Diffusion of Legal Data in the Italian System: the Foreign User and the Non Legal User

Ettore Giannantonio


1. THE FOREIGN USER AND THE NON LEGAL USER

On 13th February 1980 Mrs. Simon Veil, Head of the European Parliament officially opened the European Euronet Diane network at the headquarters of the European Parliament in Strasbourg. This is an on-line network for the transmission of scientific and technical data.

On 21st May 1981 the President of the Italian Republic enacted a Decree (No. 322) «Regulation on establishing the legal information service of CED (Computer Documentation Centre) of the Corte Suprema di Cassazione (the Supreme Court) ».

These marked two very important events. Subsequently, this legal data service, that is, consulting data banks of the Centre, is no longer reserved exclusively to legal experts, but is also open to Italian and European citizens in general.

This implies not only the need to enlarge the central system and the transmitting network due to the increase in the number of terminals, but above all a qualitative change in the data in order to satisfy the needs of the type of users and, in particular, foreign users and non legal users.

In fact, a foreign user who wishes to consult or refer to the Centre encounters three main difficulties: the difference in the enquiry system, the difference in the documentation language, the difference in the legal systems.

As regards the non legal user it is necessary to observe that the present legal data systems provide the person carrying out the research only with the documents, the basic material containing relevant information to solve the legal problem. It is the researcher's task to read and interprete the
documents in order to obtain the information for his or her enquiry. However, it is necessary to create systems which allow the user, who is not a legal expert, to obtain the correct information to solve his or her specific problem and that in other words, he or she is not supplied with the text of the document to read or with references to publications to look up or interpret. The user must know exactly whether he or she has the right or the possibility or the power to carry out a certain activity and what are the consequences of his or her work from a legal viewpoint.

This paper intends describing what can be done to overcome, at least in part, the difficulties indicated. In particular, it will describe the Euronet Network, the Common Command Language, the automatic translation of the texts, expert systems and factual data banks.

2. THE EURONET NETWORK

In June 1971 the Council of Ministers of the European Economic Community passed a resolution for encouraging and coordinating the activities of the Member States in the field of scientific and technical information and documentation (Scientific and Technical Information and Documentation – STID). Amongst other initiatives directed at stimulating progress in this sector, it was decided to create a European network for the transmission of scientific and technical information and documentation.

The setting up of a European network for the transmission of data would allow for the creation of a «common information market» giving each researcher in the various member countries the possibility, by using his own terminal, to consult the data bases existing in the Community according to the cheapest possible tariffs and, in any case, independent of the distance between the researcher and the data base.

The CIDST (Comitato per l’informazione e la documentazione scientifica e tecnica) as a subcommittee CREST (Comitato per la ricerca nel campo scientifico e tecnologico) was set up under the same resolution of June 1971 with the task of assisting and advising in carrying out the initiatives adopted.

After research lasting about three years CIDST developed a plan of action in the STID sector for 1975, 1976 and 1977. In carrying out this plan, passed by the Council of Ministers of the European Economic Community on 18th March 1975, the Commission of the European Economic Community signed a contract in December 1975 with the Administration of the French Post and Telecommunications as representative of a consortium of the administrations of the Post and Telecommunications (PTT) of the nine Member States for the designing in detail and the putting into operation of a European network for the transmission of data according to the most modern scientific and technical criteria named Euronet. Subsequently, in
June 1977, another contract was drawn up between the French Ministry for Post and Telecommunications and a representative of the committee of ministers of the other Member States and the French company, Sesa (the same company which set up the French network for the transmission of data packages «transpac») as the representative of a consortium of firms specialized in the production of hardware and software (that is, Logica Ltd of Great Britain, Sesa of Germany, Logica of the Benelux, Sait of Belgium, Christian Rorsing of Denmark, Caradata of Ireland, and Italsiel of Italy) for setting up a European network for the transmission of data shaped on the same principles as the French transpac system.

The network came into operation on 15th November 1979; was officially inaugurated, as mentioned above, on 13th February 1980; began operating commercially on 31st March 1980 and on 24th November 1980 was officially opened to Swiss users also.

On 20th December 1980 the Computer Documentation Centre of the Corte Suprema di Cassazione (the Supreme Court) was linked for the first time to the network.

2.1. The Structure


A node is defined as a packet switching exchange – PSE; a packet as a block of characters of a fixed length; a concentrator or multiplier as a processor which provides for the introduction of several lines for the transmission of data functioning at different speeds on a line at high speed; the concentrator, thanks to the fact its speed is superior to that of the lines converging on it, is capable of serving, in turn, all the various lines or channels dispatching or extracting from each one of them one packet at a time.

The nodes are linked together «in a mesh» with telephonic circuits at 43,000 bits per second (43 kb/s); the concentrators are linked to the closest nodes, the entire network is controlled by a Network Management Centre (NMC) situated in London and linked to the London node by two circuits at 96 kb/s.

Both processors and terminals by using particular protocols may be linked to the network. A protocol means a series of rules which must be observed for the exchange of data between two or more devices which converse between themselves.

In particular, processors and terminals with «intelligent» synchronous transmission, that is, capable of working on «packet mode terminals» can
be linked to the network by dedicated lines at a speed of 2400 to 43,000 bits per second using a standard packet switching control named X 25.

The processors, in order to be linked to the network, must undergo some modification of their hardware and, to be precise, some modification of the front-end of their system of teletransmission. Moreover, a particular software must also be used. It is an interface programme which obviously differs according to the type of processor.

As far as the Centre of the Corte di Cassazione (the Supreme Court) is concerned, the following has been provided for: with regard to hardware to substitute the original CTMC front end and to install a DCP Telcon front end suitable for carrying out a direct link-up with the Euronet node; with regard to software to utilize the interface created for Univac computers, by a group of technicians of Univac and of the Computer Centre of the University of Rome.

Character mode terminals with control TTY (that is, teletype compatible, in practice, almost all existing teleteliers and character mode videos) may be linked to the network by means of the public telephone network at a speed of 110 to 1200 bits (the Rome node is accessible at present with lines at 300 b/s which should soon be brought up to 1200 b/s) or by means of dedicated lines with a speed from 5600 to 1200 b/s.

In particular these are linked to the node using protocol X 28. This mode, by means of the PAD (Packet Assembler/Disassembler), packages the characters arriving at the terminal before guiding them into the network and unpackages the messages arriving from the network for the terminal and transmits them character by character.

Instead, the asynchronous video terminals may be linked to the network using the DEVIP (Data Entry Virtual Terminal Protocol). The link-up is being planned and developed by Olivetti in collaboration with the French Téléstèmes.

The list of the models of compatible terminals with the most important technical characteristics is set out in the publication of the European Community, «The Terminal User Guide», edited by Diebold Deutchland GmbH, Beratungsunternehmen für Informations Technologie, Feberbachstr.

2.2. The Host Computers

More than thirty host computers belonging to the various countries of the Community are at present linked to the Euronet network.

In particular, the following should be mentioned: Blaise (British Library Automated Information Service) of London with regard to bibliographical information and library automation, and Infoline of London with regard to engineering and science; for France, Cated (Centre d’assistance technique
et de documentation) of Paris with regard to civil engineering, Cisi of Paris, ITF (Istitut textile de France) of Boulogne sur Seine, Spi (Société pour l'informatique) of Paris, Télésysèmes of Paris, Thermodata of Grenoble in the termochemical and medical sector; for Germany DIMDI (Deutches institut für medizinische dokumentation und information) of Cologne, Fiztechnik (Fachinformatzszentrum technik) of Frankfort in the technological sector, Gid of Frankfort in the bibliographical sector, Inka (Fachinformationszentrum Energie, Physik, Mathematik) of Karlsbuch with data banks in the field of mathematics, physics, astronautics, energy and space; for Holland, Epo (European Patent Office) of Aia, for Belgium, CTI (Centre de traitement de l'information) of Brussels with regard to statistics in the sector of the public economy and in nuclear chemistry and pharmaceutics; for Denmark, Data Centralen of Copenhagen with regard to the environment; for Luxembourg, Echo Service (European Commission Host Service) for experiments, statistics and information on the Euronet network; for Italy, CCI (Centro di calcolo interfacoltà) of Rome, Cilea (Consorzio interuniversitario lombardo per la elaborazione automatica) of Milan, Iris (Information retrieval service), an office of the European Space Agency of Frascati which has about 7 million documents on line relating to physics, electronics, EDP, ecology and astronomy, CEC Joint Research Centre of Ispra (Varese) in the nuclear sector and CED of the Corte Suprema di Casazione in the legal sector.

Overall there are more than thirty centres and about one hundred data banks covering a wide variety of fields: a complete list is to be found in the special publication of the Community Diane Director, edited by Euronet Diane Information DG 13, Commission of the European Communities, BP 1907, Luxembourg.

The host computers which have adhered to the network and have, for this purpose, signed a memorandum of understanding addressed to the Commission constitute the Diane (Direct Information Access Network for Europe) group, with the task of dealing with common problems. Some of these centres have, also, constituted a special association called Ehog (Euronet Host Operators Group).

CRID (Centro di riferimento italiano Diane) has been set up by the Istituto di Studi sulla ricerca e documentazione scientifica of the Italian National Research Council.

2.3. Ways of linking

To become a Euronet user it is necessary:

1) to have, either by renting or in ownership, a compatible terminal and the relative modem at one's disposal and request a line from SIP for the transmission of data;
2) to request a code for access to the network from the Ministry of Post and Telecommunications;
3) to request a contract of use and the codes of access from the host computer for the databases one wishes to interrogate.

As far as the lines of transmission are concerned it is necessary to keep in mind that the link up may come about in two different ways: or by using the circuits of the public telephone network or by using the circuits designed for the exclusive use of a determinate processing system. In the former case the link up is made by means of the various urban, district and compartmental switching centres of the public network and is therefore defined as a link up on the switching line; in the latter case, instead, the link up is made directly between the terminal and the computer and is defined as a link up on a private or dedicated or point to point line. In both cases the use of special apparatus is necessary. These are the so-called modem (modulators — demodulators) capable of converting digital signals in analogic signals capable of being transmitted on telephone lines and vice versa.

The different types of modem are characterized by their speed of transmission and the kind of transmission (simplex, halfduplex or fullduplex) which can be carried out. It may be internal or external to the terminal.

The switching line represents particular advantages. Above all, it allows for a link up between the computer and a large number of terminals. It is sufficient that each terminal has its own telephone number. It also allows each terminal to be linked with a large number of computers. Finally payment is made on a basis of the time and the distance of the call made and not, as with a dedicated line, on the basis of a fixed price for renting the line independently of the frequency of use.

On the other hand, the link up on a switching line does not permit a high speed of transmission and is, moreover, not always available. The user, in fact, could find the number of the node, or of the concentrator, or of the data bank engaged by other users. Finally, in order to carry out the link up a manual call is necessary using a sequence switch disk or the ordinary lifting and rotating selector of a normal telephone. Nevertheless, attachments exist which permit the connection between two points situated in the switching network without the need for any manual intervention:

a) a dialing adapter which gives the computer the possibility of automatically dialing the telephone number of the desired terminal;
b) automatic answering which gives the terminal the possibility of automatically stating that it is ready to receive or transmit.

For access to the network by a dedicated line there is an annual rent for a telegraphic permit of 120,000 Italian lire for every line certified in the name of the legal residence of the user as provided for by D. M. 1 March 1976. There also an annual rent for the use and maintenance of the telephone
type circuits put at the exclusive disposal of the user for direct access to the node Euronet according to DPR 30 December 1979 n. 667.

For access to the network by a switching line which does not include an acoustically coupled terminal, the annual rent is that for a telegraphic permit.

For those with an acoustically coupled terminal the tax is 20,000 Italian lire annually.

To obtain access to the network it is necessary to request an identification code or NUI (Network User Identifier) from the Ministry of PTT, Direzione Centrale dei Servizi Telegrafici, Divisione I or IV, Viale Europa 160 00144 Rome specifying the telephone line of access to the network (whether switching or dedicated), and the type, speed and method of transmission from the terminal.

Communication costs are calculated on the basis of two elements:

a) the cost of the use of the national network including the following elements: a once only tax for the registration of the user, a periodical rental for the NUI, a fee for the use of the public telephone network for access to the node of the Euronet network situated in Rome;

b) the cost of the international network calculated, independently of distance, on the basis of the following elements: duration of link up (at present 35.30 Italian lire per minute of transmission) and volume of the data transmitted (17.65 Italian lire for 640 bytes or characters transmitted).

The service contract of the Centre of the Corte Suprema di Cassazione is at present regulated by DPR 21 May 1981 n. 322 and a request should be made to the Direction of the CED (Computer Documentation Centre) of the Corte Suprema di Cassazione (Palazzo di Giustizia, Piazza Cavour, Rome) or to the Ministry of Justice (Ministero di Grazia e Giustizia, Direzione Generale degli Affari Civili e Libere Professioni, Via Arenula, Rome). Furthermore, other contractual formulae are being developed by the Management of the Centre whilst at a community level, some research has been carried out for the harmonization of the service contracts of the various host computers of the Euronet network.

3. The Common Command Language

The first difficulty met with by the user of Euronet in consulting the various data banks of the Community is the necessity of knowing the various command systems for carrying out the search for the information required. For this reason the Community has carried out a study of the better harmonization of the various command systems, not in the sense that it intends imposing the adoption of a particular system on everyone but rather that, on the basis of an investigation of a wide series of programs applied to
the search for information, the fundamental functions of this searching and an elementary form of the commands necessary for putting it into effect have been singled out in such a way that the various Centres may have a reference point for bettering the conformity of their systems in relation to the others.

The fundamental functions have been classified into the following categories:

1) general functions;
2) functions for beginning the search;
3) functions for the formulation of the queries;
4) output functions.

It has been held that every query should begin with a symbol or a keyword and end with a fullstop or comma or a return word. Every word or symbol of the query is to be followed by a space. Every instruction is to be asked for with various synonyms in the same language or even in different languages. Every instruction is to be followed by one or more points or data of the research given in free form or in some cases, in a pre-established order. The instructions are to be abbreviated as much as possible, in general to a single character or coordinated amongst themselves. The searches are to be numbered, and also the replies.

Moreover, the user should have the possibility of asking the computer by the use of a particular command (HELP) for advice and instructions which appear preceded by the words H Prompt. For leaving the Prompt series it is sufficient to formulate the desired instruction preceded by a fullstop.

The symbol ? will be used to indicate further information on a particular aspect.

Base to identify the file requested;
Find to begin the search;
List to visualize a list of terms in alphabetical order;
Relate to visualize a series of data in logical order;
Display to visualize the text;
Format to determine the format with which the texts must be visualized;
More to visualize further documents;
Delete to cancel the previously formulated requests;
Save to retain a particular research formulation;
Own to utilize one’s own search strategy;
Stop to terminate a search.

Moreover, a series of standard abbreviations such as AU (author), TI (title), SC (source), CC (classification code), TT (thesaurus terms), LA (language), RN (number of publication), YR (year), PM (number of document), PC (country of document), JN (journal title), CO (code), AB (abstract), AN (abstract number), CN (contract number) have been formulated.
A search phase should be made up of: a) research data and b) logical operators.

A typical string could be formulated in this way: Find '..........' research datum '..........', logical operator '..........' research datum '..........'. Alternative data should be placed in brackets.

Research data could be: 1) an abbreviation recalling a previous search; 2) an abbreviation recalling one or more terms which have appeared on the terminal; 3) an abbreviation identifying a channel (AU, LA, TT, CC, etc.); 4) an abbreviation identifying an element of the document (YR, Lit, type of journal or datum).

The logical operators may be Boolean (and, or, not) or the operator of the word within the document with respect to the others in which

(W) means followed by...;
(W t n) followed by... at a distance N;
F in the same phrase;
C in the same document.

The study was carried out by INSPEC (Institute of Electrical Engineers) in 1976. The grade of standardization reached should be sufficient to permit a program, run from the host computers or the computer of a node, to automatically translate the commands used by the user in the standard channels of the network.

At present the Common Command Language may be utilized in interrogating the data bases of the German systems, DIMDI, Echo, FIZ-Technik, Gid and Inka (research system Grips/Dirs 3 of Siemens); CNUCE (research system Stairs of IBM), of the French Guestel (research system Mistral of GII-HB), of Esa Irs (research system Guest/Recon derived from Dialog of Lockheed) of Cilea (research system Unidas) of CED of the Corte Suprema di Cassazione (research system Italgiure-Find).

3.1. Examples of Retrieval Using the CCL

The following shows how retrieval in the CED banks may be carried out by any user linked to the Cassazione (the Supreme Court) system through Euronet using the Common Command Language.

First of all, any user linked to the Centre by a specific line must turn on the terminal and the modem in order to be linked up. After a few seconds the DC (digital classification) diagnostic line will appear on the display.

Instead, if the Euronet user has a terminal which has direct access to the branchpoint, it is necessary to:

1) turn on the terminale and the modem. The Euronet signal will appear on the display XXXYZZZ, where the three Xs indicate the Euronet bran-
chpoint (the Rome branchpoint is indicated by the figures 003) Y the number of the commutation module (at present 1) ZZZ the three decimal figures indicating the physical range of access to the PAD.

2) Transmitting a message of the following format by pressing the operating keys:

    NUI-NUA (CR)

that is a message made up by the address of the system whose data banks are to be consulted or referred to.

(NA: network user address); the NUA of CED of the Corte di Cassazione (the Supreme Court) is 0222 3071.

The user’s code or password (NUI: network user identifier) is given by the Ministry of the PTT to the user at the time of subscribing. It is important to keep this information confidential since invoicing is based on it.

The comma sign is to indicate the NUI CR carriage return, that is the sending of the message by pressing the right key.

Example of a call sign of the Centre of the Corte Suprema di Cassazione (the Supreme Court) by a user whose password is 222 ABCD 12345.

    222 ABCD12345-0222230710, (CR)

The string must be operated on the keyboard within 60 seconds. If format errors are made or if the string is not operated within 60 seconds the network replies Error or the three characters LIB followed by other characters which serve to specify the causes of the failure in connecting. Otherwise, it indicates that the link has been carried out by sending the COM signal.

3) Connect to the Centre of the Corte di Cassazione (the Supreme Court) by operating the Euronet string: the user’s code assigned to the user by the Centre.

For example, if the user’s code is EDD1, the user must operate the following on the keyboard:

    EURONET: EDD1

the system will reply: You are connected with the Corte di Cassazione (the Supreme Court) – Rome – Proceed well.

If the terminal is not connected directly to the breakpoint, but through the public connected network, it is necessary:

1) to turn on the terminal and the modem by turning on the specific switch;
2) to make sure that the data switch – telephone (data voice) is in the telephone position (voice);
3) select the telephone number of the breakpoint (the telephone number of
the Rome breakpoint is 06-587901 for terminals having a speed reaching 300 bit/s; 06 587901 for terminals having transmission speed ranging from 300 to 1200 bit/s;
4) as soon as the acoustic signal of the modem is heard – this operation must be carried out within ten seconds;
5) press the capital letter H and then the key (CR = the code for the Constitution) between H and CR must have a pause of not less than 200 milliseconds and not over 3 seconds. This operation serves to synchronize the speed of the terminal with that of the line, but it is necessary when an asynchronous terminal is used having a prefixed speed;
6) at this point the network will transmit indications on the active link to the PAD with the print EURONET XXX Y ZZZ. Once this has been indicated: You are connected to the Corte di Cassazione (the Supreme Court) – Rome – Proceed well.

The user may make enquiries to the banks of the Centre by either using the Italgiure-Find system or the CCL.

If the latter is going to be used, it is necessary to first of all operate the bank request on the keyboard by the base command or control, followed by the name of the bank to be consulted.

For example, if documents in the Civil bank are to be consulted, it is necessary to operate on the keyboard:

**BASE CIVILE (CIVIL BASE)**

The processor will reply: Beginning of the Civil research – Date – Hour.

To formulate the retrieval string, the user must operate on the keyboard the letter F followed by the data to be retrieved. For example, if one wishes to select all the documents containing the pseudonym word, it is necessary to operate on the keyboard:

F PSEUDONIMO; (F PSEUDONYM)

The reply will be:

RI 3 documents 0.3 sec 137785*3

To read the document it is necessary to operate the following order on the keyboard:

SHOW

In a few seconds the text of the document will appear on the display. To close the research it is necessary to operate the following on the keyboard:

STOP

If one wishes to carry out research by using a different channel from that of the text words, it is necessary to operate on the keyboard the letter F followed by a blank space and then the signal of the research channel, the
equal sign, the research data. For example, for documents classified under the big item 112, it is necessary to operate on the keyboard:

\[ \text{F SCR} = 112; \]

To select documents classified under item 112009, it is necessary to carry out the following operation on the keyboard:

\[ \text{F, SC} = 112009 \text{ (SC = complete subitem)} \]

To select all the ‘massime’ (abstracts) which refer to article 2699 of the Civil Code, it is necessary to operate on the keyboard:

\[ \text{F RN} = \text{CC 2699} \text{ (RN = Regulation references/CC = Civil Code)} \]

To select all the documents containing the word SWITZERLAND in its whole conceptual range, it is necessary to carry out the following operation on the keyboard:

\[ \text{F PK} = \text{SWITZERLAND} \text{ (PK = Semi del linguaggio, language parts)} \]

To select documents on the basis of the syntagm PUBLIC ACT, it is necessary to operate the following on the keyboard:

\[ \text{F SINT} = \text{PUBLIC ACT} \]

To link two or more data by means of logical operators AND OR and NOT, it is necessary to operate in full on the keyboard the operator one intends using. For example, to select all the documents containing the words CORRUPTION and MINOR, it is necessary to carry out the following:

\[ \text{F Corruption and minor} \]
\[ \text{or} \]
\[ \text{F (And Corruption: minor)} \]

To select documents containing two syntags PUBLIC ACT and PRIVATE CONTRACT, it is necessary to operate the following on the keyboard:

\[ \text{F SINT} = (\text{PUBLIC ACT AND PRIVATE CONTRACT}) \]
\[ \text{or} \]
\[ \text{F SINT} = (\text{AND PUBLIC ACT; PRIVATE CONTRACT}) \]

To select documents containing the syntagm PRIVATE CONTRACT, but not the syntags PUBLIC ACT and CIVIL STATUS, the following is necessary:

\[ \text{F SINT} = \text{PRIVATE CONTRACT AND (NOT SINT = PUBLIC ACT AND CIVIL STATUS)}; \]

Multi-channel research may be carried out bearing in mind that it is not necessary to indicate the channel of the text words, whereas it is always necessary to indicate the other channels.
For example to retrieve documents containing the words corruption and minor published in 1980, the following is necessary:

\[ F \text{ AN} = 1980 \text{ AND (Corruption and minor)} \]

or

\[ F \text{ AN} = 1980 \text{ AND (AND Corruption: minor)} \]

The order $\$\text{BACK}$ of the Italgiure Find system is replaced by the order $\$\text{DELETE}$. Masking of data is carried out by replacing the unknown characters with the letter X.

For example, to select all the documents of the 1970s, the following is necessary:

\[ F \text{ AN} = 197X \text{ (AN = year)} \]

The order $\$\text{Mask}$ is replaced by the order $\$\text{DEFINE}$. For example the order $\$\text{Mask}$ of the Italgiure Find system will, thus, be:

\[ \text{Define M} = Y \]

The spectral analysis is not envisaged in the Comman Command Language. To carry it out, it is necessary to refer to the Italgiure Find research system by means of the order $\$\text{OWN}$.

For example, to carry out a spectral analysis year by year, the following is necessary:

\[ \text{OWN \$\text{SPECTR:AN;}} \]

4. The Translation of the Thesaurus

One of the great difficulties of a foreign user when asking for and obtaining information from a CED data bank is the fact that generally documents are stored in Italian and, therefore, cannot be understood by the foreign legal expert.

To overcome this difficulty, it is important to have a series of language classification systems which allow the researcher, who does not understand the Italian language, to carry out research and retrieval by words in his or her own language and obtain a more or less partial translation of the documents as a guideline to their content.

To achieve this objective, it is necessary to first of all store a glossary which has the translation of the words in the documents in the language of the researcher – practically, it is necessary to store as many glossaries as the different EEC languages.
In this way, for instance, the English user, who wishes to retrieve laws, court report 'massime' (abstracts) or articles in legal literature which contain the word *fire*, must carry out the following keyboard operation:

\[
\text{F ARSON} \\
\text{or} \\
\text{F FIRE}
\]

Storing of a translation glossary of words of documents in any EEC language is not sufficient to perform good and efficient research of documents by the foreign researcher by words in his or her own language. In fact, there is no univocal relation between language and meaning.

Language characteristics – that is of any language – are variety, ambiguity and indefiniteness.

Language varies so much that the same idea or concept may be expressed in different ways and forms. It is ambiguous because the same word may have multiple meanings. It is indefinite because it does not always determine or define its subject accurately.

Variety and ambiguity cause great difficulty in the retrieval of documents by merely indicating the words in the text. In fact, variety implies that the language of the document retrieved due to the recurrence of morphological variations, the use of synonyms or circumlocutions, compound words or more specific terms, may be different from that used by the researcher. Consequently, it may result in the unsuccessful selection and retrieval of a relevant document (so called silence or failure to recall, completeness). Besides, language ambiguity may result in the selection of an irrelevant document (so called noise).

These two problems, which are features of any document research by natural language words, are more frequent in the translation of the inquiry from one language to another. In fact, the problems caused by synonymies, circumlocutions, compound words or more specific words of the original language are added to the problems caused by the synonymies, circumlocutions, compound words or more specific terms of the language of the documents. Thus, the difficulty caused by the 'polisemie' (multi-word parts) of the original language are added to those caused by the 'polisemie' (multi-word parts) of the language of the documents.

For instance, in the English language, the meaning of the term *arsen* may also be explained by terms such as *fire* or *combustion*. Likewise, in the Italian language instead of the word 'incendio', it is possible to find the words 'fuoco', 'combustione', 'autoconsumbizione' or even 'piromania'.

To avoid problems connected with research and retrieval by words of speech, a thesaurus, with original characteristics, has been compiled for the Italgiure-Find system. It is based on an estimated analysis of the language so as to provide the processor with some fundamental information of the
grammatical, morphological or semantic nature for each word. This language «processing» consists, in particular, in the following operations:

a) the ‘lemmatization’ of words. A grammatical code is placed before each lemma – that is a symbol consisting in a letter of the alphabet – this indicates to the computer which grammatical rule of the Italian language must be applied to automatically coin all the morphological variations which that lemma may undergo; b) the decomposition of words into fundamental ‘semi’ (word parts) of the language or, as linguists would say, semantic categories or first semantic factors; c) linking of more specific terms, which are not subject to decomposition into ‘semi’ (word parts) with more general terms; d) linking of all words having the same root with only one lemma known as group heading. Also words having a different root and which are its synonyms are connected to the latter. Thus, neither manual classification of each document nor reference to the Thesaurus by the researcher before every retrieval are necessary. It is sufficient for the researcher to indicate to the computer whether the selection of the documents containing the word in its textual identity or the whole conceptual text is required by the user.

The decomposition of words into language ‘semi’ (word parts) seems to be the fastest and most convenient also with regard to retrieving a document compiled in the foreign language by «key-words» in his or her own language.

In fact, whereas the translation from one language into another of words having a special meaning is complex and long (due to the great quantity of words in the most important languages), as well as being difficult (it is not always possible in another language to find the corresponding term which gives the exact full meaning), the translation of ‘semi’ (word parts) or rather language ‘semi’ is faster and easier. This is because there are only about three thousand ‘semi’. It is easier, since the ‘semi’ are represented by words having unequivocal and elementary meanings.

The translation of the Thesaurus into the different EEC languages would entail an important progress in research and retrieval of documents in the CED legal data banks by words of a foreign language. In fact, the researcher will be able to ask for, through a special channel (the PK or conceptual word channel), not only documents which contain the word used by the glossary to translate the term he or she used, but also all the synonyms, circumlocutions or compound words.

For instance, the use of the term fire in its full conceptual meaning range by the researcher will not only allow the retrieval of documents containing the word fire, but also those containing the words arson, combustion, etc. (likewise, the Italian terms ‘fuoco’, ‘incendio’, ‘combustione’, ‘ustione’ or ‘piromania’). The following string:

\[
\text{F PK = SALE AND BOVINE AND SICKNESS}
\]
will allow the selection not only of documents which contain words such as
sale, boune and sickness, but also those which mention buying and selling
cows and bulls, specific sicknesses such as tuberculosis and epizootic
aphtha. The use of the following term:

\[ F \, PK = \textit{switzerland} \]

will allow the selection not only of documents which contain the word
Switzerland, but also those which contain the adjective Swiss, the phrase
Swiss Confederation, the Ticino or Tessin Canton as it is also called, or
Zurich.

The possibility for the foreign user to use his or her own language when
retrieving documents by speech words does not exclude, however, the other
difficulty, which consists in the fact that documents, once selected, are pro-
vided to the user in Italian except for a few banks (such as Rebis, BID,
Enlex), where the documents are also stored in foreign languages.

To avoid this problem it is necessary (at least for that which concerns the
legal data banks) to provide a translation of the subject classification used
for this type of documentation. In fact, the titles or items of the subject
classification are at the beginning of each document and are, or rather
form, as a whole, with a variable part of each document, the actual title.
The translation of the subject classification will allow the foreign user not
only the use of the items of the subject classification as retrieval channels
(retrieval channels SC and SC3), but also to have the small title of the
‘massima’ (abstract) translated into his or her language and to have the
necessary guidelines, which should be sufficiently accurate, when referring
to documentation selected by the computer.

For instance, if the researcher is concerned with knowing whether a di-
vorced British subject can freely marry in Italy and whether the divorce
sentence passed in Great Britain must be examined by an Italian judge, it is
sufficient to carry out the following keyboard operation according to the
common command language:

\[
\begin{align*}
F \, PK: & \quad \text{Gran Bretagna} \\
\text{or} & \\
\text{Regno Unito}
\end{align*}
\]

or even for an English legal expert

\[
\begin{align*}
F \, PK: & \quad \text{Great Britain} \\
\text{or} & \\
\text{United Kingdom}
\end{align*}
\]

the answer or result will be:

RI 72 documents 2 sec. 140291*72
the researcher may then ask:

F Marriage
or
Divorce

the answer or result will be:

R2 2 documents 2 sec. 72*1944

If the researcher wishes to read the document, it is necessary to carry out the following keyboard operation:

SHOW

This is how the document will appear on the display console to the English language user or operator:

Division 1 Sentence 823 of 14/3/68

*ANNOTATION*

Chairman Pece
Plaintiff Rossi di Monti

Rapporteur Spagnoletti
Defendant Forbes de Grana

Family – Marriage – Contract in Italy by foreign parties – Free state or status of British or Irish citizen – Divorce decree granted abroad – Acknowledgement – Need – Exclusion – Conditions.

After having checked the free state or status of a foreigner, who intends marrying in Italy, it is important to bear in mind that both in Great Britain and Ireland a divorce decree granted abroad is valid without the need of any acknowledgement as long as it has been granted by a competent judicial authority within the territorial jurisdiction where the parties are resident in matrimonial bond.

In this way, the user has obtained an initial indication on the content of the document.

In addition, the database is ready to collaborate with the EEC multi-language teams for the application of language classification systems, such as the Systran system for automatic translation of the ‘massime’ (abstracts) text.

5. COMPUTER GLOSSARIES ON COMPARATIVE LAW

A third difficulty encountered by the foreign user when enquiring and asking for information from the legal data banks of the Centre consists in the differences between one’s own legal system and the Italian legal system. The difference, which is particularly important for Anglosaxon legal experts, may imply that in the retrieved document there is reference to a legal institution which is unknown to one’s own system and is, therefore, unknown to the user or to an institution, which although finding a formal correspondent has a structure or an aim which is essentially different.
In such cases, the English term cannot be literally translated into a corresponding Italian term or vice versa. For example, it would be a serious error to translate into Italian the term «consideration» or into English the expression «interesse legittimo». Instead, it is necessary to leave the term in the original language and provide a special file, where the user who wishes may consult the specific meaning of the term he or she has found.

Thus, a legal computer glossary is formed: a glossary where, for each term, a description of the corresponding institute is envisaged. Besides, the glossary tends to be structured on a comparative basis: that is the explanation will be given in the language of the foreign user and in order to stress the different features compared to the corresponding institutions in the system of the foreign user. This will involve many different definitions – and therefore many different glossaries – as many as the foreign legal systems taken into consideration.

6. THE NON LEGAL USER, EXPERT SYSTEMS AND FACTUAL DATA BANKS

Any non legal user who wishes to enquire and ask for information from a legal data bank meets certain difficulties for which expert systems have been created to overcome.

As already mentioned, the present legal data systems only provide the researcher with documents, material containing information which is important in order to solve a legal problem. It is in the interests of the researcher to read and interpret documents in order to extrapolate information which answers the enquiry. On the contrary, the expert systems tend to aid the researcher in the task by providing the solution to specific problems, the way a legal consultant would do. The user is not provided with the text of the document to be read or the references to publications to be retrieved or, in any case, interpreted, but the user is told exactly whether he or she has the right, the possibility or the power to carry out a certain activity and what are the legal consequences of his or her work.

To provide such a consultancy activity, the expert system uses a query or enquiry technique, which is quite the opposite to a common system of research in legal documentation: it is not so much the user who puts questions to the computer as the computer to the user. It asks him for all the factual data required to be able to say whether the situation of the researcher completes the details of the case in question described in abstract by the legislator or to be able to indicate all its legal effects-(technique of the query answering systems).

To carry out this task it is necessary for the system to have a factual data bank, that is a data bank where all the assumptions and formalities of a situation or an event may be described in the most detailed manner.
Factual data banks and expert systems, although indicating different real situations are terms which are sometimes used as synonyms: they both indicate very specialized systems in which both the explicit and implicit knowledge is emphasised, formalized and used that researchers of a particular subject use. The theoretical assumption on which they are based is that expressed by Goldstein, Minsky and Papert according to which gradual clarification and analysis of a well defined context is more useful than prior construction of universal theories.

A famous expert system, in the medical field, is Mycin, whose task consists in diagnosing infectious diseases in the blood, meningitis and in prescribing a pharmacological cure.

As regards the legal field, expert systems could be particularly useful in those sectors where typical situations occur, that is those giving rise to many problems which are similar. It is sufficient to think of taxation, of tenancies or labour law. However, they become particularly efficacious whenever they can process data supplied by the researcher (for example to calculate the amount due as tax, fair rent or superannuation).

Such systems are under way in the United States of America. In Italy, an expert system in use is the Noemi system, a system for programmed professional aid for notary work.