Information needs and database use in the field of International Relations and Security[1]

Gerd HAGMEYER-GAVERUS[2]

Introduction

Research on a specific subject requires time-consuming material searches—for data, documents, journal articles, information acquired through interviews, and so on. Efficient research therefore needs not only support infrastructure in the form of libraries or similar institutions, but also a network of 'contacts'. Information exchange is a major factor in the research process. Contact networks are usually built up through the efforts of the individual researcher, but increasing effort is being put into institutionalizing informal networks by establishing cooperation agreements between institutes and by building cooperation structures.[3] Institutional cooperation is not new, but the developments in information technology give it a new dimension. The world of information technology allows for efficient work-sharing set-ups, where money can be saved and information access improved. The information revolution has in fact provided us with an overflow of information, available through different media. Getting the full text of a recent document, like the Dayton Peace Agreement for example, could in the 'old days' take days or even weeks, depending on the user's location. Today's access time is limited to hours after release. Most importantly, it no longer matters whether the user is located in Washington, DC or a small provincial town in Central Asia, as long as there is an Internet connection. Traditional methods of information retrieval and collection have been changed, to the good and to the bad. Information is rapidly accessible but there is too much of it and in many cases it is too non-specific. Most users therefore struggle with the problem of how to find important information and how to avoid spending hours looking for it. The International Relations and Security community therefore needs a home from where short paths lead to places of specific interest. This home must be set up through a joint effort and take into account the needs of the community. This chapter tries to identify the information needs and suggests an organizing structure for information production, storage and retrieval.

Scenario for organizing Internet information for our research

Researchers, librarians and students in the field of International Relations and Security wrestle with the problem of quickly finding and accessing information for their needs on the Internet. Unmanageable personal libraries of Internet links, lack of knowledge of updated information, the traffic jams on the Net and the ever-tedious search for relevant information are the bottlenecks. It's time to put the Internet child on its feet by organizing information production, processing, exchange and retrieval.

---

[1] This paper is a shorted version of an article to be published by UNIDIR later this year.

[2] Gerd Hagmeyer-Gaverus (hagmeyer-gaverus@sipri.se) is an economist working as Information Technology Manager and Researcher at the Stockholm International Peace Research Institute (SIPRI).

[3] One example is the European Working Group on Information and Documentation in the Fields of International Relations and Area Studies, a cooperative effort of European libraries in the field not only to achieve the exchange of information, but also to bring about work-sharing.

The need for subject-relevant database systems

Subject-relevant database systems are a major part of an efficient retrieval system in our research field. Database systems facilitate not only fast retrieval of information but also dynamic processing and presentation of information. In database systems existing data can be processed through combination and exclusion, through sorting and counting and then presenting the data as text, tables or graphs. Thus the database system has qualitative advantages over plain text or document retrieval systems. Our field of research has many applications for such database systems, where the subjects could be:

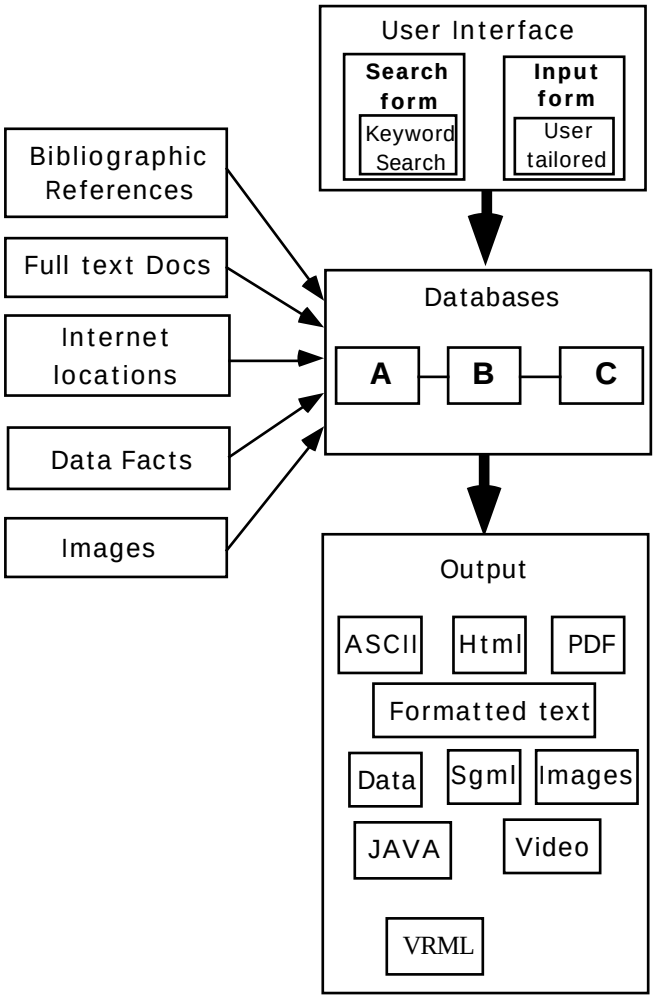
Facts and data:
- military expenditure figures by countries, economic data, arms trade, military manpower, military holdings data etc.
-data on conflicts
- data on weapons of mass destruction
Ideally, a database where a military country profile could be received would be most helpful.
- computer models
Bibliographic information:
- article references, articles-full-text, newspaper clippings, declarations, books, research reports, government reports grey literature,
Institutional information
- research institutes, general information
- research projects
- conference information
Information on individuals
- lists of experts

The other category of systems are those which do not keep the original information, as in these cases, but the Internet location, the links, to the original information data, documents, maps, sounds—whatever can be put in digital form, such as information produced and published by international organizations, governments, the media. These databases not only keep the location of specific documents at the site of the organisations and save time-consuming searches at these sites, but also keep data on the kind, category and 'value' of the information as well as on last updates. Intelligent informationcollecting tools—as described below—are used to obtain the entries to this database system.

While some efforts to create these database systems may be unilateral, efforts should also be multilateral in order to achieve a high value and efficiency and to save resources. One example of a cooperative information collection and dissemination effort on the Internet is a common database on journal articles.[4] Today, most institutes compile their own databases on journal articles. Library staff spend a lot of time collecting and indexing this information and then feeding it into a computer system. In many institutes the same information is processed, 'wasting' resources which could otherwise be used for additional tasks if institutes divided up subject areas or journal titles and fed into a

[4] In Germany institutes cooperate to feed into an existing common database, the World Affairs Online. In the framework of the European Working Group on on Information and Documentation in the Fields of International Relations and Area Studies parts of this database are, together with SIPRI's article database, made available on the Internet: http://www.sipri.se/SFgate.html.

common database. Ideally such a database would have links to the full text of the articles,[5] which are stored in another database or exist at different locations throughout the Internet.



Such databases should be the collective project of institutions and organizations in the field and could be in many different locations. The conference 'Institutes and the Security Dialogue', held in Zurich in April 1994, laid the foundation for such a collective institutional effort. One of its goals, to be of assistance and support to the new institutes evolving in the East, could be at least partly achieved through such an effort.

The advantage of the Internet is not only the on-line use of these kinds of database but also their on-line feed. In a work-sharing effort these databases can through specific user-tailored input forms be updated by select institutions, organizations and individuals. The input is secured through identification and password access and is organized through a list of input criteria. Reliability is the most important criterion for such a collection. However, published information, in whatever form, must meet a certain standard. It is not quantity but quality which must be the goal of such a project.

A new way to retrieve information

A researcher in the field needs user-friendly access to the information necessary for his work. He shouldn't have to deal with the technology of the system, only the result is

[5] Legal problems concerning copyright laws must be solved in this framework.

important. A search on a subject—by keyword or concept—should provide him with all the necessary information. Today's search on one of the global search engines often produces, as we have seen, thousands of 'hits' which are cumbersome to look through, because the indexing procedures of the engines are crude and produce keyword lists which are extraneous. It is still necessary to 'surf' the Net to find all the information needed. One needs to jump from one location to another, often getting lost on the way, even forgetting where you started. Since our institutes are looking for the same or at least similar information, we would collectively save time if there were a place on the Net where we could pose our search questions and get usable answers—one link instead of many links; one search instead of many; a retrieval machine which is focused on the subjects of our research field and does intelligent silent or background searches and produces a HTML result page with the links or sends an e-mail including the links; a retrieval machine which searches in several databases for references, full text articles and documents, video clips from news groups (for example CNN), the spoken word from sound archives, pictures, maps—in other words, all the elements of our multi-media world. How does it work?

1.) A user-friendly interface allows for input of keywords (with Boolean operators), citations, whole sentences, a glossary where several concept entries can be chosen of or a combination of all these. The better the search is defined for the user, the better the search result. Various input pages can be designed for that purpose.

2.) The input information is processed by a program locally or sent off to be processed remotely. The request is split up and sent off to different databases to collect answers to the requests.

3.) The databases process the requests and send additional requests to other databases with a flag to send back the results of the requests either as HTML or as e-mail, including links to HTMLs or full-text ASCII, attachments of documents, spreadsheets, sounds or video clips.

The basic element of such a system is the databases, as described above—databases which keep original information and data on locations of original information. These links must be either added manually to the database or collected automatically by using an intelligent system which filters and sorts relevant information. The most promising technology for this task is the so-called intelligent agent.[6] Intelligent agents are one of the fastest growing applications today. They are 'intelligent' software pieces to perform certain tasks on the computer or on a network like the Internet. These programs allow for an Internet search not only in static documents but also in databases or other application

---

[6] France Telecom has launched project Pyramide, which tests some of these agents mainly to book travel and to monitor stock exchanges; others use these agents to assist Internet shopping. However, their usability is still limited. For examples see the following URLs: http://home.mcom,com/comprod/mirror/client_download.htmle or http://genmag.com

For a general discussion see also: Ram Kumar:Internet Information Resource Discovery Tools: Current Status and Future Trends, Fujitsu Australia Software Technology Pty.Ltd.Australia;http://www.cs.umbc.edu/~cikm/iia/submiited/viewing/kumar2a.html

Keith Clark, APRIL: a language for building intelligent distributed information retrieval systems, Imperial College; http://www.cs.umbc.edu/~cikm/iia/submiited/viewing/clark.html

Mark Nissen, IntelligentAgents: A technology and Business Application Analysis, http://haas.berkeley.edu/~heimann/agents/

files. Intelligent agents ‘travel’ through the Internet and make, according to a set of rules, ‘independent judgements’ on the value of the information found on the Internet and return relevant information back to the user. They can be used for background searches or monitoring processes. Such background searches would feed the databases to keep information up-to-date at all times. They could also be invoked directly by a submitted search form. Another, more focused approach to feed into the databases is to index—on a regular basis—the predefined set of Internet servers.

Let us assume that all the servers in our research field are part of this effort, then all the documents of thess servers are indexed and the indexed documents are collected in a database. The search engine connected to this database would produce only relevant results. A program called Harvest[7] is now available to organize such a search routine Each site connected to the Harvest system indexes the information on its server and leaves it to the site to decide what information shall be included in the index. This means that irrelevant information such as published telephone lists, announcements etc. can be excluded from that index database. Specific databases on conferences or telephone lists can be created additionally.

Automated indexing would also replace the function of today’s list servers. While list servers provide a load of information on a subject area, most of the information is irrelevant. We still need to sort out irrelevant information and keep the rest. Automatic monitoring would free us from this extra work, and would regularly check the predefined Internet space for relevant information. However, no one has the resources to get all this information together in one place, in one large database, which is why a cooperative effort is needed. The Internet gives us the opportunity to put cooperation on a totally new level.

Result presentation

Today the Internet allows several formats for full-text presentation. One standard is the HTML[8] standard, another is PDF,[9] and both can be used. However, the simplest approach would be to present the articles as graphic images. The advantage is that any non-Latin character can be presented and these documents can be produced by a simple scan without a time-consuming OCR (Optical Character Recognition) process. The output can have several forms which partly are the choice of the user. They can be either pure text, standard Internet formats such as HTML, PDF or the like, or they can be standard word processing formats such as Microsoft Word or WordPerfect documents or images and sounds in Internet format.

Information will be directly displayed on the user’s screen or can be downloaded to be presented with the respective applications. The exchange of information in different languages, including those with different character sets like Cyrillic, Arabic, Chinese and Japanese, is possible as standard Internet image files.

New Internet tools allow for dynamic presentation of information. One of the latest is the Java script—a computer platform-independent script language which allows programs to

---

[7] http://harvest.cs.colorado.edu/

[8] HTML (Hypertext Mark-up Language) is a page description language for Internet documents.

[9] PDF (Portable Document Format) allows for the exchange of documents produced by any word processing or layout program, keeping the original format.

be sent over the Net and executed on any computer. Java and other new technologies like Shortwave allow for specific presentations of data. Java makes it possible to present data in a dynamic way—as spreadsheets, modifiable graphs, or animations. Datasets on International Relations and Security such as the international arms trade, military expenditure, size of arms forces and military holdings can thus be presented not only as static Internet pages but also in dynamic, more pedagogical form. Shortwave allows for multimedia presentations of subjects—e.g., presentations of UN peacekeepingoperations including video clips, pictures and animations, which are targeted mainly for the general public and the media. It also provides a unique possibility to present the institutes and their research results in a more 'popular' way.

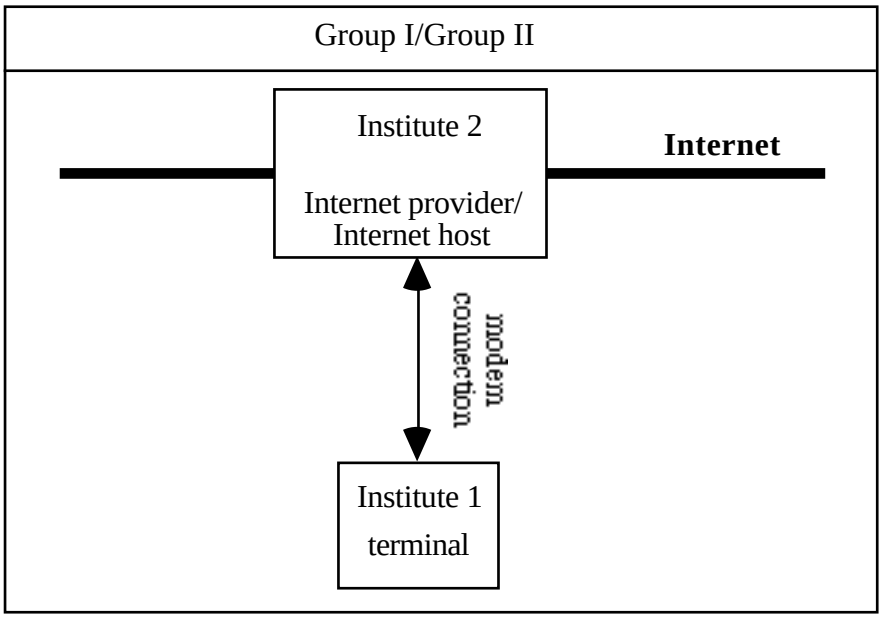
The infrastructure

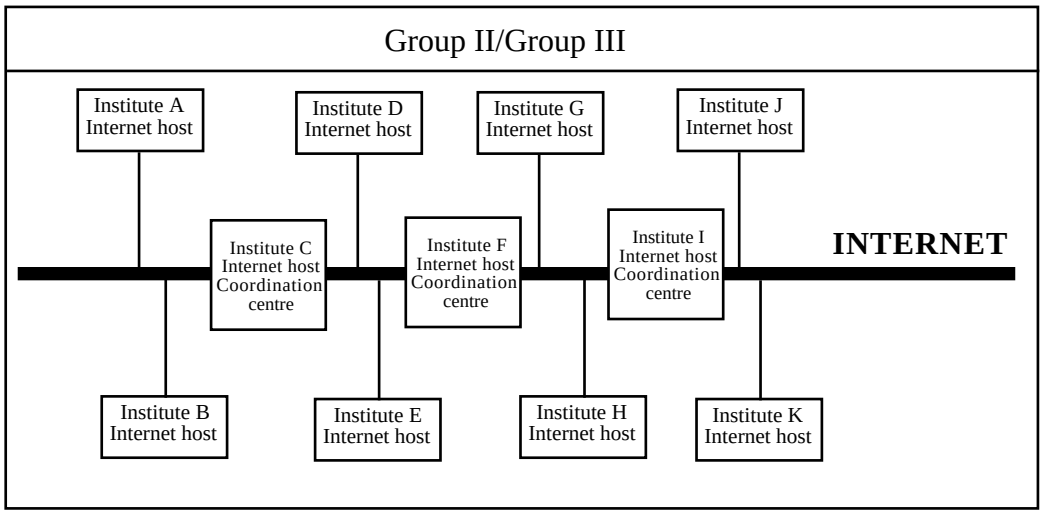
Creating an organizing structure for information exchange is the main task in our information collection and dissemination effort. Those with access to information in the form of printed documents, data, bibliographic references etc. should be encouraged to share this information with the rest of the community. Agreements need to be made on work-sharing in order to avoid duplication of effort, most obviously in the case of bibliographic information collection. A structure is needed for collecting, producing and processing information and to filter and redirect specific information to specific groups and users.

First we need to build a support structure for all those institutes which do not have the advanced technology for information processing (referred to as Group I institutes). General access to the Internet is inexpensive today: what is needed is a simple terminal-type computer, a browser and other communication software, a modem and a telephone connection to any Internet provider: this guarantees access to the full power of the Internet. New institutes, independent researchers and the like, with very limited financial resources, could afford this set-up or could be sponsored through a special program. For the cooperation effort they need additional scanning equipment in order to convert printed information into electronically usable format.

To establish electronic conferences, common databases and so forth it is necessary to set up a series of Internet servers. Since the Internet is discriminant on the location, these servers could be located at any institute or in any country. The only criterion is sufficient capacity for signal traffic, i.e., enough bandwidth in the communication lines and enough speed and storage capacity on the server machine.

The institutes with more sophisticated computer infrastructure would be such Internet servers, also called Internet hosts (Group II institutes), and should serve as support centres by publishing the information of Group I on the Internet and serving as mirror sites for other hosts. The form of publication can be in any standard Internet format or as an entry in a database system. The Group I institutes either log on to these computers by modem and input information via input forms direct and on-line into the respective databases or document locations or supply the information through e-mail or by electronic file transfers.





Group III institutes run servers which the meet high-level requirements for on-line communication like sufficient communication bandwidth and computing capacity for heavy traffic and which also have a specific set of software. Group III institutes serve as coordination centres and are the main target for users to search information.

A coordination centre with an Internet server would be in charge of maintaining databases, and the cooperating institutes together with the coordination centre would feed and update the databases. The detailed set-up and the division of work have to be decided at meetings.

The amount and coverage of the information included in such a set-up are almost unlimited. It is only a question of resources. As a research tool, however, it would be of enormous value. Any topic or subject could be part of such a set-up, but for greatest efficiency it should be established by means of cooperative efforts and implemented as shown in the figure above:

The coordination centres keep an update of all information kept in the network by regularly indexing the information stored in Group II servers. They serve as mirror sites for other network hosts in order to keep access times low. They also provide the necessary service for searches and for the use of intelligent agents. They regularly produce statistics on use and make regular evaluations of the system and make suggestions for improvements and for what additional information is needed (Statistics on searches by keywords and their 'hits' show where information is lacking to satisfy requests.)

Regular meetings of representatives of the institutes would discuss and decide the further development of such a network: additional information needed, work-sharing arrangements to be changed or established, etc.

Good management is important for a network to succeed. This would include setting goals according to user needs, developing a structure where all participants are actively involved, ensuring efficient implementation, and regular evaluation of the project/system/network.

Quality before quantity

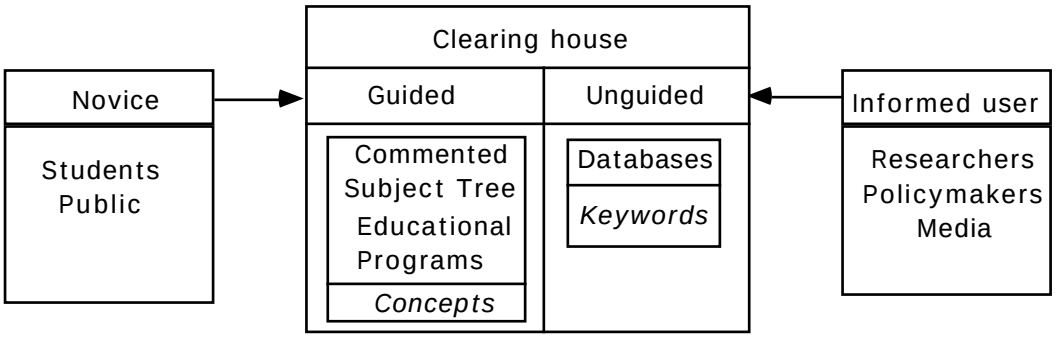
Let us say that I would like to know in a search the quality of the documents I have been presented with. So far, a keyword search provides me only with the information on of how many times the word or words I searched are represented in the document. No indication is given of how significant this document or its author is in the political debate. Whether an article on American foreign policy is written by Bill Clinton or Michael Jackson is never discriminated for. A sorting mechanism referring to a database on relevant authors (e.g., politicians in office or eminent scholars) could be one solution, but the decision regarding what type of author is significant is an important one, as is the value assigned to documents. Again, there are many alternatives: different databases can be accessed, average values of judgements can be used, etc.

What is to be included in the databases and Internet document collections must be evaluated. The research community must determine criteria and perhaps establish an editorial board on electronic publishing. Electronically published papers should go through the same vetting procedure as other published papers, but since electronic publishing is not subject to the same cost restrictions as traditional publishing, other papers could still be made available for interested readers, but should be marked somehow. A grading system could for example be established.

Other demands on information retrieval

Most experts in the field need short paths which lead to places of their specific interests. This short path is guaranteed by database systems, although the layman may get lost in information he cannot use or even judge its relevance. This is why a second and parallel gateway is needed. A clearing-house is needed to filter and redirect specific information to specific groups and users—a supportive effort to guide users quickly through the information. The clearing-house should help the uninformed user in a specific way and the informed user in another way. The less informed user needs a different approach to his or her question. He needs to be guided through the subject, he needs comments, he needs to be educated. The International Relations and Security Network (ISN)

homepage[10] at the Center for Security Studies of the ETH Zürich is a first step in this direction. On the other hand the informed user doesn't need guidance. He wants direct access to information. He knows exactly what he is looking for. This is why two approaches are proposed to this clearing-house function: one is for the uninformed user, the novice, the other for the user looking for specific information by searching in keyword-indexed databases.



What is the advantage of such a set-up and why would we like to join?

Efficiency may be increased through coordination efforts in the areas of both the retrieval of information and the production of information. Emphasis must, however, be put on the production of information, because qualitatively high-level information can only be produced and offered by the institutions dealing with the subject in question.

This is not an effort to organize the Internet—an impossible task—but an effort to organize ourselves. Our work has at least two limiting factors: time and money—and with both we have to be economical. Information retrieval costs time and sometimes also money, and information production costs time and money—at least in the form of additional human resources. Work-sharing cuts on spending, while an organized information structure in our field gives us more time. Each of us has to stop and recall how much time we spend to find the precise information we need for our research in order to properly judge how valuable such a system would be. A rough cost-benefit analysis at SIPRI shows that the net benefits of such a system would translate into savings of about $90 000 a year if retrieval and production time is cut by 50%, which I would regard as a minimum value. This means that I would find my information twice as fast and the library staff would have to feed only half the number of references into the database.

Whatever sophisticated search tool is available, the result is dependent on what information is available. One can only find those subjects that have been put there. This is where the efforts of institutes come in. We have to provide the information which is specific and make it available to others. Two scenarios are envisaged. We either a) make it available to the general public, meaning everyone, or b) make it available only to members of a network. The latter implies that we grant limited access to information we put on the Internet. The goal should be to make it generally available, but there are valid reasons for at least partially limiting access to certain groups only.

What about the costs?

[10] http://www.isn.ethz.ch/

For the most part, the technical infrastructure already exists. Some of the coordination centres may need hardware and software upgrades, which could be organized through sponsorships from the still blooming computer industry. The human resources that may be needed—especially for extra projects to scan documents—could be taken either from free capacities made available through work-sharing or by seeking funding for specific information processing projects. There are many important documents in our field which are still not electronically available.

In the future such efforts may even be self-funding. Once the problem of electronic money security and administration are solved, small user fees could be charged to non-members of the cooperation structure. If administrating and collecting such fees can be automated and the break-even point is very low, which means it is profitable to charge very low and affordable fees, and if use of such a system is sufficiently high, then it may become a self-runner. Thus only the initial costs have to be funded.

Conclusion

Mutually beneficial cooperation between researchers in the East and the developing world and those in the West should be promoted, primarily with a view to the new opportunities opened up by developments in communications technology. It is crucial that we learn to reap the advantages offered by these developments, that we are aware of the pitfalls, and that we adopt a generous, open approach to sharing information. The benefits are indeed mutual: there are distinct advantages to be gained from an exchange between the industrialized and the less industrialized world through institutional cooperation.

Use of existing technology and the Internet is a major task for the future. The daily business of International Relations and Security can draw great advantages. Making a common effort in efficient dissemination of information will not only provide users—whether they are researchers, politicians or media people—with their required working material, but also create bridges between institutions and countries, and can even serve as a kind of development aid. The outline presented here and the description of existing examples will hopefully lead to further development of such efforts. It is not to be seen as a final design, but as a first step towards a workable approach. Databases are now known to us, we know how to work with them. We are familiar with the Internet, too, but interfaces must be created which allow us to be much more efficient—not only interfaces to handle our computers better but also interfaces which help us to process ‘relevant’ information.

For the creation of an Information Network on International Relations Research together with SIPRI, experience can be drawn both from the work of the IntSecNet group and the European libraries cooperation efforts. The technology exists and is available at moderate costs. What is needed is a cooperative approach and an initiative.