INFORMATION RETRIEVAL

FOR LITERATURE AND FACTUAL DATA

IN PHOTOGRAMMETRY AND REMOTE SENSING

Prof.Dr.Jürgen Hothmer
Chairman WG VI - 6 (Bibliography)
Rheinlandstr.19, 6231 Schwalbach - FRG

ABSTRACT

Information retrieval requires a database and a host computer. Most scientific
fields dispose nowadays on their own database. In contrast, we do not have as
yet a comprehensive database in our domain. Information on literature in our
field is scattered over many databases which makes retrieval cumbersome and
expensive.

The paper analyses the present state and summarizes details for establishing
ISPRS - IRS, the Information Retrieval System of the International Society for
Photogrammetry and Remote Sensing, which is tailored to cover surveying &
mapping including remote sensing. ISPRS - IRS should be available online, and
also free-of-charge by post card approach. For deciding whether there is at
present really a need to implement ISPRS - IRS, the next phase should be to
collect data for assessing capabilities, efficiency, reliability and costs of
available bibliographic and factual information.

1 Feasibility of computerized literature retrieval

The ultimate objective of publishing is that the content be read. However, the
quantity of published material is ever increasing according to an exponential
law (Hothmer 1982, 1983), yet the capacity of readers is limited. Efficient
capabilities for selecting relevant literature can be considered as an escape
from this dilemma.

Bibliographies are a well known tool to allocate appropriate literature sources.
In the field of photogrammetry, the 'ITC International Bibliography' started
1954 with a literature card indexing system. This type of literature document-
ation ceased to exist due to lack of demand. Abstract journals are another mode
to register literature. Yet, this type as well as card indexing bibliographies
have the inherent disadvantages that retrieval through tying together several
specific keywords is hardly possible. As an example, the query : " Analysing
stressed trees in a tropical urban environment " requires entries under several
headings. Searching for all keywords separately is a tedious and time consuming
undertaking. Even more important is that those literature documents can not be
retrieved which combine two or more keywords, and that is what actually is
required. In contrast, this demand can be accomplished by applying electronic
data processing.

It is, therefore, not a surprise that quite some scientific disciplines have
meanwhile switched to computerized literature retrieval. Several hundreds of
literature databases cover many fields. As an example, the chemistry professionals
can retrieve in a database comprising 7,6 mio literature documents. The growing

rate in information business has two digits. An international database exhibition and convention INFOBASE is being held annually in Frankfurt - FRG with half of all earthwide available 3.ooo databases being represented.

2 Structure of computerized information retrieval

The structure of information retrieval is shown in figure 1. Essential elements are database, host, telecommunication link, and the user.
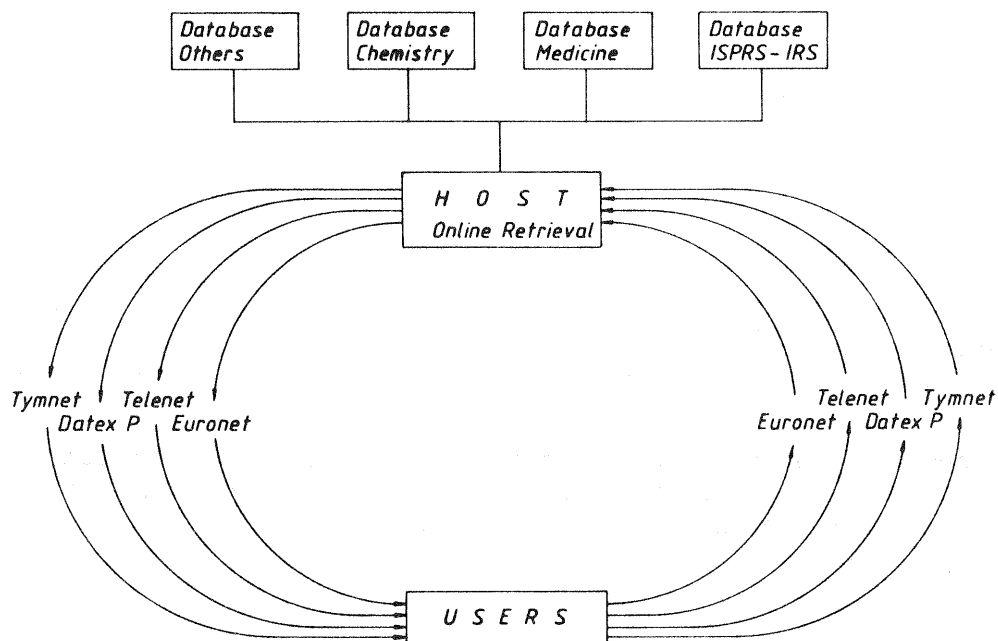
Figure 1

Structure of information retrieval

2.1 Database

Information is stored in databases. Information can be literature documents or factual data. An example of factual data is information on all companies in a country allowing, inter alia, to retrieve all enterprises offering remote sensing services. Presently, there are 3.ooo databases available earthwide. A database may have a storage capacity of up to 1oo mio words.

The focal point of any information system is the database. Any retrieval can be just as good as is the database. Consequently, the staff for establishing and maintaining a database for a specific field should be professionals in that discipline. As they should analyse the relevancy of any literature document, they may be called Analysts.

A database producer must dispose on hardware. In our field, a medium size computer with appropriate storage capacity is considered to be a suitable equipment. Analysts maintain the database. Any Analyst should dispose on a terminal to key in documents. All inputs are transferred weekly or monthly to the host, online or by mailing a magnetic disc or tape.

Database producers license potential hosts as online vendors. Hence, a database may be accessible at several hosts.

2.2 Hosts

To allow for efficient retrieval, a host must dispose on a powerful computer with a very short cpu time, and on sophisticated software. As both are very expensive and, furthermore, as various peripheral units for the benefit of users are costly as well, it is feasible that one host accomodates many data-bases. As an example, ESA - IRS in Italy, the host of the European Space Agency, offers some 12o databases; this computer has entry ports for 2oo users simultaneously in time sharing, nevertheless guaranteeing response time less than 1o seconds, also when retrieving 7,6 mio documents.

There are earthwide 45o hosts. Any host necessitates a retrieval language which are, unfortunately, not standardized. Thus, searching in different hosts requires the capability to manage different retrieval languages. This is, doubtless, a handicap for users. To counteract, there is a trend to diminish the number of retrieval languages. A good example is that the hosts STN International (Scientific & Technical Information Network) Columbus - USA and Karlsruhe - FRG and Tokyo - Japan agreed to use just one retrieval language. That allows to operate on a compound communication whereby the user has access to any database being implemented on either one computer. The user dials up one host for a search but he may automatically be switched via a gateway to another host. He does neither know nor care whether his search is operated by the computer in USA, FRG or Japan.

Online access is accomplished with a telephone link. In some countries, however, the telephone network is not always reliable with the effect that a word gets sometimes lost in oral communication. This has little significance as the redundancy in speaking is considerable. It is completely different for online retrieval. If just a single bit gets lost then the computer may produce odd results. Hence, a practical point of quite some impact for users is that the host be linked to a technically reliable telephone network.

Presently, hosts report that databases are used by 5oo.ooo professionals in USA, 2.ooo in UK, 1.5oo in France and 5oo in FRG. The 1987 sales are $ 3,2 bio in USA, $ 8oo mio in Europe with $ 72 mio in FRG.

2.3 Telecommunication from user to host

The telephone is used for this communication with a modem as interface. Several links such as Tymnet, Telenet, Euronet and Datex P are available as is shown in figure 1.

A modem (modulator - demodulator) is a device which translates digital data, being sent from the terminal, into a frequency signal which can be transmitted over a telephone line to the host. Modems are being built according to the transmission speed, e.g. 3oo bits per second (bps or baud) or 1.2oo bps. Standards of the Comité Consultatif International Télégraphique et Téléphonie (CCITT) for modems have been adopted globaly. When operating at a transmission rate of 3oo bps, the terminal screen will be filled with a speed allowing to read any word at the moment of first display. When operating with 1.2oo bps, then all of the screen is filled immediately. The present telephone system allows to transfer 4.8oo bps as a maximum.

Meanwhile, a completely new transmission system is being implemented in some countries, viz. the Integrated Service Digital Network ISDN (Hothmer 1984). All signals must be digitized before transmission. As an example, for digitizing the human voice its amplitude is measured 8.ooo times per second within an ISDN chip. When disposing on optical fibres, the binary digits are transmitted as light impulses. Light has a high frequency (10^{15} Hz), hence the transmission capacity is tremendous. The CCITT, being responsible for international

standardization, has set the transmission rate at 64.000 bps within the ISDN.
So far, we had to dial one number for the telephone, and other ones for telex,
telefax etc. The great advantage of ISDN is that it can transfer data, text,
voice and images on one single line.

The Interactive Videotex is developing rapidly. As this system will allow
communication round the clock 'from an armchair at home' with any remote com-
puter using the TV set added by a keyboard, we will provide some details.

Videotex services were launched in Europe : PRESTEL 1979 in the UK, VIDITEL
1980 in the Netherlands, BILDSCHIRMTEXT 1981 in FR Germany and TELETEL 1981
in France. Experiments are running in Canada, Japan and USA. Three videotex
systems are in global use :

+ CEPT, an alphanumerical standard used throughout Europe on the basis
  of an agreement of the Comité Européenne de la Poste et Télégraphie facing
  interchange of vidoetex systems of European countries. Joints of all hard-
  ware manufacturers within CEPT are standardized providing compatibility
  also with external computers, thus enabling users to access hosts via
  videotex.
+ TELIDON, an alphageometric standard developed in Canada.
+ CAPTAIN, an alphaphotographic standard developed in Japan is
  specifically adaptable for the japanese characters of writing.

The format of the CEPT screen comprises 24 lines with 40 characters each. Hence,
it differs from the computer screen, and this may cause difficulties.

Interactive Videotex is, so far, not successful in all operating countries. The
most advanced is France with 2,6 mio users (Bohm 1987).

2.4 User's hardware

The user needs for online retrieval a telephone, a modem, a terminal being tele-
type compatible, thus applying the ACSII transmission code, and a printer. A
PC can be added for having an 'intelligent terminal' for more efficiency.

2.5 Costs for users

The costs for online retrieval have three components :

+ Telecommunication charges to access the host;
+ charges for the services of the host, and
+ fees for the database, normally being included in the invoice of the host
  who transfers this license fee to the database producer.

The costs depend on the time consumption. A very rough estimate for one online
retrieval is $ 100.

2.6 Acquiring full documents

The ultimate objective of literature retrieval is to dispose on the full text
of relevant publications. It does not make much sense to accomplish retrieval
within minutes whilst having to wait half a year for obtaining the full docu-
ment. To counteract this dilemma, an indispensable requirement for efficient
retrieval should be that the database includes information on where the full
document is available, and that the software of the host facilitates ordering.
Tehnzen 1982 published details with prices and practical hints for users from
the library point of view.

Meanwhile, Philips and Sony have invented optical storage on compact discs (CD) with a capacity of 55o Megabyte which is equivalent to 15o.ooo single-spaced typewritten pages A 4. The access time to any data on the CD is less than one second. This CD may revolutionize our subject as it is nowadays technically feasible to have full documents available of all items within a literature database. At present, the European research project ADONIS is in full swing; libraries in USA, Australia and Japan participate. CD - ROMs (Compact Disc Read Only Memory) are used to test materializing immediate full document supply (Tehnzen 1987).

3 Present state of literature documentation in our field

Hereunder, an attempt is made to review several well known sources. Although any care has been taken to be comprehensive, omissions can not be excluded.

3.1 Non - computerized systems

3.1.1   Zeitschrift für Vermessungswesen, FR Germany, publishes as from 1875 once a year (July issue) about 2.6oo literature references without abstracts. Photogrammetry and remote sensing as well as the wide field of surveying & mapping are covered. 2oo journals and series publications of all continents and originating from 14 different languages are evaluated.

3.1.2   The Chinese Society of Geodesy, Photogrammetry and Cartography publishes a trimonthly abstract journal covering their field including remote sensing. 63 periodicals and series publications of all continents and originating from six different languages are evaluated.

3.1.3   VINITI, USSR, publishes as from 1954 the abstract journal GEODESIA Y AEROFOTOSJEMKA covering geodesy, photogrammetry and remote sensing. VINITI is an institution with a staff of 1o.ooo being engaged in literature searches for all fields including translations into Russian.

3.1.4   The AIDOS database, Germany DR, covers surveying & mapping and remote sensing (Peschel 1983). AIDOS includes also summaries. Any language of litera-ture is considered. Entries in the database are in English and German. There are 3.5oo new entries per year. Online access is not offered. Retrieval is in German by writing to VEB Kombinat Geodäsie und Kartographie, Gohliser Str. 4, Leipzig 7o22 - GDR. The cost per query is DM 15o.

3.2 Systems with computerized online retrieval

3.2.1   GEO Abstracts, Norwich - UK, publishes bimonthly different issues of an abstract journal for different disciplines. Their main concern is geography. Issue G covers surveying & mapping and remote sensing with 2 - 3 ooo new entries per year. They had 1987 six issues which sell for $ 51. Summaries are included. All languages of literature are considered for the database. The entries in the database are in English only. Online access is available via DIALOG. The host provides no assistance for acquiring full texts of retrieved documents.

3.2.2   The GEOLINE database, Hannover - FRG, covers several geosciences including surveying & mapping with remote sensing (Nowak 1982). Publications which are available at the Institute for Applied Geodesy in Frankfurt - FRG will be docu-mented with summaries for the database. Entries are in English and German. Online access is available via STN Karlsruhe - FRG with English as retrieval language. The host provides assistance for acquiring full texts of retrieved literature references.

3.2.3   RESORS (Remote Sensing Online Retrieval System), Ottawa - Canada, is a database concentrating on remote sensing and touching surveying & mapping only as related to remote sensing. No summaries are included. The database contains 6o.ooo entries with 5.ooo added each year. Languages of literature considered for the database are primarily English and French. Cracknell et al 1987 published some 6.ooo documents of the period July 1985 - June 1986 with addresses of authors. Search requests can be mailed to RESORS, 1547 Merivale Road, 4th Floor, Ottawa K1A OY7, at no costs. The only host available for online retrieval is the computer at RESORS itself with an individual retrieval language. No assistance is rendered to acquire the full text of retrieved documents.

3.2.4   The NASA - USA and PASCAL - France databases include some documents on photogrammetry and remote sensing. The AGRIS database from FAO - Italy is on agriculture and has also references on remote sensing with abstracts for some entries.

3.2.5   The USA National Technical Information Service NTIS database was established and is maintained by the US government to make publicly available reports on research projects (Clark 1987). Each year some 6o.ooo reports are added from 2oo US Federal agencies, along with 15.ooo reports from agencies from other countries. A journal is published twice a month to identify new reports entered into the database. Online access is available at the hosts DIALOG Information Service, ORBIT Search Service, and BRS Bibliographic Retrieval Service. The full text of 1,7 mio different technical publications is for sale at the NTIS.

3.3   Summary of the present state

3.3.1   Comparing printed information and online information retrieval, a book with an alphabetical register is the less expensive source for successful search if the user knows exactly what he is looking for. In contrast, when searching for a subject which can not be defined easily, then computerized online retrieval in a database is more successful as it permits to search for various words and, above all, allows to combine several keywords. Online retrieval is the more expensive system but allows more demanding queries.

3.3.2   Valuable literature documents for online retrieval in our fields are scattered over many databases. That presents some disadvantages : (a) The user has to access different hosts. Further, he has to be familiar with the individual retrieval languages of the hosts.   (b) Assuming within a database on geography nothing but 5% is of concern to our fields. Nevertheless, we have to pay for all the content of the database when doing online retrieval.

3.3.3   The reliability of most databases for our discipline is not known. A test of reliability of NASA, PASCAL and GEO Abstracts databases has been conducted by ten Haken 1982. We quote his conclusion : "The results achieved indicate that the investigated databases do not constitute a fully reliable retrieval system for scientists working in the field of photogrammetry and remote sensing".

3.3.4   Some databases do not offer summaries which are normally considered to be an essential feature when retrieving literature. The capabilities of these databses should be assessed.

3.3.5   Access to databases via online retrieval requires hard currency and is, hence, available only for a priviledged group of users, merely excluding poor countries.

3.3.6   Although most photogrammetrists and remote sensing specialists are well familiar with computers, they are hesitative towards online retrieval of literature. Continuous marketing is necessary, particularly at universities. As it stands now, present days students are the non - users of online retrieval of tomorrow.

The foregoing reveals that in our field an adequate and reliable literature retrieval system does not exist in any country earthwide. Therefore, ISPRS embarked to fill this gap. Details of what has been done have been published by Hothmer 1984. The following is a summary report of that paper.

The basis for the operation of ISPRS - IRS are the detailed Specifications as attached to the quoted publication.

The substantive set up is shown in figure 2. An ISPRS - IRS Office is responsible for the database. Co-operation should be established with as many institutions as is feasible and possible. Co-operation was settled with the Sister Socie- ties International Society for Photogrammetry and Remote Sen- sing (ISPRS), International Car- tographic Association (ICA), International Association of Geodesy (IAG), and International Federation of Surveyors (FIG), and this writer was assigned Con- vener ISPRS - IRS. Co-operation with other Societies promoting and applying remote sensing was negotiated.

Co - operation with International Societies

| ICA | FIG | IAG | ISPRS |

Others particularly within the variety of fields applying remote sensing

ISPRS - IRS OFFICE responsible for database

| RESORS Canada | VINITI USSR | ISTIC China | Japan | IFAG FRG |

| ZFV FRG | GEO Abs. UK | IAG GDR |

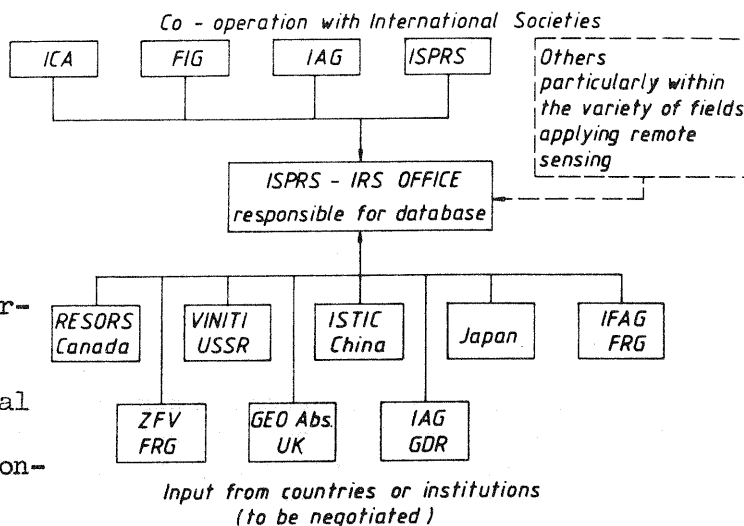Input from countries or institutions
(to be negotiated)

Figure 2

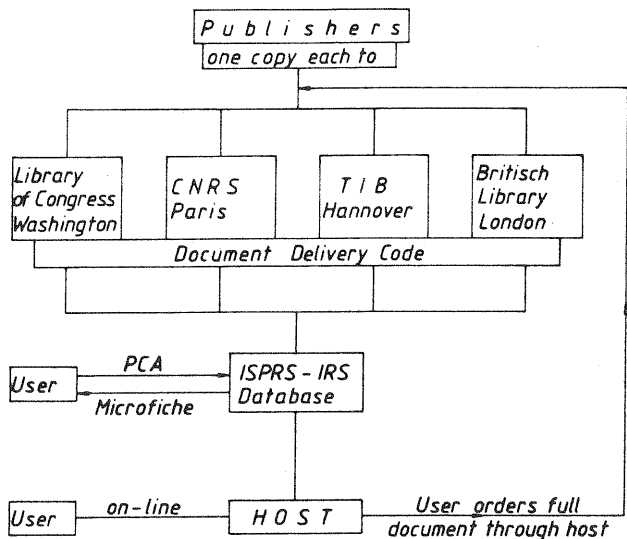Substantive support for ISPRS - IRS

The range of subject fields shall comprehend surveying & mapping including remote sensing. Some commentators proposed to consider imple- menting a separate database for remote sensing as this technology serves many scientific disciplines. Being aware that most results of remote sensing re- quire presentation in a thematic map, the outcome was that remote sensing may best be served within the ISPRS - IRS Database.

The content of the ISPRS - IRS Database shall include bibliographical data with summaries being subsumed from conventional and from grey literature, as well as factual data. Full document delivery can be provided according to the outline presented in figure 3.

Input selection is anticipated to prevent that users be submerged by too many computer outputs. The procedure is delineated in figure 4.
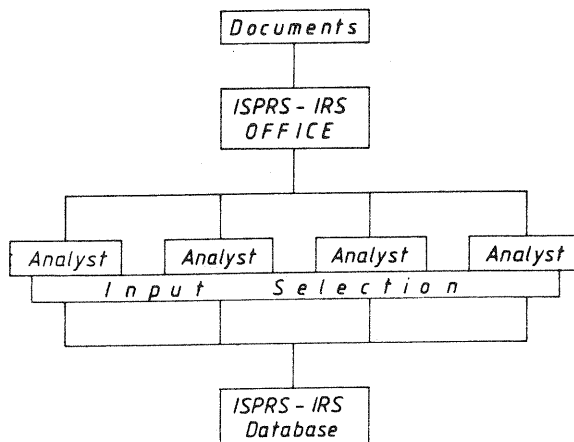
In pursuance with requests having been put forward at the ISPRS Symposium Mainz 1982, the powerful hosts ESA - IRS in Frascati - Italy, and STN Karlsruhe - FRG (including STN - USA and STN - Japan as well) agreed to negotiate on the grounds of the Specifications to host the ISPRS - IRS Database.

Any achievement of our International Society should be available to all Member Countries. Bearing in mind that most online retrieval requires payment in hard currency, and envisaging that quite some Member Countries, particularly de- veloping nations, can not afford spending hard currency for this purpose, ISPRS - IRS has been designed so as to allow a free-of-charge access via the 'Post - Card - Approach' (PCA), compare figure 3.

Full document delievery within ISPRS - IRS

Figure 3

Input selection to prevent users being submerged by too many computer outputs

Figure 4

There are, inter alia, two different modes for implementing and maintaining the ISPRS - IRS Database. The first and easiest mode would be to have it done by a database producer in the USA or Europe on commercial grounds. As an outcome, any user would have to pay in dollars. Further, this solution encounters that the database will possibly not be established by professionals in our field which is an inherent disadvantage.

Much better appears to be a second mode with financial support from one country for maintaining the database allowing both, online access to a host computer, and the free-of-charge PCA. Financial constraints of governments of many countries are, however, such that one country can hardly carry the burden alone. Therefore, a joint venture of European countries had been initiated to be channeled through the Commission of the European Communities (CEC). Being aware that other scientific disciplines would also approach governments to support information retrieval systems for their fields, our request to European countries via the CEC has been based on two reasonings :

+ In contrast to e.g. chemistry and medicine, being commercially structured resp. financially self supporting, a particular feature of surveying & mapping is that the majority of users of an information retrieval system are govern- ment employees who do not dispose on funds to utilize ISPRS - IRS. Consequently, implementing ISPRS - IRS on commercial grounds (first mode) would require altering the budgets of all government organizations engaged in surveying & mapping. This appears to be a tedious resp. unrealistic undertaking. Instead, it will be more efficient if five governments share expenses for the ISPRS - IRS Office.

+ Surveying & mapping, and particularly remote sensing, have quite an impact on developing countries. Hence, financing the ISPRS - IRS Office as a joint ven- ture will be a contribution for technical co-operation for development.

The structure of financial support is outlined in figure 5. A ISPRS - IRS Office carries responsibility for establishing and maintaining the database, and responds to PCA queries. Four Analysts are required to do the work. One country each is expected to carry expenses for one Analyst. A fifth country carries

Convener
initiates

Comission of the
European Communities
recommends

| Country 1 | Country 2 | Country 3 | Country 4 | Country 5 |

c a r r y   e x p e n s e s   f o r

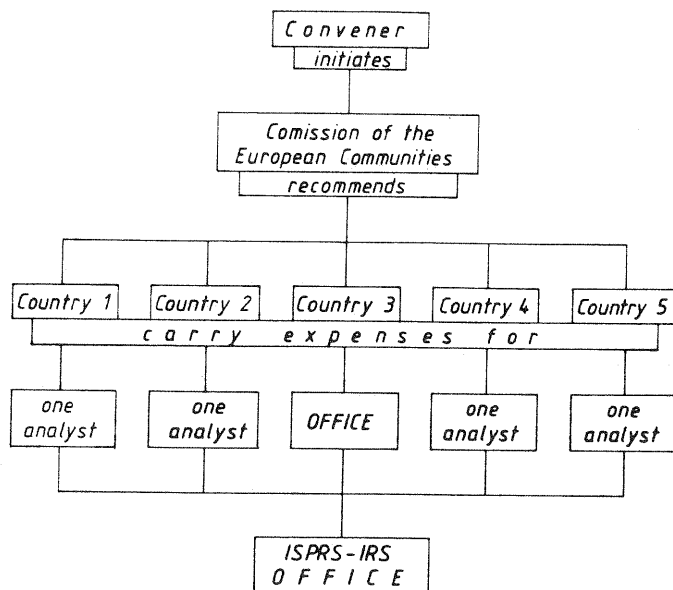| one analyst | one analyst | OFFICE | one analyst | one analyst |

ISPRS - IRS
O F F I C E

Figure 5

Financial support for ISPRS - IRS

expenses for office expenditures, hardware and software, and for a director.
All four Analysts are working in the ISPRS - IRS Office to be established in
the supporting country.

5 Outlook

Financing and operation of the ISPRS - IRS Database as a joint venture of
European countries did not materialize so far. Since establishing the ISPRS -
IRS concept 1984, four years have passed. An enormous advancement has been
achieved particularly in the online scenery during this time span. Databases
grew bigger; they may be more powerful now, yet they may as well submerge users.

After all, we need to know where we presently are. Therefore, it is feasible to
start after the ISPRS Congress Kyoto July 1988 an inquiry on the capabilities,
efficiency, reliability and costs of available bibliographic and factual infor-
mation.

References

Bohm, J.P.A., 1987. Value - added - services. In : Convention Proceedings
     INFOBASE, pp 12 - 16, Frankfurt - FRG

Clark, J.E., 1987. Accessing the resources of the National Technical Information
     Service. Photogrammetric Engineering & Remote Sensing, LIII / 4, pp 395-397

Cracknell, A. and Hayes, L., 1987. Remote sensing yearbook 1987, pp 160 - 498,
     Taylor & Francis, London - UK

Haken, H.ten, 1982. An investigation of available databases in the field of
     photogrammetry and remote sensing. In : International Archive of Photo-
     grammetry and Remote Sensing, 24 - VI, pp 82 - 91, Mainz - FRG

Hothmer, J., 1982. Scientific communication in surveying and mapping facing the
     year 2ooo. In : Proceedings of the Canadian Institute of Surveying Cen-
     tennial Meeting, Ottawa - Canada

Hothmer, J., 1983. Wissenschaftliche Kommunikation durch Zeitschriften und graue Literatur. Zeitschrift für Vermessungswesen, pp 234 - 242, Stuttgart - FRG

Hothmer, J., 1984. The information retrieval system ISPRS - IRS for literature and factual data. In : International Archive for Photogrammetry and Remote Sensing, 25 - A6, pp 155 - 169, Rio de Janeiro - Brazil
(The present paper is an update of this quoted paper)

Nowak, J., 1982. GEOFIZ, the information for geosciences, mining and water management. In : Ibid, 24 - VI, pp 38 - 42, Mainz - FRG

Peschel, H., 1983. Report of the permanent Commission VI 'International Bibliography Geodesy' . In : Proceedings of the General Assembly of the International Association of Geodesy, Hamburg - FRG

Tehnzen, J., 1982. Access to literature having been retrieved. In : International Archive of Photogrammetry and Remote Sensing, 24 - VI, pp 57 - 62, Mainz-FRG

Tehnzen, J., 1987. Von der Bibliothek zur Discothek ? ABI - Technik, 7/2, pp 87 - 94