



Up To infotechnologies srl

Via G. Battistella 4/2 * 31053 Pieve di Soligo (TV) Italia

Tel./fax: +39-0438-842337 * E-mail: info@uptoit.org * <http://www.uptoit.org>

Italian Databases 2001:

A Baseline Market Research Report

Numerical evaluation of the Italian database industry,
with international comparisons and a typical user's profile

Banche dati italiane 2001:

***analisi quantitativa-qualitativa, differenze con il mercato internazionale
e profilo dell'utente tipo***

15 July 2001

Table of Contents

Executive Summary-----3

Riassunto -----4

Italian Databases -----5

 Databases defined and classified -----5

 Data collection and analysis -----7

The State of the Italian Database Industry, May 2001 -----9

Analysis and Previsions for the Future ----- 15

 International Comparison ----- 15

 Who’s Using Italian Databases and Why: What the Numbers Tell Us ----- 16

 Future Expectations ----- 17

About Up To infotechnologies ----- 19

Executive Summary

In order to determine the nature and dynamics of the Italian database industry, Up To infotechnologies undertook a study to identify and characterize databases containing information predominantly about Italy. *Italian databases* are important knowledge tools for business operators, business consultants, accountants, lawyers, scientists and anyone else who needs comprehensive and reliable information about Italy. This study gives industry operators and database users an overview of the market offer, its strengths and weaknesses, and some expected changes and trends. Up until now, little attention has been given to the Italian database industry by the popular or specialistic press, and the present study is the first of its kind.

The market availability of databases about Italy has been monitored since 1998 for the online “Guide to Italian Databases” (<http://www.uptoit.org>). The present study, current as of May 2001, cites 166 Italian databases. Of these 66% are available “online” (via Internet), 25% are distributed on CD-ROM or other removable computer media, 7% are available in both formats and 2% are searchable with assistance of trained personnel. The products cover a wide range of subjects, but the most popular topics are public records and legislation. About two-thirds of the products are available upon subscription, while the remainder can be used at no cost. The vast majority of Italian databases is produced by Italian companies, and is in the Italian language. However, a few products have been produced by companies headquartered in other countries. In addition, some products are available in Italian and English, only English, or multiple languages. Substantially more Italian databases (99 of 166) have been produced by commercial enterprises, although a significant amount has resulted from the efforts of the public administration (46 databases), universities (11 products) or non-profit entities (10 products).

These and other data are illustrated throughout the following pages. In addition, the market offer of Italian databases is compared with that available worldwide, according to M.E. Williams’ article “State of the Database Industry”. Finally, a profile of the typical Italian database user is offered. The information in this study is offered in the belief that it is interesting to database producers, distributors and users, and in the hope that awareness of the market’s characteristics, strengths and weaknesses will lead to an overall improvement of the quality and availability of Italian databases.

2001 © All rights reserved.

Riassunto

La Up To infotechnologies ha intrapreso uno studio per identificare e caratterizzare gli archivi digitali che contengono informazioni sull'Italia, con lo scopo di conoscere il mercato italiano delle banche dati, le caratteristiche dei prodotti e la dinamicità dell'offerta. Queste *banche dati italiane* rappresentano un importante fonte di informazione a disposizione di chiunque abbia bisogno di raccogliere in un modo efficace informazioni mirate. Lo studio, unico nel suo genere a trattare questo nuovo mercato, presenta un panorama dell'offerta, ne descrive i punti di forza e le carenze, ed offre alcune previsioni per il futuro.

Dal 1998, il mercato italiano delle banche dati è stato regolarmente ma informalmente recensito dalla Up To infotechnologies nella "Guida alle Banche Dati Italiane", presente in rete nel sito <http://www.uptoit.org>. Il presente studio, aggiornato a maggio 2001, cita 166 banche dati italiane, di cui 66% è accessibile in rete ("online"), 25% è distribuito su CD-ROM o altro supporto magnetico (portatile), 7% è disponibile in entrambi i formati ed il rimanente 2% è accessibile indirettamente con l'assistenza di un operatore. I prodotti trattano un'ampia gamma di argomenti tra cui spiccano legislazione e dati ufficiali. Le banche dati sono per due terzi disponibili via abbonamento, mentre un terzo è accessibile gratuitamente. La maggior parte delle banche dati sono "made in Italy" – prodotte da aziende italiane in lingua italiana. Alcune banche dati italiane sono offerte da aziende estere, ed alcune – sia prodotte in Italia sia all'estero – sono scritte in lingua straniera (inglese ed altre lingue). Di tutti i prodotti catalogati nel presente studio, 60% sono stati creati da aziende commerciali, 28% dalla pubblica amministrazione, 7% dalle università e 6% dalle associazioni senza fine di lucro.

Questi ed altri dati vengono presentati ed illustrati sulle seguenti pagine. L'offerta di banche dati in Italia viene confrontata con quella internazionale (M.E. Williams, "State of the Database Industry"). Infine, in base alle osservazioni sul tipo di mercato, viene proposto un profilo del probabile utente.

2001 © Tutti i diritti riservati.

Italian Databases

The market availability of Italian databases has been monitored by Up To infotechnologies since 1998 in a regular yet informal manner. In this period, significant changes have been observed in the characteristics of the market offer: new databases appeared, others disappeared, prices increased or decreased and some databases became free, products migrated from telnet to Internet or from CD to DVD, and some subscription policies relaxed. Overall, the market has proved itself to be dynamic, experiencing mutations, transformations, migrations, fusions and, almost, revolutions.

With the most recent update (in May 2001) of Up To i.t.'s "Guide to Italian Databases" (<http://www.uptoit.org/monitor.htm>), it was decided to characterize and quantify this Italian industry in terms of the number of databases, sectors of interest, modes of access, etc. This decision was motivated by the belief that such information is useful and interesting to producers, distributors and users of databases in Italy, and in other European markets.

The following analysis describes the Italian offer of databases, and refers to the situation in May 2001. This baseline study will be the basis for future analyses of trends in this industry. Before presenting the results of the baseline study, descriptions of the classification scheme and definitions of the terminology utilized are provided.

Databases defined and classified

Databases may be defined as electronic archives containing information that can be searched and retrieved using appropriate software. The purpose of a database is to store and organize great amounts of information, and to facilitate the retrieval of such information.

For reasons of practicality and marketing, the content of individual databases is typically restricted to a narrow subject matter. Worldwide, databases are available for almost all subject matters: they collect laws, demographic statistics, newspaper articles, documentary film clips, lists of the chemical properties of drugs, images of trademarks, etc. *Italian databases* also contain information on any subject, but their geographical coverage is limited to Italy. Italian databases collect Italian laws, demographics about the Italian population, articles of newspapers published in Italy, video clips or films produced in Italy or about Italy, chemical properties of drugs authorized for sale in Italy, etc. Italian databases are not only "made in Italy" but they are also produced in other countries, and their content can be in the Italian language or in any other language. However, the

content of Italian databases is always focused on Italy. Therefore, databases that contain information about Italy within a European or international context have not been considered Italian databases, even though they may be useful tools for researching about Italy. Similarly, databases having local interest only, like a university's archive of research projects or a town's hotel listing, have also not been considered Italian databases for the present purposes. Finally, databases of international character that are utilized in Italy have similarly not been considered in this report.

As the previous paragraph illustrates, subject matter represents a first method of classifying databases. A second method of distinguishing databases refers to the format, or type of digital media, in which they are stored. Regarding media type, two main classes of databases are recognized:

- *Online databases* are stored on a computer and are interrogated remotely by users. Today, the majority of online databases are distributed on the World Wide Web. But before the wide availability of Internet, online databases could be accessed through a direct modem connection between the user's computer and the remote server.
- *Portable databases* are interrogated locally by users on their personal computers. However, to be sold or distributed, portable databases are transported on removable computer disks, such as CD-ROMs, DVD-ROMs or floppy disks, or they are downloaded in their entirety as a file from Internet to the users' workstations.

Essentially, the content of an online database is accessed only in part by the user, while a portable database is entirely present on the user's computer, no matter how it was transferred there. For the purpose of this report, a third format of database has been recognized: the "assisted-search" archive. These are electronic archives that may not be searched directly by the user, but instead are searched by trained personnel upon a user's request. Although these databases are not actually available on the market, it was decided to include them in the present analysis for the important information they contain.

A third distinguishing characteristic of databases refers to the nature of the information they contain:

- *Word-oriented* (or text-based) databases are the most common type, having been the first type of database to be developed. Word-oriented databases can contain lists (e.g. directories,

yearbooks or catalogs), bibliographic citations to printed works, or the “full text” of articles, laws, public records or other official documents, for example. For the present study, word-oriented databases have been further classified as directory-catalog, bibliographic, bibliographic plus abstracts (summaries) and full text.

- *Numbers-oriented* databases contain only or mostly numerical data in lists or in tables. Financial and market data and statistics are typical contents of numerical databases.
- Databases classified as having *audiovideo* content can contain images (e.g. diagrams, still photographs, film) and sound recordings (e.g. music, speech). Often audiovideo databanks also contain simple texts and descriptors (key words) to facilitate searching.

Clearly, some databases contain mixed types of records. For example, a database of patent information will include text descriptions of each invention but also drawings and diagrams. However, the characterizing feature of this type of database is that it contains the full text of patents. Therefore, for the purposes of this report, databases with mixed content have been classified according to their most distinguishing feature.

Other features of databases examined in the current study include language, cost (free vs. for-fee) and updating schedule (how often new data are added or old data are modified). In addition, the basic characteristics of database producers, including country of origin and economic sector, have been recorded.

Data collection and analysis

Criteria for selection. To be included in the “Guide to Italian Databases”, a database had to be geographically focused on Italy. If a database’s actual coverage of the country was only partial, it was included if it did not exclude a priori other regions. Additionally, databases had to permit searching according to user-defined criteria. The evaluation of “searchable” archives was quite generous, and some products were included for the value of their information even though they only offered “guided browsing” rather than searching.

Sources. Italian databases were identified, over a period of 3 years, from public sources, including Internet, newspapers (in particular *Il Sole 24 Ore*), and advertising. Therefore, the present collection may have inadvertently omitted some Italian databases, for the fact that they were not identifiable by these common information sources. It is hoped that, as a result of publishing this

report from the “Guide to Italian Databases”, producers of the more obscure databases will contact Up To infotechnologies in order that their products be inserted in future updates.

Data storage and manipulation. Data were collected from public sources as previously described, and entered into a proprietary database for ease of management, updating and statistical elaboration. For each product, the following details were recorded: means of access, subject matter, type of information, subscription policy, language, and updating frequency. In addition, the geographic origin and economic category of the producer were recorded, as was the route by which the producer distributed the product to the user.

The State of the Italian Database Industry, May 2001

As of May 2001, there were 166 products meeting the aforesaid criteria to be considered an *Italian database*. Of these, 110 (66%) were online products, 41 (25%) were portable, and 11 (7%) were available in both online and portable format. A further 4 databases were classified as assisted-search archives, in that they were not directly available for searching by the user. The distribution of media types for the 166 Italian databases is shown in Fig. 1.

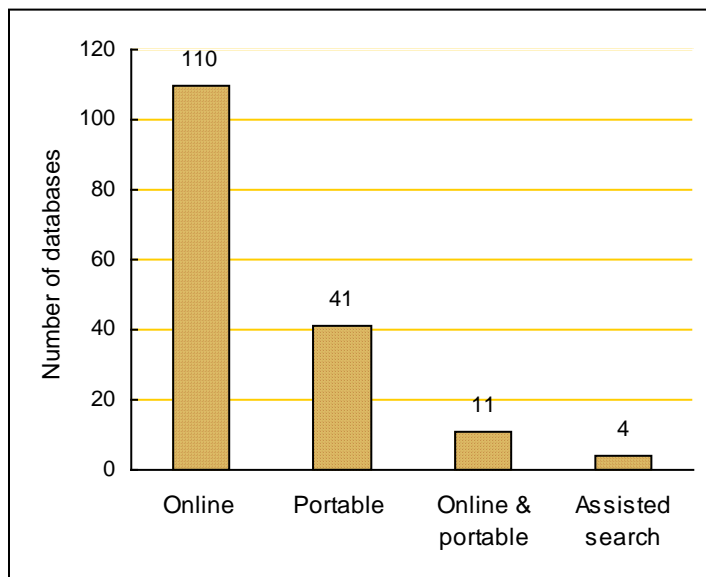


Fig. 1. Media types (means of access) for 166 Italian databases

As expected, Italian databases cover a wide range of topics and interests. When broad subject areas were evaluated, the areas having the greatest number of databases were business and public administration (Fig. 2).

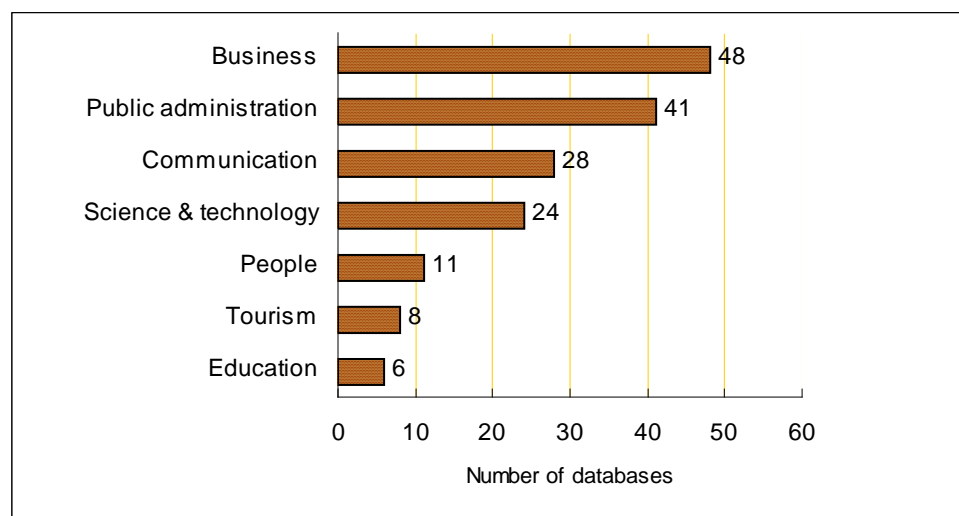


Fig. 2. Subjects of Italian databases, by broad topic area

When categorized into specific arguments (Fig. 3), the greatest numbers of databases were found in the categories public records (business) and legislation (public administration).

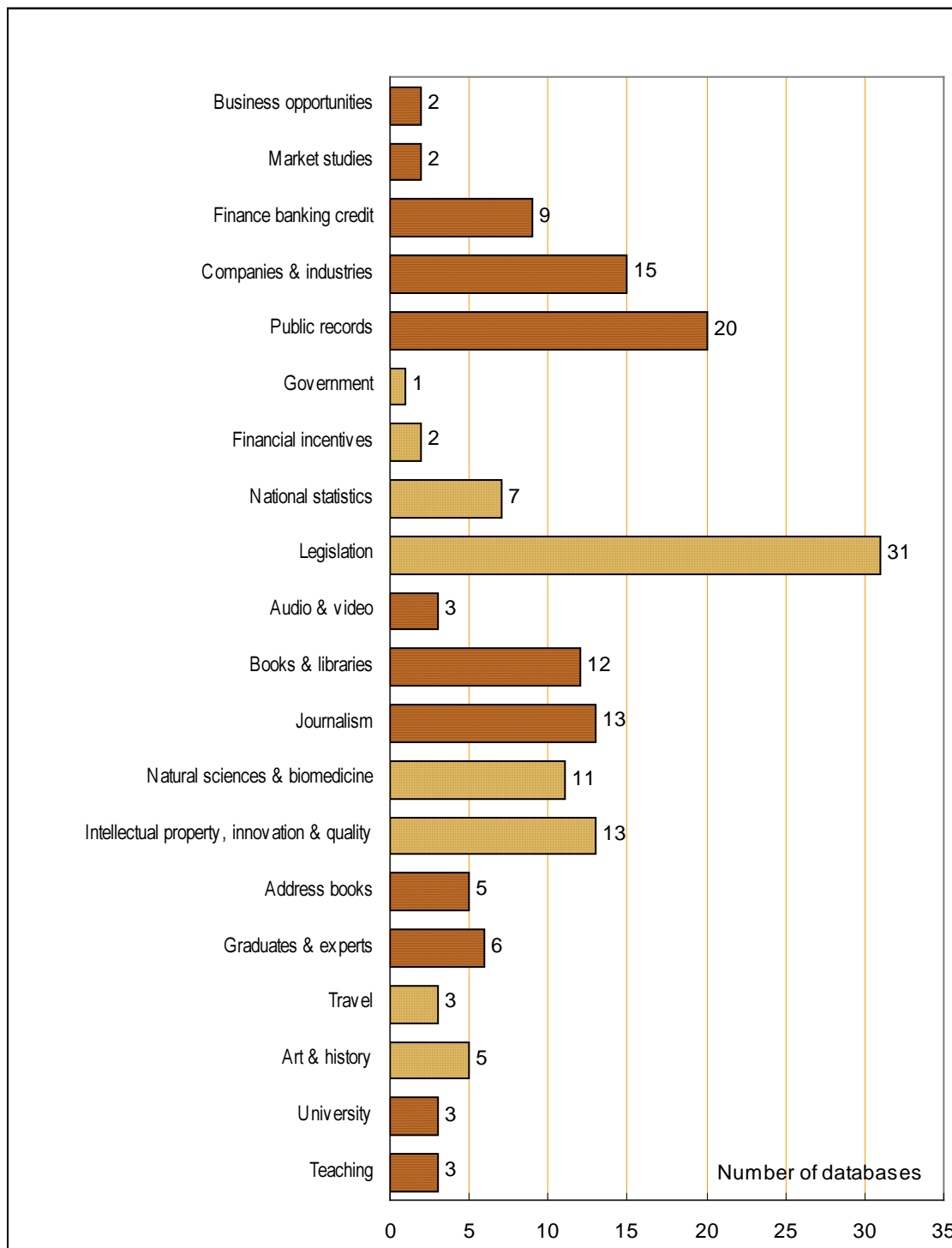


Fig. 3. Subject of Italian databases, by specific argument

In terms of the type of information memorized in the database, most products ($n = 142$) identified in this study were word-oriented (or text-based), but there were 21 numbers-based databases and 3

databases with audiovideo content (Fig. 4a). Among the audiovideo databases, all media types were represented: still photographs, film, and sound recordings. Among word-oriented databases (Fig. 4b), there was a similar number of full-text and directory-catalog products, while there were fewer bibliographic databases ($n = 13$). Of bibliographic databases, about half contained abstracts.

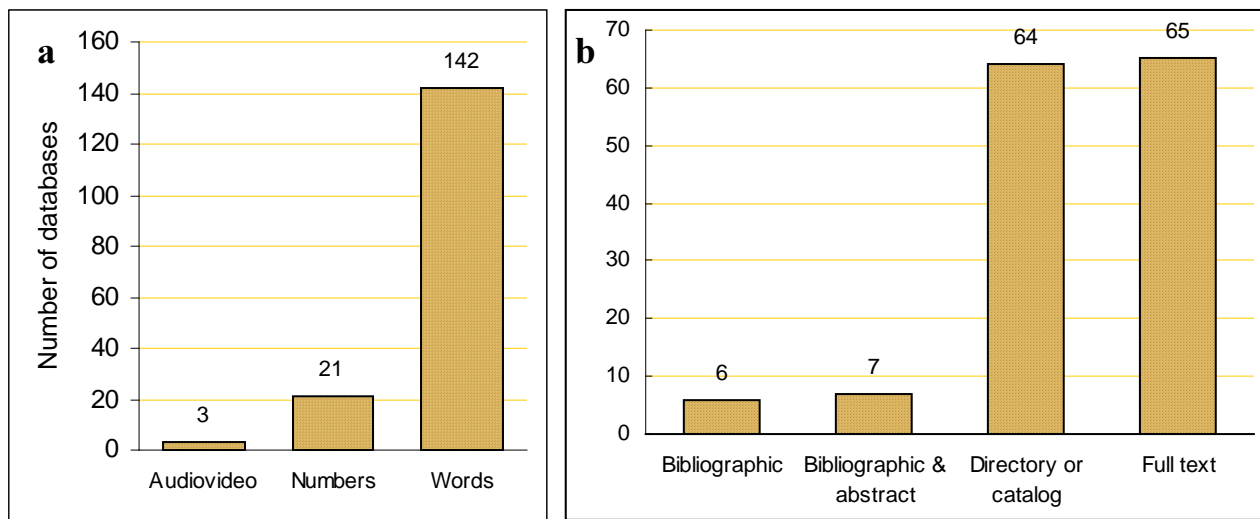


Fig. 4a,b. Classification of Italian databases according to the characteristics of the individual records. a) According to three major categories. b) Subdivision of text-based products

Regarding costs for use, about two-thirds ($n=101$) of Italian databases identified in the study required a paid subscription, but there was a significant number of databases available for free (Fig. 5). Two databases produced by associations were free to members, one of which was also available to non-members for a fee. Almost all the portable databases required a subscription, while the online databases were rather evenly divided between being free or for-fee. Almost all the free databases are available online, but there was one free CD-ROM (ConISTAT) and one free assisted-search databank (FENDAC's Managers and Markets). Of the eleven databases available both online and on CD-ROM, four were free in Internet but a fee was requested to purchase the portable form. In these cases, the producers indicated that the portable version had superior search facilities.

Of the 101 databases requiring a fee for use, only 15 allow the user to pay for the actual amount of information acquired. The others request a financial commitment from the user before providing access: 23 databases require a non-refundable prepayment towards use (sometimes expiring after 1 year), and for 51 products the user must pay for full, unlimited access to the databank regardless of actual usage; this latter category includes the 39 portable for-fee databases. For 12 fee-based products, it was not possible to discern the type of fee schedule offered.

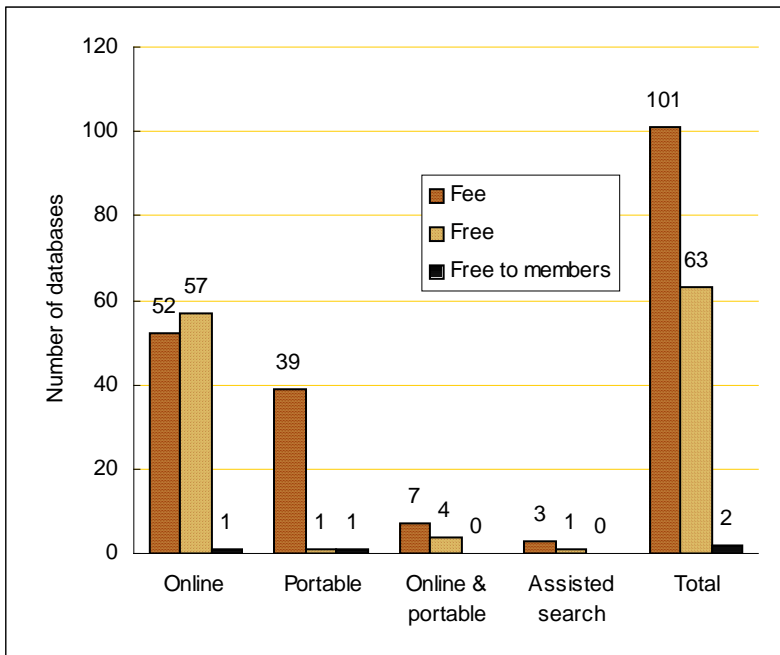


Fig. 5. Subscription policies for accessing Italian databases, by type of media. Of the 166 databases, a total of 101 are fee-based, 63 are free and 2 are free to members of the producing association

The vast majority of Italian databases has been produced in Italy (Fig. 6). Of 166 products, only 5 have been produced by companies based outside of Italy (all available on a for-fee basis), and another 5 by the Italian affiliates of multinational companies (3 of which are available for free).

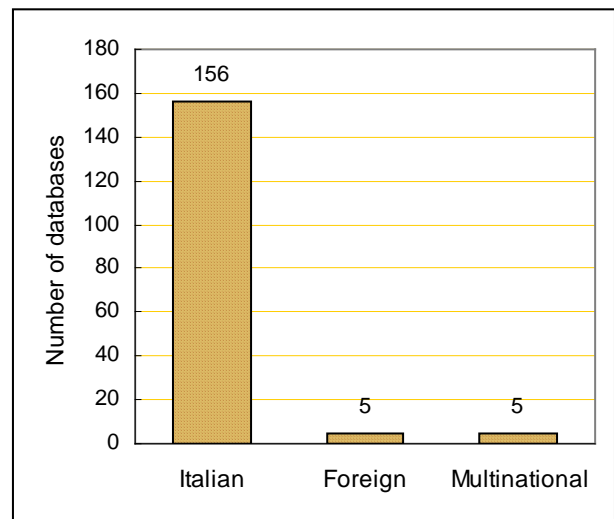


Fig. 6. Geographic origin of the producers of Italian databases

A majority of Italian databases is produced by commercial firms, although many databases have resulted from the efforts of the public administration (Fig. 7). The greatest number of Italian databases has been generated by Infocamere (n = 10) followed by UNI Ente Nazionale Italiano di Unificazione (n = 7). Not surprisingly, commercial firms produced more for-fee databases than free ones (78 vs. 21), while producers from the non-profit, public administration and university sectors

produced more free than for-fee databases (42 vs. 23, with 2 databases being free to members). Commercially generated databases that are offered to users at no cost are supported by online advertising (e.g. SEAT's Pagine Gialle) or are seen as a means of marketing and publicity for the producer.

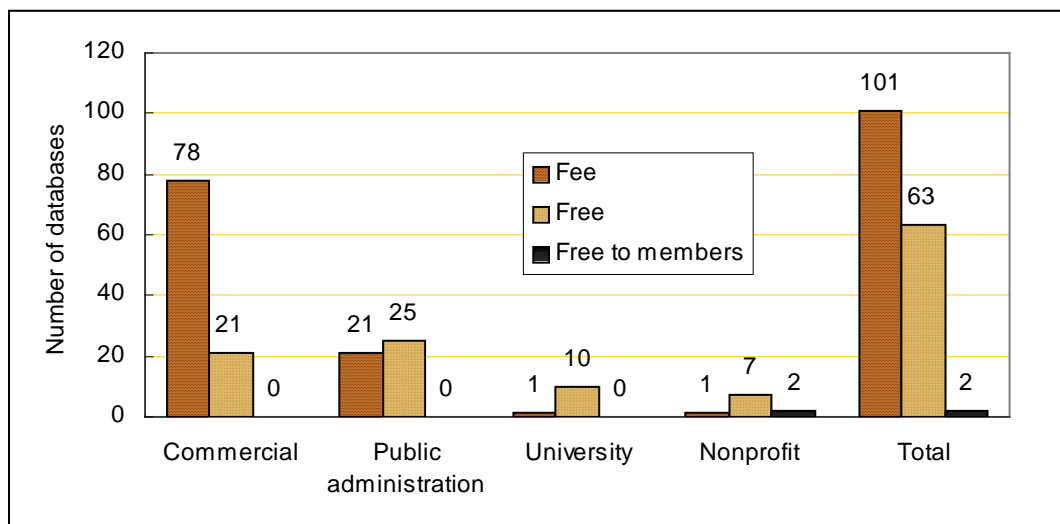


Fig. 7. Economic category of producers of Italian databases, and the subscription policies that they adopt for their products. Of the 166 databases, a total of 99 are commercially made, 10 are from the non-profit sector, 46 are produced by the public administration and 11 are made by university-affiliated institutes or departments

Of a total of 166 Italian databases, 140 (84%) are distributed to users directly by the producer, while the distribution of the remaining products is managed by third-party firms. Among the 101 databases that require a fee for use, 78 (77%) are distributed directly by the producer. Among Italian databases distributed by third parties, 7 are available from the affiliated distributors of Infocamere and 9 products are available by the major database hosting service Datastar.

Not surprisingly, most (90%) Italian databases are written and searchable in the Italian language (Fig. 8). However, some Italian databases are only available in English and others are in multiple languages. Of the 6 databases in the English language, two have Italian-language counterparts (Bridge and Cartesio of Cerved; ITIE and SDOE of Infocamere). Of the 10 databases not produced by Italian companies, 3 are available only in Italian, 1 is only in English, and 6 are at least in Italian and English (3 have multiple languages).

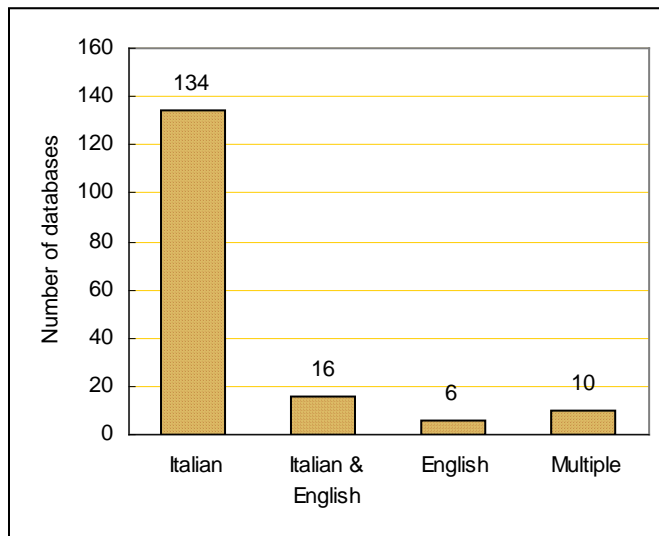
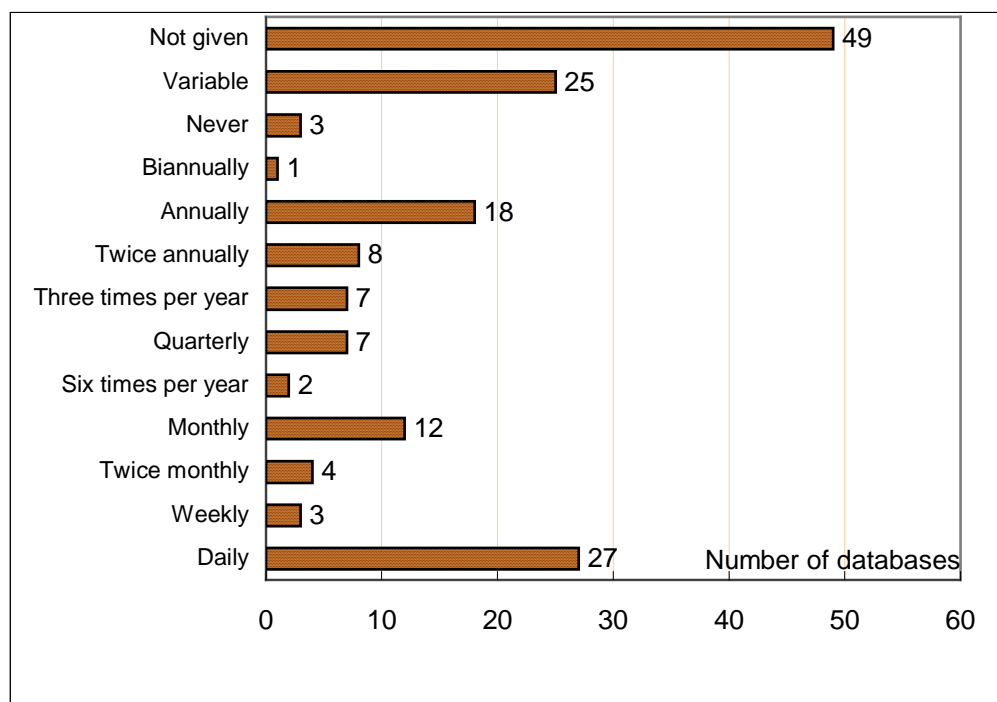


Fig. 8. Language of Italian databases

The updating schedule of Italian databases is quite varied, and ranges from daily to biannually (Fig. 9). However, for the majority of products (30%) it was not possible to discern the updating schedule from the information made available to the public. Three databases are not subject to updating: in two cases because the database contains the results of a terminated study (Intermediate Census of Industry and Services and CLIO Italian Books of the 1800s), and in one full text database because the corresponding print journals are no longer available (MNDO Mondo Economico & L'Impresa).

Fig. 9. Updating frequency of Italian databases



Analysis and Previsions for the Future

International Comparison

A quantitative analysis of the world market for databases has been carried out for several decades by Prof. Martha E. Williams of the University of Illinois, and published annually in the *Gale Directory of Databases*. Prof. Williams has identified and characterized over 10 000 databases worldwide, the majority of which come from North America (63%) and 30% are from Western Europe. Several interesting comparisons can be made between the present analysis and that of Prof. Williams.

Worldwide, 66% of all databases are word-oriented, 17% are numbers-oriented and 15% have audiovideo content; an additional 2% refer to electronic services and software (Table 1). In Italy, these values are 85%, 13% and 2%, respectively. Thus, there are fewer audiovideo databases available in Italy compared to the international market.

Table 1. Italian databases compared to the international market offer, by type of information memorized. Values are percent of all databases

	International market offer (n = 12 417) ^a	Italian databases (n = 166)
Numbers-oriented content	17	13
Audiovideo content	15	2
Word-oriented content	66	85
Bibliographic	22 ^b	9 ^b
Full text	55 ^b	46 ^b
Directory-catalog	20 ^b	45 ^b

^a From M.E. Williams (2000) *The State of Databases Today, 2001*. In: *Gale Directory of Databases*. Gale Research Group, Farmington Hills, MI

^b Percent of word-oriented databases

Among word-oriented databases worldwide, 22% are bibliographic, 55% are full text, and 20% are directory (also indicated are 2% dictionary and <1% patent-trademark, considered full text in the present study). For Italy, these values are 9%, 46% and 45%, respectively. Thus, there are notably fewer bibliographic databases for Italy and notably more directories. The lack of bibliographic databases stems in part from the fact that scientific literature is often indexed in databases having no geographical boundaries. For example the medical literature, including that produced in Italy, is indexed in Medline, Embase and other international databases.

In the world market for databases, the most popular subject areas were science-technology (30%), business (24%), and law (12%). For Italy, 29% of databases refer to business, 25% to public administration (three-quarters of which are law), and 15% refer to science and technology. The lower proportion of scientific databases may stem, in part, from the international nature of scientific research, as mentioned previously.

Referring to media type, international databases are 46% online, 36% on CD-ROM, and a further 18% on other types of portable media. In Italy, 73% are online and 31% are portable (this total is greater than 100% because 11 databases are both online and portable). Therefore, the present survey of Italian database identified a much higher proportion of online databases than were found by Prof. Williams in her study published in September 2000. The reason for this difference is not clear. In part, it may reflect how the database industry in other countries (in particular, in the USA) had its major growth in the pre-Internet decade when portable databases replaced online products accessible by direct modem connection.

Among database producers at the international level, 82% were commercial, 8% governmental, and 8% non-profit/academic (2% had mixed characteristics). The present survey found that 60% of Italian databases is produced by commercial operators, 28% by the public administration (government), and 13% by non-profit and academic sectors. Thus, the Italian database market has a greater contribution from the public and non-profit sectors than from the commercial sector, compared to the international situation.

Only 10 of the 166 Italian databases are distributed via international hosting services (e.g. Datastar, Questel-Orbit). These products make a small contribution to the hundreds of databases that these services offer to their clients worldwide. The reason for this is unclear, but may relate to language, market interest or weaknesses in inter-business communications. It is also unclear, in the future, whether Italian databases will be acquired by such hosting services, or whether an Italian aggregator of databases will emerge. Since these database-hosting companies developed before the wide availability of Internet, it is possible that their services are not needed in a market (like that of Italy) which has experienced its most significant growth in the Internet era.

Who's Using Italian Databases and Why: What the Numbers Tell Us

Knowing the dimensions of the Italian database industry can help create a profile of the typical client, i.e. the "Italian database user". This profile, of course, is valid only if database producers are

sensitive to users' preferences, so that the market offer reflects the demand. Considering the predominance of for-fee products, we assume that the industry is sufficiently sensitive to market requests, and that the present offer reflects, at least in part, what current-day users are interested in. Here is what the numbers tell us:

- The typical user has Internet access, and is able to perform online searching and to manage subscriptions to online accounts (121 of 166 products are available online).
- Users are particularly interested in law and public (company) records. These subjects represent 19% and 12%, respectively, of all Italian databases. The next most populous category is companies and industries, representing 9% of all databases. Therefore, users of Italian databases likely work in law or accounting, or are business operators or consultants.
- The user recognizes the importance of full text information, and has a small interest in bibliographic databases. This fact reflects the relative lack of public reference libraries and document delivery services in Italy. For most users, bibliographic searching is a dead end. This is particularly evident in the medical field where many physicians know that Medline is available free online, but they are inhibited from obtaining the full text of articles because of the high costs of purchasing single articles from journals.
- The user is willing to pay for comprehensive information collections. Of the 166 databases, 61% require some fee for use; of these, 85% require prepayment towards use or prepayment for full access, irrespective of the amount of information the user actually consumes. These subscription policies favor the regular user, while the infrequent user is discouraged by the high initial costs.
- The typical user of Italian databases is knowledgeable of Italian, and is likely to be working in Italy. Few products (19%) are in languages other than Italian, and fewer still (6%) are distributed in international packages of databases.

Future Expectations

The present report provides an overview of the actual market offer of Italian databases, but is unable to quantify the changes in this industry that were informally observed over the past several years. This study, however, is the basis for an analysis of trends in the future. Certain trends and

innovations in the Italian database industry can be foreseen, based on world trends and on European initiatives presently in act:

- The overall number of Italian databases is expected to increase. According to Williams, the total number of databases available worldwide has increased at an average rate of 6% per year since 1994.
- The proportion of audiovideo databases may also be expected to increase. According to Williams, databases with multimedia content first became publicly available in the late 1980s, and have shown a steady growth in total number and in terms of percent of all products.
- The availability of Italian databases in languages other than Italian is anticipated to grow. This expectation derives from European Union initiatives to promote public access to digital content, through programs like the Information Society and EContent. Additionally, commercial database producers are beginning to recognize the international interest for information about Italy and are thus making English-language or multilingual products (for example, Cerved's BRIDGE service). The greater availability of multilingual products will encourage users in countries other than Italy to search for information about Italy.

As the Italian database market grows and matures, further changes may be anticipated regarding costs, subscription policies and distribution channels. The nature of these changes will be revealed in the next annual report of the Guide to Italian Databases.

About Up To infotechnologies

Up To infotechnologies offers information research, analysis and technical editing, with a dual focus. All general commercial-technical information about Italy is handled, especially for an English-speaking clientele. In a second specialization, Up To i.t.'s qualified operators offer medical bibliographic research, medical literature analysis and medical report editing to a professional clientele, mostly in Europe. Up To i.t. is actively involved in working to promote, analyze and improve the offer in Italy of information sources in all subject matters. To this end, Up To i.t. created the "Monitor of the Italian Information Industry" and its "Guide to Italian Databases", both available online at <http://www.uptoit.org>.

President Valerie Matarese, Ph.D. brings to Up To i.t. over 10 years of research experience in various environments, spanning from biomedicine to bioinformatics. She was the 2000 recipient of the M.T. Grenier Award given by the Association of Independent Information Professionals (<http://www.aiip.org>) to an aspiring new member. She is featured in a forthcoming volume of the *Super Searchers* series of books published by Information Today (<http://www.infotoday.com>). She is author of numerous articles of scientific nature, and has contributed articles about the information industry to specialty journals published in the USA, UK and Italy.

For more information about Up To infotechnologies and its services, please send an e-mail message to info@uptoit.org, visit our Web site at <http://www.uptoit.org> or telephone +39-0438-842337.