

# The Librarian as Interface between Users and Electronic Information

© 1995 Ken Varnum

No Reproduction without Authorization

**About the author** : Ken Varnum is the Electronic Services Librarian at the Open Media Research Institute in Prague. He received his BA from Grinnell College in American history and Russian language (1989) and masters degrees from the University of Michigan's School of Information and Library Studies and Center for Russian and East European Studies (1994). He has been working at OMRI since it opened in January 1995. He can be reached at varnumk@omri.cz.

**ABSTRACT** : This paper explores some of the issues involved in creating digital collections, based on the author's experiences at the Open Media Research Institute, where he has been creating an electronic periodicals department for use by a diverse group of individuals and organizations. The creation of this electronic archive is presenting many opportunities for improved access to information, but also presents a number of challenges in terms of organizing, storing and providing access to it.

The Information Services Department (ISD) of the Open Media Research Institute (OMRI) is the custodian of the Radio Free Europe/Radio Liberty (RFE/RL) archives and library. In addition to these historical collections, OMRI is also developing and expanding the information resources available to its users by creating a large-scale collection of electronically stored periodicals and other information resources. OMRI is taking advantage of the Internet and other digital means to transfer information from providers to the library and from the library to users. OMRI is based in two locations in Prague and one in Budapest. In Prague, there are approximately 30 analysts as well as staff and students of the Central European University (CEU) who use the main library at the OMRI headquarters and an additional 270 broadcasters and news writers at

RFE/RL. In Budapest, there are many more users from the CEU's main campus in Budapest as well as researchers and scholars who use our materials in both cities.

Furthermore, the library and archives are themselves divided among the same three locations. The RFE/RL library, over which OMRI has custody through a unique public-private venture between the U.S. Board for International Broadcasting and the Open Society Institute, is split between Prague and Budapest. About 45,000 volumes are available in Prague, either in OMRI's main library (which has 30,000 volumes) or our branch in the RFE/RL building (where there are about 15,000 volumes). An additional 50,000 volumes are kept Budapest at the CEU-OMRI Archives. The RFE/RL archives have been similiary divided, with the most recent five years of materials stored in Prague and the remainder available in Budapest. As materials age, they will be shipped from Prague to Budapest. Electronic storage and transfer are valuable assets because of our geographically dispersed users and resources.

One of OMRI's strongest assets is its collection of periodicals, journals and monitoring of print and broadcast media. OMRI currently subscribes to more than 1000 newspapers and over 400 periodicals, most of them from the new countries of the former Soviet Union. Many of these, to our knowledge, are not available elsewhere in Central Europe. ISD also receives more than 700 pages a day of press clippings and transcribed television and radio broadcasts which we generally receive by fax from the capital cities of the region. These materials are supplied by indepenent contractors scattered around Eastern Europe and the former USSR. We receive an additional 250 pages per day of such materials via the Internet.

Although we do not have a large number of regular users, they nonetheless present an array of challenges in terms of how and when they need information. Timeliness has become a critical factor in ISD's information delivery, and is the driving force behind our emphasis on digital delivery and storage of information resources. Users from OMRI and RFE/RL have very different needs from those at the CEU and the general public. These needs can be separated into three groups, each with

different requirements for timely information delivery.

In the first group is OMRI's staff of analysts. OMRI publishes the OMRI Daily Digest, a 4500-word summary of news and events of Eastern Europe and the former Soviet Union. This publication is distributed to be available in the eastern United States as the business day starts there. This deadline means that the analysts must submit their stories to the editorial staff in Prague by mid-morning, which in turn means that ISD must provide them with information by about 9:00 AM at the latest or the information cannot be incorporated into that day's publication.

The second group of users, the news writers and broadcasters at RFE/RL, have a steady demand for new information from the region. Since each language services broadcasts at different hours throughout the day, each service has its own deadlines for news to be written. A steady flow of timely information is the key element for them, particularly because it is difficult to find secondary information resources from many of these remote countries. Broadcasters rely heavily, often exclusively, on the electronically delivered information OMRI provides them. If it is late, there frequently are not adequate resources to cover their region comprehensively.

The third user group consists of students and faculty from CEU and the users of OMRI's reference service (requests either in person or by phone, fax or e-mail). The requirements for this class of user are analogous to those of users of any reference service--information requests with varying degrees of time sensitivity and research needed to answer them. While questions posed by this group may require electronic resources to answer, there are generally other information resources that can also be used.

The focus at OMRI, then, is on rapid delivery of information to meet a variety of needs. Because of the scarcity of resource to meet these needs, the timeliness of electronic information becomes the driving factor behind OMRI's reliance on electronic information acquisition. Having set the stage, we will now turn to an exploration of the issues involved in creating and maintaining electronic periodical collections.

In the electronic realm, the acquisitions process involves identifying both user needs and appropriate electronic periodicals to meet those needs. As is evident from the number of pages transmitted to OMRI by faxed from all over Eastern Europe and the former Soviet Union every day, telecommunications costs are a significant portion of ISD's total budget. Given a choice, we prefer delivery via the Internet because it drastically reduces (and often eliminates) telecommunications charges. We encourage our suppliers to obtain access the Internet if they do not already have it because the increase in speed of delivery and decrease in telecommunications costs outweigh the costs associated with hooking up a supplier to the Internet. And as Internet connectivity becomes increasingly available throughout the former Eastern bloc (the annual rate of growth of Internet hosts in Eastern Europe in 1994 was the second highest in the world, at more than 130 percent), we have found it is increasingly possible to shift from fax to Internet as the primary means of delivery. OMRI currently receives electronic material from about 35 news sources in six languages from a dozen countries, a total of about 1 megabyte of information a day. We are still receiving faxed material from an additional dozen countries, although we are working on switching to Internet delivery in these cases, as well.

Once we have received the information, it must be prepared for storage in our "Electronic Archive". The files need to be decompressed and frequently must go through further conversions to make them readable. Much of the incoming monitoring is created on IBM computers, and OMRI and RFE/RL use Macintoshes. We have developed means of converting the incoming materials from IBM formats and IBM fonts. The issue of converting fonts is significant because we receive information in a number of languages with Cyrillic alphabets. Unlike plain ASCII text, which is a worldwide standard, there is not a universal Cyrillic font, even within a single computing environment such as Macintosh or IBM. We have often found it difficult to display files properly. Ukrainian monitoring, for example, is compiled in Kiev on an IBM clone and sent to us via the Internet, where we display it on a Macintosh. After having a special translation program written, we can now view it properly, but for several months it was impossible to do so.

Once whatever necessary conversions have been made and the files have been rendered readable, they are stored in the Electronic Archive, a file server to which users at OMRI have access. At present, there is no way for people at RFE/RL to have direct access to this file server, but that will soon change. In the meantime, some materials are sent via electronic mail directly to the individual users at the RFE/RL building as well as at the OMRI and RFE/RL offices in the United States who need it.

So far, so good. But there are still a number of challenges to face, problems to solve, and improvements to be made. User training is, of course, an on-going issue. Most of our regular users at OMRI and RFE/RL are familiar with computers, or at least are comfortable enough that, given clear instructions, they can access a specific directory on the file server, locate and display the file they need. However, CEU students and visiting researchers cannot be expected to have the same skills, and will have to be trained. A second training issue relates to the procedures involved in downloading and converting the incoming materials. This has been more difficult than expected and has involved writing a very detailed and constantly updated procedures manual. Hardly a week goes by that we do not subscribe to a new service requiring the development of a new procedure or that an information provider suddenly changes formats without notice. The important part is to ensure that users are not adversely effected when I am away from the office on business or on vacation, or when formats change. No matter what the source or the process, the files should be in the same spot and readable with the same software.

Another problem is the Internet itself. While Internet delivery is a large improvement over fax delivery, there are a few drawbacks. While Internet delivery is cheaper, it relies on the proper functioning of many connections between sender and recipient (over which neither party generally has any control). Although service is constantly improving, still more improvements are needed. Faxing, while significantly more expensive, tends to be also more reliable because it depends on a point-to-point telephone call which, while subject to a number of potential problems, is generally of a higher degree of reliability. Also, when a

phone call does not go through, it is instantly obvious to both sender and recipient. With the Internet, it is not clear if the message has been transmitted or not. These factors will undoubtedly become less important as time goes on and as Internet connectivity in the region improves, but they are to be considered in the short term.

ISD also must ensure that a variety of electronic connections are up and running properly. We rely heavily on a dedicated phone line leased from Czech Telecom for our connection to the Internet both for receiving and redistributing electronic information. We also rely on a Local Area Network within the building to allow users access to the file server. We have a dedicated VSAT link between Moscow and Prague that allows information to be transmitted much more reliably than the Internet connection we formerly used. In addition, There is also a radio wave link in place between the OMRI and RFE/RL buildings to speed delivery of materials to their offices. And, finally, there will be dedicated links between OMRI's Prague and Budapest facilities to allow research materials to be transmitted almost instantaneously from one site to another, whether for broadcaster, analyst or Central European University use.

All the technical issues of receiving information aside, probably the most important issue is the basic task of finding relevant information in the electronic archive. We currently have about a gigabyte of text information, an amount growing by about 1 megabyte a day. Unfortunately, we do not yet have a satisfactory searching tool even for the English language collection, but we are eagerly awaiting the arrival of a new software package that will allow Boolean searching of the text archive. The problems posed by our multi-language collection are proving more difficult to solve; we have not been able to identify a search tool that can perform free-text searches across alphabets and languages, or even within one family of Cyrillic fonts. This has proven an ongoing frustration, both for users and for the library staff.

The absence of good searching software will become an increasingly onerous problem over the next few years as we begin to rely more heavily on electronically delivered information and less on faxes through the

increased use of Optical Character Recognition (OCR) software. So far, reliable OCR software does not exist for many of the languages in which we are dealing--the accents and diacritical marks in Polish, Romanian and Czech give current software a very difficult time--but it is just a matter of time before it does. At that point, we will be faced with mountains of electronic materials we cannot search electronically, or at least, not with a single search program, which would, of course, be best. And then there is the language barrier itself--a searching tool that can search in both Russian and Albanian will need to be able to translate, at least at some basic level, from one language to another.

We hope to make all of our materials available to the public over the next few years, but we are also faced with a number of copyright issues that might prevent or limit access to certain parts of our collection. The *samizdat* collection, part of the RFE/RL archives, along with the book and periodical collections, can of course be made available for use. But the electronic resources present problems of copyright. Can we make this information available electronically for use beyond our three sites? Some material that we archive is available by subscription, just as a traditional magazine or journal. Unlike a paper publication, though, there is no way to prevent copies from being made of electronic materials. If they are displayed on a terminal somewhere, a copy has already been made. Creating a publicly-accessible library that includes this material would deprive publishers of potential subscribers, and will surely be prohibited. Other information resources are created expressly for us (much of the monitoring we receive falls into this category). While we pay to use this material in-house, a separate licencing arrangements will have to be made to allow us to further distribute these resources. We are faced with similar problems in our plans to digitize our collections of historic archives of newspapers and periodicals from the region, a collection which covers almost the entire Cold War period. We would like to create information resources based on this material, such as on CD-ROMs, but we do not hold copyright over much of it so our activities will be restricted.

And finally, there is the issue of making sure that these electronic resources are still useful in years to come. This means that the materials

must both be machine readable through future generations of Macintoshes and transfers to other platforms, and that the materials themselves are preserved. Frequent backups of the electronic archive are made to ensure against accidental erasure. But deletion is a short-term threat. Perhaps more damaging is the problem of guaranteeing future use of the materials over the coming years and decades. Since not all files are in plain ASCII text, but are saved in various word processing standards (especially files in various Cyrillic alphabets, which cannot be stored in an ASCII file and be readable in the original language), over the long term we will need to periodically transfer the data from one word processor to another to ensure that whatever the current word processing standard is, it can read the files. This issue is one that we librarians face along with archivists, who are discovering that much of the electronic data created in the 1970s now in archives is unreadable, either because the machines that stored it have been relegated to the landfill or the media upon which they were stored is no longer viable.

In conclusion, our library is facing a new range of issues in dealing with electronic publications. User training has so far been surprisingly simple, but this is likely because we created procedures designed explicitly to make the user end of the process as simple as possible. Staff training has been much more difficult because of the complexities of simplifying a variety of file types and creating programs. Much of the work in building an electronic collection has been in ensuring that the user--analyst, broadcaster, news writer or student--does not have to work too hard to get at the texts she needs. As we expand our electronic emphasis over the next few years to include, hopefully, all incoming information that is currently faxed, we will be faced with a number of challenges to keep the process simple for the end user. It is this, I believe, that is the central role in the library-patron interface in electronic records: ensuring that information is not hidden by the technology that creates and preserves it.