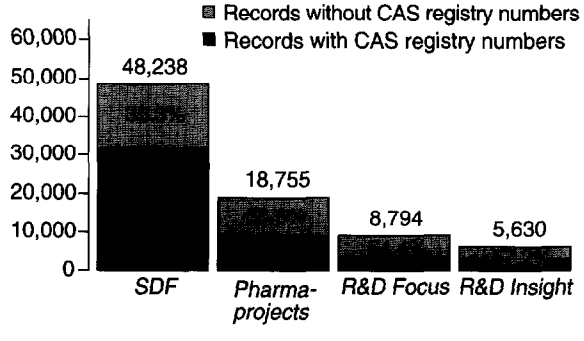

Figure 5. *Number of records and coverage of CAS registry numbers in the* SDF, Pharmaprojects, R&D Focus *and* R&D Insight *files.*

In summary, *R&D Insight* offers some unique features as an information source on competitor products from a publisher renowned for quality products. However, its value could be enhanced if some of the above points were to be taken into account.

Comparison of selected files

The user has to question whether all of these files are really needed – is there not a vast degree of overlap? In Figure 5, statistics on the *Pharmaprojects*, *R&D Focus* and *R&D Insight* databases are presented alongside information on the *SDF* (*Standard Drug File*)[8] or *WDI* (*World Drug Index*) from Derwent Information Ltd. (Derwent House, 14 Great Queen St, London, UK WC2B 5DT), which includes more than 48,000 biologically active substances but lacks information about developer or development status[2]. The comparison presents the various files in terms of numbers of records and CAS registry number content. The latter is an absolutely essential feature for the unequivocal identification of a development compound in the various files.

The CAS registry number content of these files varies between about 35% and 65%. Of course, in the last three files shown in Figure 5 many compounds are in the preclinical phase, where the chemical structure has not been disclosed, and therefore no corresponding CAS registry number is available. Nevertheless, the share of CAS registry numbers is too low and represents a source of frustration for many serious users.

Overlap of CAS registry numbers

Pharmaprojects, *R&D Focus* and *R&D Insight* together include about 13,000 different CAS registry numbers. Remarkably, however, a mere 1,487 are common to all three files (Figure 6). The overlap would be somewhat higher if different forms of the same basic chemical structure, such as salts, had been taken into account.

Figure 7 illustrates the CAS registry number content and overlap of the files. Derwent's *SDF/WDI* file has again been included even though it is really only a dictionary of biologically active compounds rather than a competitor tracking file[2]. This gives a total of about 38,000 different CAS registry numbers of which only about 1,000 are common to all four files. The high degree of overlap between *Pharmaprojects* and the *SDF* is readily explained by the number of older established launch products in both files as well as those in the category 'discontinued projects' in *Pharmaprojects* (see Table 1).

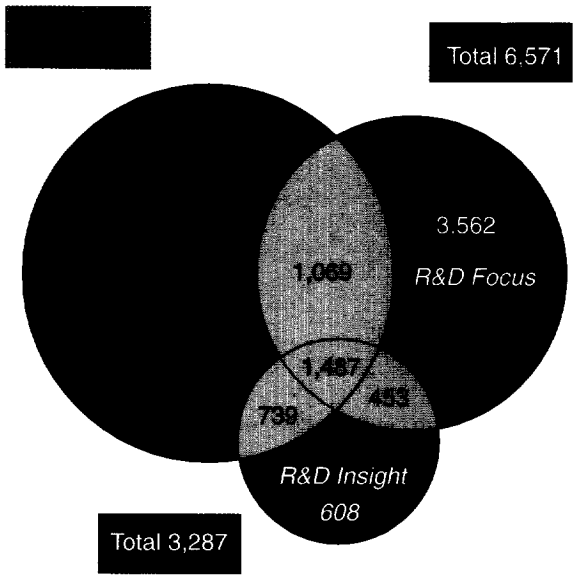

Figure 6. *Overlap of CAS registry numbers in* Pharmaprojects, R&D Focus *and* R&D Insight.

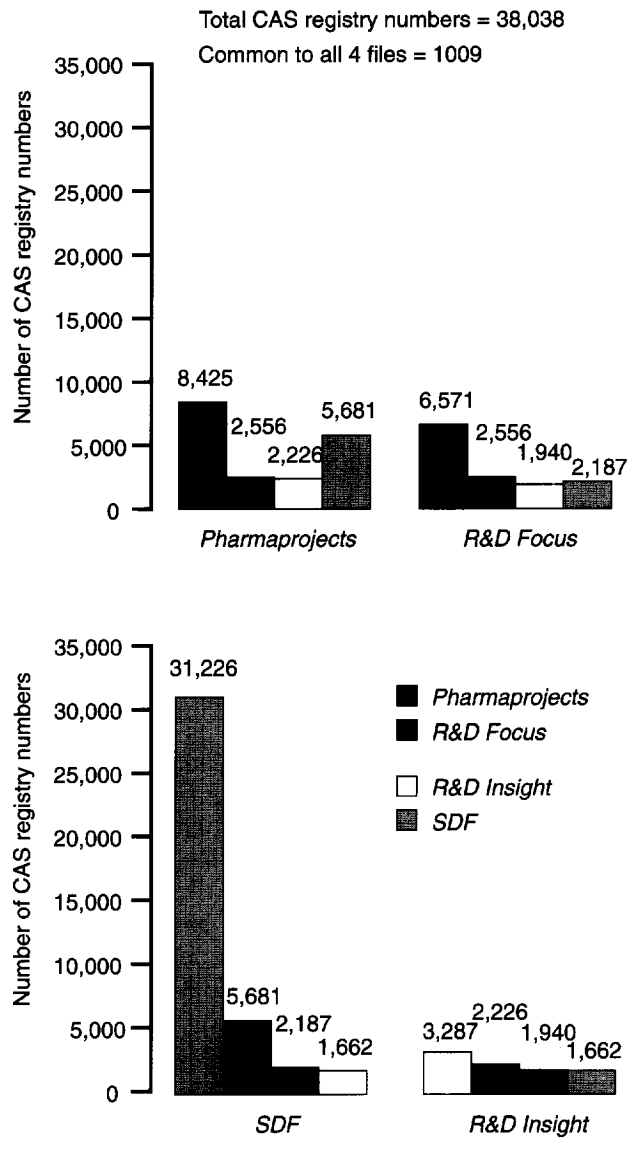

Figure 7. *Relative overlap of CAS registry numbers in* Pharmaprojects, R&D Focus, R&D Insight *and the* SDF *files. The leftmost column of each group indicates the total number of CAS registry numbers in the file specified, the other three columns of the group show how many CAS registry numbers the file has in common with other files.*

Upgrading of data

Data are raw unconnected pieces of knowledge and processing is the intellectual process whereby raw data are sifted, evaluated and upgraded into intelligence.

The enhancement of raw data is becoming increasingly important in the competitive environment of the pharmaceutical industry because of the ever-pervading information explosion. However, those files that provide current information and maintain high-quality control checks and indexing standards are amenable to statistical analysis. The *Pharmaprojects* file has been used for that reason in the example shown in Figure 8.

Analysis of 'memory enhancer' projects

To view which companies are actively involved in memory enhancer projects, a spectral map analysis[2] was carried out based on developer and development phase – the relative positions of the circles to the squares are important. This type of presentation rapidly indicates which organizations (circles) are mainly involved in the preclinical phase or in one of the later development phases (squares). The area of the circles is directly proportional to the number of development products of a particular company.

It is also possible to look at the relationship between the number of projects, their mechanisms of action and development phase in order to make transparent the new and established R&D approaches to this therapeutic area. The statistical manipulation of this type of data demands very high standards from the database producer as far as coherent indexing and quality control are concerned. As suggested above, not all files are suited to such treatment.

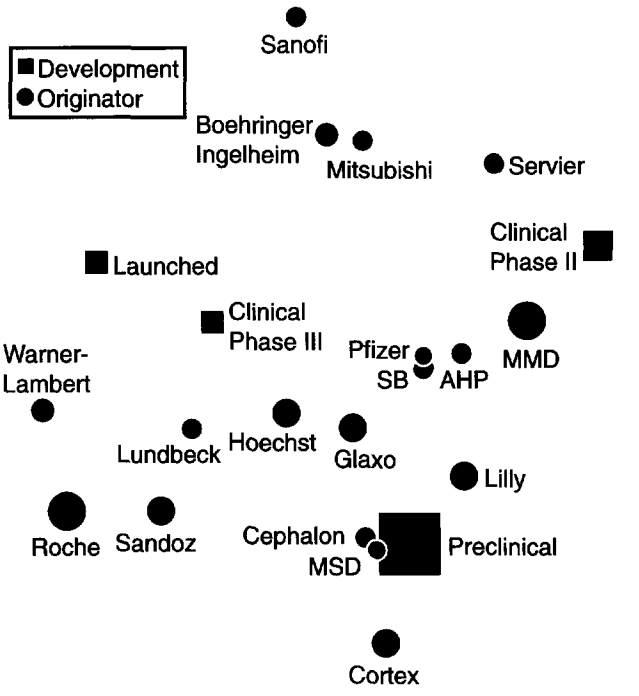

Figure 8. *Spectral map analysis of memory enhancer projects based on number of compounds in different development phases according to originator. The relative position of a circle (organization) to the squares (development phases) indicates the development focus of the organization's memory-enhancer projects. The area of a circle is directly proportional to the number of development products of a particular company.*

Open issues

Some paramount concerns of users within the industrial R&D environment relating to these databases[2,9] are compiled in Figure 9. These issues warrant full discussion but will be alluded to only briefly here.

In some of the files, company names are infrequently updated and it takes considerable time before mergers are taken into account, sometimes companies are even listed that no longer exist.

A common link between the substance records in the different files can, and should, be provided by the CAS registry number which, as already stated, is 'unique' for a specific chemical structure. It would be a welcome step if the policy of CAS were to be modified to enable and encourage a more widespread use of the CAS registry number designation in commercial files of the type under discussion.

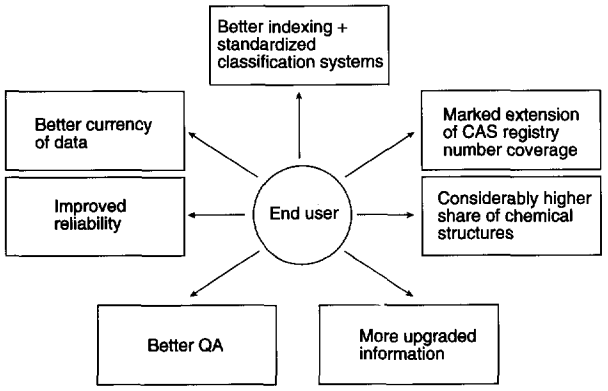

Figure 9. *Unfulfilled needs of end users of competitor information databases.*

There is a strong case for more value-added information, meaning more intelligent analysis of the substance data rather than mere documentation. The concept of a 'one-stop shop' for information on competitor products is attractive, even if no database producer has so far attained this objective.

Concluding comments

In many ways, pharmaceutical R&D is well served by the present range of databases available from reputable database producers. This article has attempted to highlight areas where there is still a need for improvement. Nevertheless, we are grateful that database producers provide this service for the industry and look forward to cooperating with them through organizations such as the PDR[4,6,7] in order to have resources that more fully meet user needs.

REFERENCES

1 *Statistics '96* (1996), p. 30, Verband forschender Arzneimittelhersteller e.V. (VfA), Bonn
2 Mullen, A. *et al.* (1997) *J. Inf. Science* 23 (1), 9–23
3 Busch, T. *et al.* (1990) *Proceedings of the Montreux 1990 International Chemical Information Conference*, pp. 159–176, Springer-Verlag, Berlin
4 Mullen, A. (1996) *Drug News Perspect.* 9, 58–60
5 Mullen, A. (1992) Aims and Activities of the PDR – Pharma Documentation Ring. *Report of the Derwent Literature Subscriber Meeting*, pp. 41–56 (private publication)
6 Dubosc, Y. and Mullen, A. (1994) *Drug News Perspect.* 7, 551–555
7 Dubosc, Y. and Mullen, A. (1997) *Drug News Perspect.* 10, 61–64
8 Mullen, A. *et al.* (1990) Drug Information Files for Inhouse Use. *Report of the Derwent Literature Subscriber Meeting*, pp. 45–59 (private publication)
9 Mullen, A., Blunck, M. and Möller, E. (1996) *211th American Chemical Society National Meeting*, New Orleans, Abstracts Part 1, CINF 20