

New Challenges : digital documents in the Library of the Friedrich-Ebert-Foundation, Bonn

Rüdiger Zimmermann / Walter Wimmer

**„Archives of the Present“ : from traditional to digital documents. Sources
for a history of contemporary social movements.**

02. – 03. dicembre 2004
Milano Sala Conferenze
Archivio di Stato di Milano
Via Senato 10

Basic situation

Today I would like to talk about the activities of the Library of the Friedrich Ebert Foundation concerning the integration of digital documents into the Library's range of services. My report is partly based on a script of our system administrator Walter Wimmer who is therefore mentioned as author as well.

The Library of the Friedrich Ebert Foundation (<http://library.fes.de>) is one of the large special libraries in Western Europe in the field of labour movement history. The library is rooted in the tradition of the great library of German Social Democracy, although its holdings were scattered during the time of Fascism and the Second World War. Many of you may know that the SPD library kept the complete private libraries of Karl Marx and Friedrich Engels. After the war the SPD built up a new library which in 1969 was handed over to the Friedrich Ebert Foundation, a non-profit organisation.

Later on the DGB as umbrella organisation of the German trade unions handed over its library holdings to the FES Library, as well as many single trade unions, international trade union organisations and organisations of the workers cultural movement. Nowadays the library comprises a total of 700.000 volumes.

Together with the Archive of Social Democracy the library forms a research centre with more than 70 employees.

The library collects printed documents (newspapers, journals, brochures, books) while the archive collects organisational files, personal estates, photographs, leaflets and posters.

In Germany librarianship has a federalist and decentralised structure. Many libraries co-ordinate their collecting activities according to a clearly defined plan, the controlling authority is the Deutsche Forschungsgemeinschaft (German Research Foundation) that provides additional financial support.

Our library is supported for building up its collection of publications of political parties and trade unions in Germany and Western Europe. These collecting activities comprise conventional as well as digital documents.

Since the middle of the nineties a growing number of publications of trade unions and political parties is published digitally. Many publications appear in printed as well as in digital form. But the number of publications that are only available electronically grows steadily.

As an additional challenge the German Research Foundation demanded that sponsored libraries should engage themselves in the field of retrospective digitisation of primary sources in order to re-publish them on the Internet.

Consequently, our library had to cope with problems in terms of content, copyright law and technical difficulties.

Problems in terms of content

Today, we as librarians hold the opinion that our tasks in the contemporary digital age do not differ fundamentally from those in the "paper age". Our collecting activities comprise conventional as well as digital documents and both types of documents are indexed according

to almost the same standards. And we attempt to integrate the digital documents into the existing structures of German librarianship.

The library collects brochures, press services and journals in digital format. The archive files complete Internet pages. This means that it is possible to browse through the library's digital journals as if they were conventional documents. This is not possible with the archive's Internet pages though.

The retrospective digitisation of documents

The electronic conversion of important documents for scientific research in Germany was supported by a special financial support programme. The systematic conversion of conventional documents into digital documents is a central aspect of our library's activities.

To meet this challenge we followed certain principles:

- The documents must be of special significance for academic research.
- They must be rare.
- The online presentation should lead to a certain added value.
- All questions concerning copyright law must be verified.

Within the Library of the Friedrich-Ebert-Foundation we thought quite a long time about the questions which documents we could recommend for the national project of retrospective digitisation. The chance to be financially supported attracted us, of course. After thorough consideration we decided on the Social-Democratic Press Service in the years 1946 to 1995 (<http://library.fes.de/cgi-bin/populo/spdpd.pl>) and Press releases and information of the SPD („Pressemitteilungen und Informationen der SPD“) from 1958 to 1998 under different titles (<http://library.fes.de/cgi-bin/populo/spde.pl>).

In the past there were two decisive facts that imposed a restriction on the use of these excellent sources: On the one hand no library in Germany possesses a complete copy, the Library of the Friedrich Ebert Foundation is the only one with an almost complete copy. On the other hand the documents are only inadequately indexed and catalogued. Neither a weekly nor monthly nor an overall index exists. Thus the scientific utility value is extremely limited.

These deficits have been erased completely through the retrospective digitisation. The only copy that remained preserved has been microfilmed by the Library of the Friedrich Ebert Foundation in accordance with the corresponding DIN-rules. With the help of an external service provider (the Dutch company DMP) the complete microfilm has been retrospectively digitised as „digital images“. This means an electronic copy has been made available in the World Wide Web. This service alone though would not have improved the user-friendliness.

Therefore all parts of a single article that describe its content have been „transcribed“ anew, stored in a database and finally linked with the electronic copy. Moreover, a librarian has revised this information with special emphasis on the explanation of abbreviations and pseudonyms. Besides this, personal names have been replenished with extra added biographic information.

On the screen the early years of the Press Service don't look particularly attractive at first sight. But the original printed version isn't really any better. And a document printed out at

home with a standard printer is of better quality than a photocopy from the original document. The text is completely readable.

Together both services add up to a total of 160.000 political full-text documents. They form an outstanding source for post-war history of the German Social Democratic Party and the German labour movement.

For me the presentation of a rather conventional project was important because I think that the problems of retrospective digitisation do belong to this conference's subjects as well.

The collection of current Internet sources

With the onset of the Internet age also the Library of the Friedrich Ebert Foundation began to consider how to handle monographs and periodical documents published on the Internet, in order to save them and provide permanent access for scientific research. We have also been on some wrong tracks in this issue, I do not want to go into our mistakes or the technical details though.

We regularly save Internet-based monographs and periodicals on our library's server, especially when it is obvious that these documents are not permanently saved. We know for example that Internet server are completely shut down on the occasion of trade union mergers. In other cases where we think that the documents are indeed permanently safe only the URL is catalogued, which is especially true for periodicals. The German librarianship's infrastructure with a central database for periodicals and another database for electronic periodicals that is been set up at the moment supports our work here.

We are not allowed to re-publish documents on the Internet that have been downloaded and saved by the library. Both the European and the German copyright law make this almost impossible.

I would like to demonstrate the difficulties by the example of periodicals of international trade secretariats. Although they allowed us to download their periodicals and present them on the Internet again we were unable to do this due to the pictures in these periodicals. Normally, the institutions acquire the publication rights on the pictures from certain agencies for a single purpose. To present a downloaded copy of a document on the Internet again means to re-publish this document. And for this re-publishing the agencies would have to be paid again a lot of money.

Which solutions did we find?

We catalogue the sources according to the cataloguing rules and so produce entries in our online catalogue. As long as there is a stable connection with the Internet server of the particular organisation the catalogue user can access the original online document. If the organisation removes the document from the original Internet server the user has the opportunity to order the publication online as we still possess the locally archived copy of this document. This procedure might be a borderline case as far as copyright law is concerned, so far nobody complained though. Let me demonstrate this procedure by the example of the title "Teamwork and older workers". The cataloguing of Internet resources certainly requires a reorganisation of our library's conventional work routines, as all the relevant servers have to be searched and analysed under the principle of job sharing.

The database of electronic press releases of political parties and trade unions (<http://library.fes.de/cgi-bin/populo/presen.pl>)

The difficulties caused by copyright law made us consider which form and procedure we could choose to extensively save Internet-based resources of political parties and trade unions, and re-publish them on the Internet as collections. Finally, we decided to archive press releases of political parties and trade unions. This decision was influenced by the philosophy of the International Institute of Social History in Amsterdam.

Archiving of press releases – a sensible addition to conventional collecting activities (<http://library.fes.de/cgi-bin/populo/presen.pl>)

The decision to collect press releases of political parties and trade unions systematically is based on different ideas:

- As press releases are explicitly intended to be openly and widely published there are no copyright problems in the redistribution on the Internet. These press releases usually do not contain any pictures that – differently to written information – often cause copyright problems.
- Press releases are naturally predestined to be offered on the Internet. Therefore many organisations choose to publish theirs this way.
- Several archives of press releases of single organisations relevant for the project can be found on the Internet, but there are no cumulative archives that contain press releases from different organisations. Now it is possible to compare directly different opinions and statements which offers many advantages to the user of this service.

Conceptual consideration on the indexing of the archived materials

The most obvious idea concerning the indexing of the archived documents is the use of a fulltext retrieval programme. The project does offer this, but only in addition to the search with meta data or – speaking in librarian terms – catalogue records.

In certain cases fulltext retrieval is not always sufficient, especially in those cases in which the search is restricted to the finding of documents of a certain period. Fulltext retrieval systems that stay within the library's financial limit do not offer this.

So there was the idea to produce meta data of the archived documents in order to take advantage of the library's efficient cataloguing of locally saved press releases is done with the help of programmes written in Perl. This method is based on the fact that nearly all press releases on the WWW are presented in the form of lists. In many cases these are dynamically developed HTML pages that are generated by database inquiries.

This kind of presentation demands a HTML source code that is also evenly structured. On this basis it is possible to write Perl programmes that extract the kind of information from the source code which is used to produce catalogue entries automatically. Once the information is filtered it can be transferred into a structural format and administrated in a database. Again Perl programmes are used in this process. The database software used in this project is the library's cataloguing and database programme Allegro. But it would also be possible to modify the produced data for administration in relational database management systems like

Oracle. These automatically generated data sets are transferred into special databases by their specific update routines.

The programming that forms the basis of the realisations is done in the library. It is a great disadvantage that for each press release a new subroutine is necessary. This subroutine has to be adjusted to the layout of each single press release. If the layout changes the subroutine also has to be modified.

This preliminary work is the basis for the archiving which is done by a student assistant.

The technical aspects of this concept were presented by my colleague Walter Wimmer on the occasion of the 2001 IALHI conference in Tampere. Therefore I would like to skip these details. (<http://library.fes.de/fulltext/bibliothek/01103.htm>)

By the example of the keyword “Cancun” I would like to demonstrate the different possibilities for a database enquiry.

Fulltext retrieval as a supplementary option for searching

The disadvantages of fulltext retrieval have already been mentioned. Nevertheless, it makes sense to also offer this kind of search. At first it was planned to extract all meaningful words from the archived press releases in order to combine them with the generated records in our library database Allegro. Unfortunately the programming of the batch files necessary for this approach took too much time.

In the present version of the project the fulltext retrieval was implemented as an additional search option. For the indexing of the fulltexts the programme Htdig is used, which is also used in the Occasio project run by the International Institute for Social History. Htdig is a Unix based freeware programme. Within this project Htdig runs (as well as the Allegro database offered on the Internet) with Linux.

For example, a database enquiry with the keyword “overtime” results in approximately 400 hits. These hits are arranged according to their relevance. Of course there are full-text documents among the hits that do not contain “overtime” in their titles. This kind of retrieval though does not allow more specific restrictions for queries (for example documents published by political parties). Nevertheless, our users are satisfied with this service.

At the moment we have saved approximately 150000 full-texts, all of them comfortably accessible by full-text retrieval. So far documents of 100 organisations have been collected. Theoretically, this number could be increased considerably.

At the beginning of the twenty-first century libraries need sophisticated strategies to save digital sources and to provide access to these sources in the future. Digital sources belong to the collective memory of mankind. But: for every institution it is impossible to handle all the problems alone. Co-operation is essential and must not fail despite the existence of national borders.

Perhaps this conference is a milestone on our way to a successful and productive co-operation.