

Computerized support for research and publication in contemporary history

Databases at the 1956 Institute

The 1956 Institute was established in 1990 to research Hungarian history since the Second World War, including its international aspects. The primary subject of research was the Hungarian Revolution of 1956. This has steadily been extended back to the end of World War II and forward to the collapse of the socialist system.

Importance was attached from the outset to keeping a record of the documents encountered during the research. Data on the various source materials and literature were stored for later use by Institute staff or others, using the information systems available at the time. This was done in such a way that the database could be kept continually up to date and was sufficiently flexible to allow for changes of criteria or approach or the relinquishment of earlier criteria.

These requirements were considered while the databases for bibliography, video materials and oral history interviews were compiled. The document descriptions went well beyond the bounds of plain bibliographical data. With each type of document, related persons, institutions and events were recorded in a structured form, as 'related records' containing a biography, the date and place of an event, the persons concerned in it, and so on. Care was taken that either broad or detailed searches could be performed, for any of the recorded data or any combination of it, or for any words appearing in the data. Output for a required purpose can be prepared by defining various listing and printing formats. As for the structure of the databases, they could be operated either with the DOS operation system or in a network (Novell). However, this system was not suitable for storing or handling long texts such as oral history interviews.

The database for oral history interviews

The Oral History Archive at the 1956 Institute contains about a 1000 life interviews. A little more than half of these are interviews made with people who took part in the 1956 Revolution. The rest, labelled 'leader interviews', were made with people who later held leading party, state and economic positions under the Kádár regime. The database so far has processed the 1956 interviews, by making abstracts of 5–10 pages that can be incorporated into the

by Zoltán Lux*

database structure. Apart from the technical and administrative data (interviewer, length, accessibility *etc.*), three main groups of data have been recorded: 1. A description of the interviewee's life before the revolution (previous life). 2. A description of the interviewee's life since the revolution (subsequent life). 3. The interviewee's

activity during the revolution (events). The first two groups of data consist essentially of free text, with a maximum length of 4000 characters, in which it is possible to search for expressions marked and of course standardized when the text was introduced into the database. The 1956 events—activities during the revolution—consist of finely structured records (geographical location, exact location, institutions, participants, date, time, side-events *etc.*) These allow a very accurate search for memories of specific events.

Trial records

As research continued, demand for incorporating other types of document into the database arose. About 13,000 trials took place in the country during the period of reprisals that followed the 1956 Revolution. These concerned more than 25,000 people, including a high proportion of the participants in previous and subsequent political and intellectual life. The trial records provide a great deal of biographical data on those prosecuted (origin, schooling, wealth and so on). Requisite analysis of a well-structured database allows otherwise impossible examinations to be made of participants in the revolution and characteristics of the reprisals. (For instance, what social groups and strata were worst hit? How were foreign observers able to keep track of the sentencing in protracted trials?) The first to be put into the database were the trial records of those who were executed and the data of those who were sentenced to death but the sentence was never carried out. At present, the database contains data on more than two thousand people concerned in the '56 trials.

Chronology

The events are of interest to historians not just in the context of some document, but as distinct chronological events. So the databases have been compiled from the events as well as the documents. In other words, the events were the records incorporated into the database when a

period or subject was processed, and links were made, in many cases, to existing records of persons, books, articles, photographs and so on.

The photographic database

Various audio-visual items—photographs, sound documents and films—appeared by chance with the documents. Apart from their undeniable valuable as source materials, there is strong outside demand for these as illustrations for publications of various kinds. Initially, only descriptions of these were incorporated into the database. Systematic processing and digitalization began about three years ago, under various cooperative agreements, along with the establishment of the appropriate links to events and biographies.

Establishment of a new database

The rapid development of the World Wide Web was perhaps the main instigating factor when the institute began to rethink its procedures around 1995 for archiving data and develop an adaptable digital archive to meet the changing requirements.

The institute had several databases at that time. In some cases there were structural problems. The records of the same item document found in the different databases were not given in a uniform way.

Some of the tentative requirements of the new system were the following:

- * Link all the associated matters found into the single database.
- * Store and provide search facilities for long text documents as well.
- * Include the audio-visual documents.
- * Allow very fine tuning of the accessibility of the database and entitlements to use it.
- * Allow some of the data to be public, or even available on the Web.
- * Enable statistics to be produced from the database without the user seeing any specific records (to protect personal rights).

With these and many other, similar requirements in mind, the institute began in 1996 to develop a database based on Oracle software and transfer the data to it. It was found during these operations that there was no established practice for describing, examining or storing the various kinds of document. The basic concept adopted was to choose the structure that allowed the broadest description to be made, in which some of the data need not be given.

Data did not simply reach the database by being entered, within the institute. With the bibliographical data on books and articles, existing descriptions were imported using MARC format, before institute staff subjected them to further processing (linking them with events and persons, for instance). In the absence of a standard, the same procedure was adopted with some descriptions of photographs as well. Cooperation with other institutes and archives meant that data could be entered once, jointly. The sources for compiling them could be provided jointly, rather than the institute having to buy the data. (For instance, some of the photo documentation in the database is being developed further in conjunction with the Budapest Archives, in other words, part of it is shared.)

The database operated and compiled in this way assists in the work of the 1956 Institute in two main ways: database support for research and database support for publication.

Database support for research

Researchers can have access to data entered by institute staff. The database can be used in the same way by everyone, rather than being dependent on special recording procedures. Unfortunately, there is a long way to go before the database entirely replaces the card indexes. Not every researcher possesses a research notebook computer, and some of the data unearthed is still jealously guarded by researchers. However, I think this will change. The stocks of libraries and archives are being digitized at an increasing rate, so that even if the whole is not available, some description of each item can be found on computer, and often in a restricted form on the Web as well. So the decisive factor in finding the right document increasingly becomes the ability to compose good search questions, rather than chance. Furthermore, the 1956 Institute is among several places developing intelligent software that is able to search for intelligent information on local databases and on the Internet. Then it can deliver this information to potential users based on user behaviour and the questions that have been formulated. This forms part of the knowledge-processing procedures for the social sciences (which will not spread so fast, of course, as they have in the commercial sphere). This is the direction pointed by the demand by researchers that the data (from which certain conclusions can be drawn) should be fully accessible, so that statements based on them can be verified, or altered as subsequent information emerges. However, this conflicts with the requirement that research findings and source data should be kept secret.

The Internet allows researchers to make use of more than one database. Cooperating institutions can compile databases jointly or open their databases to each others' researchers. These days particularly, the discovery of successive new potential sources of data must be expected. I am thinking here, for instance, of the potential role of corporate archives or of the opening of party and state-

security archives in the former socialist countries. The archives of the Gauck Office in Germany and the Bureau of History in Hungary, for example, are being rapidly explored.

The second way the new database helps researchers is by giving support for publication activities.

Database support for publication activities

The findings of the research done at the 1956 Institute appear in a variety of publications. Even with a printed publication, it is a great help in preparing the chronology or bibliography, for instance, if the draft text can be compiled from searches in a database that always represents the most up-to-date situation. This applies still more if each item has to be arranged from several points of view.

Meanwhile several new media have appeared in the information society. These, in my view, do not replace the old media (books and films), but for certain purposes, a CD-ROM or a Web site capable of presenting audio-visual information may be more appropriate than printed publications. This is the case, for instance, with encyclopaedic publications containing large volumes of textual data, in which database handlers are the only means of making a rapid, detailed search. It also applies to works that set out to present a period of history in the most comprehensive possible way (including the arts, way of life, historical characters and so on).

It is important to prepare publications of this kind. Computers and Web usage are part of everyday life for the generation growing up today. We cannot forego the opportunity to convey cultural values and scientific findings through the media they understand best.

It is also imperative for scientific findings to reach the various levels of education. Students need textbooks and teachers need teaching aids. Research institutions also have a responsibility to ensure that information can be transferred rapidly. It is important to have rapid access to the source of authentic, up-to-date information, or to the source with the broadest knowledge of how to obtain it.

The database at the 1956 Institute serves as the basis for all its publication methods and objectives. It lies behind the various publications, including the Internet series on contemporary Hungarian history since 1945, aimed especially at secondary-school students, and the associated, encyclopaedic CD-ROM series for researchers. (The second disc is to appear in 2000 and covers the 1945–56 period.)

The part of the database containing the chronology and the photographic and textual documents will be made accessible by degrees to a limited extent. Rather than remaining a closed database, it will gradually develop and

alter. The photographic database includes TIFF format photo files for press use, which are not freely accessible on the Internet. The aim in the longer term is to cover some of the mounting expense of maintaining and developing the system by charging fees for press use of the photo files. This will call for developments in electronic trading that the institute eagerly awaits.

* Paper presented at the IASSIST Conference, June 9, 2000, Northwestern University, Evanston, Chigago.

Zoltán Lux, luxz@helka.iif.hu, Institute for History of the 1956 Hungarian Revolution, H-1074 Budapest, Dohány u. 74. Tel.:36-1-322-5228 Fax: 36-1-322-3084. Internet: www.rev.hu

