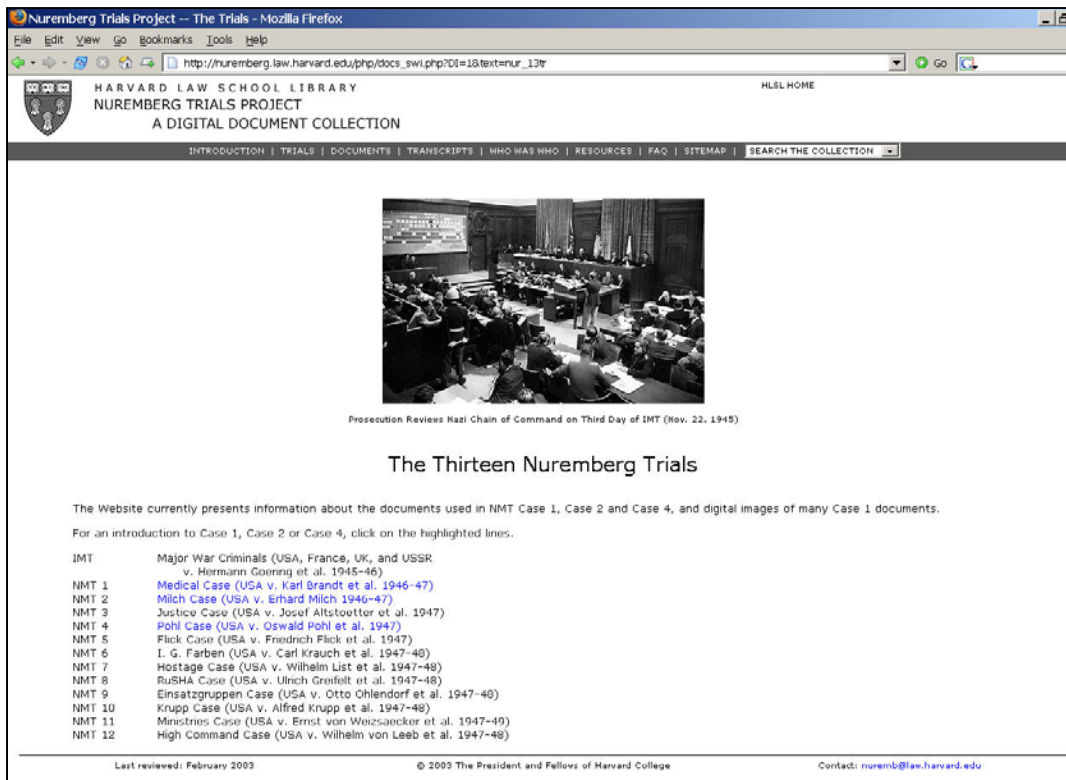


Harvard University Library
Library Digital Initiative Grant

Final Report

Harvard Law School Library

Nuremberg Trials Project:
Completion of Case 1



The screenshot shows a web browser window with the URL http://nuremberg.law.harvard.edu/php/docs_swi.php?DI=18;text=nur_13tr. The page header includes the Harvard Law School Library logo and the text "HARVARD LAW SCHOOL LIBRARY NUREMBERG TRIALS PROJECT A DIGITAL DOCUMENT COLLECTION". A navigation menu contains links for "INTRODUCTION", "TRIALS", "DOCUMENTS", "TRANSCRIPTS", "WHO WAS WHO", "RESOURCES", "FAQ", "SITEMAP", and "SEARCH THE COLLECTION". The main content area features a photograph of the trial proceedings with the caption "Prosecution Reviews Nazi Chain of Command on Third Day of IMT (Nov. 22, 1945)". Below the photo is the heading "The Thirteen Nuremberg Trials". A paragraph states: "The Website currently presents information about the documents used in NMT Case 1, Case 2 and Case 4, and digital images of many Case 1 documents. For an introduction to Case 1, Case 2 or Case 4, click on the highlighted lines." A list of cases follows, with Case 1, Case 2, and Case 4 highlighted in blue. The footer contains the text "Last reviewed: February 2003", "© 2003 The President and Fellows of Harvard College", and "Contact: nuremb@law.harvard.edu".

IMT Major War Criminals (USA, France, UK, and USSR v. Hermann Goering et al. 1945-46)

NMT 1 [Medical Case \(USA v. Karl Brandt et al. 1946-47\)](#)

NMT 2 [Milch Case \(USA v. Erhard Milch 1946-47\)](#)

NMT 3 Justice Case (USA v. Josef Altstoetter et al. 1947)

NMT 4 [Pohl Case \(USA v. Oswald Pohl et al. 1947\)](#)

NMT 5 Flick Case (USA v. Friedrich Flick et al. 1947)

NMT 6 I. G. Farben (USA v. Carl Krauch et al. 1947-48)

NMT 7 Hostage Case (USA v. Wilhelm List et al. 1947-48)

NMT 8 RuSHA Case (USA v. Ulrich Grafelt et al. 1947-48)

NMT 9 Einsatzgruppen Case (USA v. Otto Ohlendorf et al. 1947-48)

NMT 10 Krupp Case (USA v. Alfred Krupp et al. 1947-48)

NMT 11 Ministries Case (USA v. Ernst von Weizsaecker et al. 1947-49)

NMT 12 High Command Case (USA v. Wilhelm von Leeb et al. 1947-48)

Last reviewed: February 2003 © 2003 The President and Fellows of Harvard College Contact: nuremb@law.harvard.edu

Submitted by Paul Deschner
Harvard Law School Library

February 17, 2006

About the Collection

The Harvard Law School Library possesses approximately 1,000,000 pages of documents in both German and English relating to the trial of military and political leaders of Nazi Germany before the International Military Tribunal (IMT) and to the twelve trials of other accused war criminals before the United States Nuremberg Military Tribunals (NMT) immediately following WWII.

The documents, which include trial transcripts, briefs, document books, evidence files, and other papers, have been studied by lawyers, scholars, and other researchers in the areas of history, ethics, genocide, and war crimes, and are of particular interest to officials and students of current international tribunals involving war crimes and crimes against humanity.

To preserve the contents of these documents -- which are now too fragile to be handled -- and to provide expanded access to this material, the Library has begun a digital project to create and present the English-language versions of these documents -- some 650,000 pages -- on the Internet, along with analytical information about each document, general information about the trials, and search tools offering focused access into these materials. Due to the massive scope of the project, and the fact that the large funding amounts required must come from outside donors, the work has only been able to proceed in modest and carefully targeted stages, as funds have become available.

The first stage of the project was envisioned as a pilot to work out all the technical, infrastructural and procedural requirements for the project as a whole by completing all the processing of a single trial's materials -- the so-called "Medical Case," which was Case 1 of the NMT trials. This would allow us to more accurately determine the staffing, funding and timeline requirements as well as feasibility for completing the full project while also offering a coherent and completed resource for researchers of the trial selected for the pilot. Equally important, the completed trial would also serve as a means to graphically present the project to prospective sources of outside funding.

This first stage began in the late 1990's and reached its first milestone with the mounting of the Nuremberg Trials project website in July 2003, offering users the full array of data and display formats and searching capabilities we had envisioned for the full project. Though the materials presented were representative enough of Case 1 to allow a user to sample the site's full functionality, their scope fell far short of being comprehensive for that trial, with just under half of the 13,904 pages of evidentiary material and little more than five per cent of the 11,538 pages of the trial transcript online.

All materials are accessed through the Nuremberg Trials Project website at <http://nuremberg.law.harvard.edu>. By means of a web browser, the online user can view page images and corresponding descriptive analyses of evidentiary documents and conduct both fielded and keyword searching of these descriptive analyses. The user also is given access to the trial transcript for the trial in question, which serves as a roadmap to the trial's evidentiary documents, introducing them in the context of the prosecution's or

defense's case at that point in the trial. Full-text searching is available for the trial transcripts. All of these materials are cross-linked in such a way that a user can retrieve the images of an evidentiary document, accompanying descriptive analysis and transcript citations of the document regardless of which one of these is selected as an entry point.

To create the digital files of evidentiary documents, project staff assemble documents into batches of several hundred pages and deposit them with the Harvard College Library Digital Imaging Group (DIG) where these are digitized as image files in both archival (TIF) and delivery (JPG) formats and deposited into the Harvard Office of Information Systems Digital Repository Service (DRS) for storage, from which copies of delivery-format images are subsequently retrieved by project staff and mounted on the Nuremberg web server. To create the full-text transcript, project staff outsource keying-in to a third-party vendor, then proofread and edit the text themselves before posting the text on the Nuremberg website.

Once the evidentiary materials have been digitized, they are reinserted into their appropriate archival boxes and returned offsite to the Harvard Depository. The transcripts are stored in their own boxes together with the evidentiary materials. Due to the fragile condition of these materials, they have been withdrawn from public access.

The Nuremberg website is hosted on a server maintained by the Harvard Law Library's Automation Department. The website itself has been developed and is maintained by project staff at the Harvard Law School Library. Project documentation -- including backgrounders on the Trials and their central figures, how to use the website's search engines, links to related Web resources -- has also been mounted on the website.



German online news article about Nuremberg website launch.

When the website launched, the Library distributed announcements to a number of major listservs dealing with domestic and international law, as well as to Harvard-internal news

sources. The response by the mainline press both here and in Western Europe was immediate and very supportive, with numerous online articles introducing the project and its website appearing in the weeks that followed. This also generated in some cases lively discussions in online forums about issues raised by the materials presented on the website. During this immediate post-launch period, the project was also profiled by New England regional and German TV news and by local public radio, including interviews with the Director of the Library and project staff.

Currently, the website logs 200-400 web visitors per weekday (somewhat fewer on weekends), with activity occasionally spiking to as much as 2,000 visits per day in conjunction with various external events, such as Nuremberg-related conferences, online media coverage of the website's launch, etc. Project e-mail queries and weblogs indicate that users from within the U.S. and Western Europe comprise the largest group of website visitors, including a large number of researchers or students at academic institutions.

After dozens of e-mails to the project's contact address in the six months following the launch, the site now attracts about three messages a week. These are almost always from descendants of Nuremberg participants requesting information about their relatives' role in the Trials and from students desiring information for secondary school or university projects.

The project has attracted contributions of research items to the Library relating to Nuremberg. For example, Nuremberg prosecutor Drexel Sprecher has donated his files relating to the Trials, and photographs of the I.G. Farben Trials were contributed by prosecutor Belle M. Zeck.

About the LDI Grant Project

We were anxious to be able to finish processing of the full Case 1 trial and applied to LDI for funding to complete digitization of the evidentiary materials and staff a one-year editing position to complete proofreading and editing of the trial transcript. The project was awarded and funded in April 2004, the project staff (editor) began in September 2004 and worked through August 2005. Work was completed in March 2005, while the budget was closed out in November 2005 after the final HCL-DIG invoice.

Case 1: The Medical Case

The Medical Case (U.S.A. v. Karl Brandt et al., also known as the Doctors' Trial) was held in 1946-1947 and involved 23 defendants accused of organizing and participating in war crimes and crimes against humanity in the form of harmful or fatal medical experiments and other medical procedures inflicted on both civilians and prisoners of war.

The materials which we have digitized for mounting on the Nuremberg website are of two types: evidentiary materials, which have been imaged by the Harvard Library Digital Imaging Group, and the trial transcript, which has been re-keyed, proofread and edited.

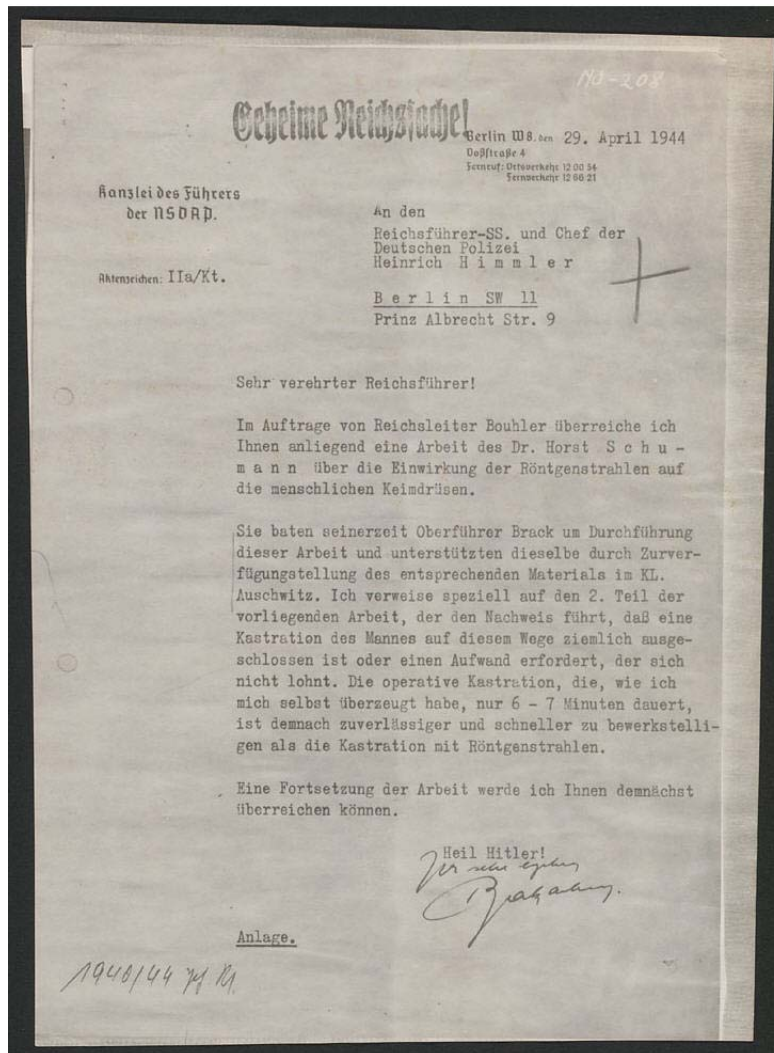
Case 1 Evidentiary Materials

The evidence documents consist of two types: "trial documents" submitted in evidence by either prosecution or defense during the trial, and "evidence file documents" from which the trial documents were drawn. The evidence file documents were compiled by the Allied war crimes staff at the end of the war from German government ministries, private papers, and other sources and were placed at the disposal of the Nuremberg legal staffs to use as evidence during the course of any of the thirteen Nuremberg tribunals. These files typically came in up to four versions: a photostatic copy of the German original together with German and English typescripts, and a staff evidence analysis prepared by the war crimes staff. Most of the prosecution exhibits and some defense exhibits were drawn from these documents, the prosecution often simply making a copy of the English typescript and submitting that copy into evidence.

Since it is often the case that the Library's collection of trial documents for any given trial is incomplete, the evidence file collections have proven invaluable as stand-ins for the missing originals. Our evidence file collections have proven to be sufficiently complete to allow presentation of either the original evidence file document from which the trial document was copied, or failing this, some other version closely related to it. For example, if the original was the English typescript and that is missing, our collection would usually have the staff analysis, the German typescript or a photostat of the original.

For this reason, it was decided to digitize our full collections of both document types for all trials. Our Case 1 trial document collection, which is roughly 75% complete, has been made functionally complete through the availability of the evidence file documents from which the Case 1 trial exhibits were derived. Of the approximately 14,000 evidentiary document pages analyzed for Case 1, roughly half are trial exhibits and half evidence file materials. But it is also of great benefit to have available evidence file documents relating to a given trial document even when the trial document is present, since the evidence files offer further perspective, usually by offering the original German-language document or a staff analysis of the content presented in the trial document.

It was decided to digitize these documents as images rather than as full text, since the mimeographed typescript originals are often of sufficiently poor quality to allow reasonably efficient optical character recognition processing, while the other materials -- photostats of handwritten material and some photographs -- can not be adapted at all to this form of processing.



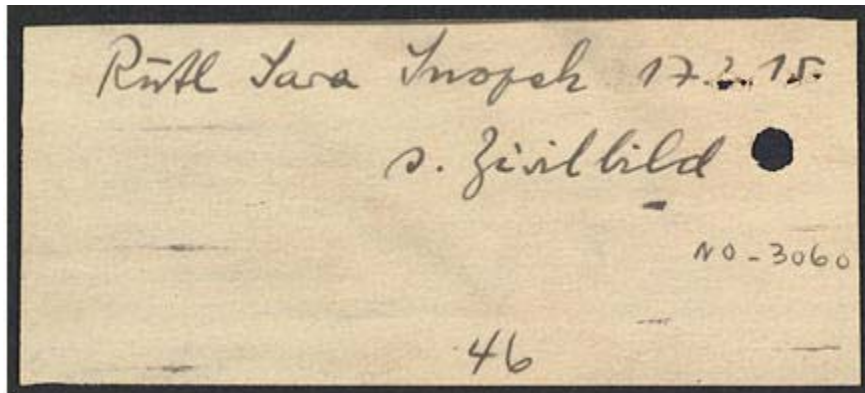
Digitized evidence file document NO-208.

Before commencing the LDI-funded completion project, Case 1 digitization of evidentiary materials was roughly half completed, with 6,755 out of a total of 13,904 images processed.

The LDI-funded one-year editorial staff position was to spend roughly 10% of his time with assembling of batches of these documents for transmittal to Harvard Library's Digital Imaging Group (DIG) for processing and, upon completion, with retrieval and recollection of the documents into their storage boxes. With 7,149 pages to be digitized distributed over 17 batches, this worked out to about 420 pages per batch from November 2004 through June 2005.

At DIG, original document pages were digitally photographed, producing an archival TIFF as well as a derivative delivery JPG at 100 dpi of each page for online presentation.

When digitization for a given batch was complete, DIG electronically transmitted both types of image objects for each document page to the servers of the Digital Repository Service (DRS) at OIS, together with the structural metadata for each document which we had supplied. We then extracted the delivery images from DRS using the DRS Web Administrative Interface tool and mounted the images on our Nuremberg website.



Digitized photograph and associated note of Buchenwald inmate Ruth Sara Snopak.

The number of images remaining to digitize for Case 1 was difficult to estimate, since without actually compiling the batches, we could use only rough estimates based on round figures maintained by our document analyst when analyzing the documents in an earlier phase of work. While we had originally estimated that in the Case 1 completion project we would need to digitize 6,245 document pages, the final number was 7,149 document pages, which, at DIG's processing fee of \$4.50 per document page, amounted to a cost increase of \$4,068 over against the projected cost of this component of the project. In addition, not anticipated in our cost projection was the presence in the materials scheduled for digitization of around 20 large-format items, such as charts, which required special processing and was charged at the rate of \$75.00 per item.

Case 1 Trial Transcript

It was decided to have the 11,538-page transcript keyed in to allow machine-readable processes to manipulate the text. As with the mimeographed evidentiary materials, the transcript also consists of mimeographed typescripts of sometimes very poor quality, so automated processing was not an option and the text was sent in microfilm form to China where it was keyed-in by hand. Full-text functionality is important for the transcript, since as a "roadmap" to the nine months of trial proceedings it is important for a researcher to be able to search for the context in which trial exhibits were introduced and be able to link out of the transcript to the images and descriptive analyses of the documents. To enable fielded searching into the transcript, it was also decided to have the keyers mark up the text with XML according to a template we delivered to them.

After the keyed-in transcript was returned to us, we discovered that extensive editing was required to correct numerous errors that had been introduced at every stage of production of both the original as well as the keyed-in versions. The keyers had been sent over ten reels of microfilm from the National Archives and Records Administration, corresponding to the 23-volume paper set the Library possessed. After arrival in China, the microfilm was segmented and then distributed for keying in.

By far the main source of textual corruption occurred during the process of re-keying. First, we discovered that roughly 130 pages at over 20 locations in the original transcript had been entirely omitted in the keyed-in version, most likely corresponding to microfilm segments which had been lost during the keying-in procedure. Second, there were a number of places in the keyed-in transcript where the text on entire pages had been truncated before the right margin, with whole words or parts of words missing as lines abruptly ended. Third, there were the inevitable typographical errors introduced by the process of manual keying in. Fourth, and most likely due to the fact that the keyers were non-native or non-fluent in English, there were numerous word-recognition errors in the transcript, particularly at places where the original words were difficult to make out, with some constituent letters so faded as to be impossible to interpret without an understanding of the context.

```
<TP>The statements of Professor Schreiber about the experiments co??  
ducted in my institute at Posen are not true. I have already said that in  
January 1945, when the Russians came to Posen, the institute was not eve??  
finished; that at this time there were about 300 workers still working th??  
that nothing was ever done at this institute as far as research is concer??  
and no airplane experiments were conducted at Posen. During my whole per??  
in office not a single such airplane experiment was carried out, not eve??  
other areas. The experiment which Schreiber mentions was one done before ??  
was given my assignment, which Professor Kliewe once told me about, but ??  
was not for an offensive purpose and the result of the experiment was ab??lutely  
negative.</TP>
```

Example of truncation by keyers at right-hand margin of transcript text.

There were also different types of errors introduced into the original transcript during the original process of production. This process became complex particularly when the language of the proceedings switched from English to German, typically when a defense counsel or a witness spoke in German. During these passages, the original German spoken at trial was rendered on the fly by a court interpreter into oral English, which a court stenographer with headphones then recorded in handwritten format. This was then later deciphered and recorded by a typist, mimeographed and eventually bound into the volumes in the Library's archive. It is easy to see by comparing the German typescript at these points that the interpreter occasionally misunderstood the context or specific words of the original (which were often technical medical terms) and produced a rendering which is incomprehensible or misleading in the final printed version. We decided in these cases to supplement the original text with editorial comments.

There were also numerous errors concerning evidence file document code numbers which were introduced into the original text. One common example resulted from the misinterpretation of that class of evidence file documents belonging to the "NO" archive in such a way that when, for example, evidence file "NO-641" is introduced into evidence, the original trial typescript erroneously refers to "Document No. 641" -- mistaking the archive prefix for the abbreviation "no." for "number." Since in the online transcript we link from the mention of such a document to the digitized images and analytical description of it, it is important to correct these errors so that proper matching can take place.

A third area of errors was generated by inconsistencies in the spelling of personal and place names in the text.

The editing workflow divided into an initial proofreading and subsequent editing phase. Proofreading was done on a paper print-out of the digital keyed-in files (ASCII plain text), and focused on identifying clear typographical errors and those errors based on a clear misunderstanding of the text by the keyers. Additionally, the proofreading phase tried to resolve the various lacunae within sentences and words, where the keyers could not make out the words clearly, typically indicated by them with a double question mark ("??") in the text. Related to this was the keying in from whole cloth those gaps in the transcript which resulted from pages having been missed by the keyers in China, around 130 pages. Any questions whose resolution depended on a deeper knowledge of the text or required further research were to be annotated and left to the attention of the editing phase. All annotations were done in pencil on the paper print-outs.

Before the Case 1 completion phase, we had proofread 9,214 of the transcript's total of 11,538 pages. The LDI-funded position completed the proofreading of the remaining 2,414 pages before proceeding with the more demanding editorial phase. The initial proofreading pass also served as an excellent introduction to the text and the sorts of problems which were to be encountered later in the editing phase. This proofreading work was to take up approximately 20% of the total hours allotted to the position.

After this phase was completed, the editing phase began. This work divided into three main parts: verifying the correct code numbers of evidence file documents and trial document exhibit numbers, resolving the deeper textual problems which had surfaced during the proofreading phase, and inputting into the original keyed-in electronic ASCII plain text files all corrections and editorial comments as well as some extra XML markup we had decided would be useful.

The nature of the first two sorts of problems has been sketched above. As evidence file documents were introduced into evidence in the trial and were assigned exhibit numbers, they were identified by codes in the transcript which we needed to verify in each case to allow proper linkage to our digitized document page images and the descriptive document metadata accompanying them, which are identified by the same codes.

Likewise, it was during the editing phase that all errors, inconsistencies and confusions introduced into the text by the process of its original production needed to be resolved. It was a boon to the project that our editor was conversant in German and could resolve numerous textual problems by consulting the corresponding passages in the German Case 1 transcript. He also made extensive use of the Web, background reading and some of the online resources available on our Nuremberg website as well as ever deepening familiarity with the transcript text to resolve the other problems. Finally, he input the corrections and editorial comments into the electronic texts which were to be presented on the Nuremberg website. He also supplemented the pre-existing XML markup with a few additional tags to aid in document linkage, page formatting and pagination.

Before commencing the Case 1 completion phase we had edited 500 of the 11,538 transcript pages, and had hoped to complete editing of the full text by the end of the editor's one-year appointment. But due to problems which had not surfaced before, such as numerous gaps of pages which the keyers had entirely missed, as well as our editor's unexpected competence in German, which allowed an area of textual enhancement not possible with our previous editor, this proved optimistic. Our editor completed editing of 3,409 additional pages of transcript by the end of his appointment, which brought our total to 3,909 pages.

```

during the experiments -- also shared this impression?</TP></Sp>
<Sp>
<TP><Spkr>A:</Spkr> Yes.</TP></Sp>
</ProceedingText>
</ExaminationSection>
</WitnessSection>
<ProceedingText>1
<Sp>
<TP><Spkr>THE PRESIDENT:</Spkr> Is there any further cross examination of the witness? If not, the witness will be
excused.</TP></Sp>
<Sp>
<TP><Spkr>MR. McHANEY:</Spkr> If the Tribunal, please, we had concluded yesterday's session with the introduction of
Document <DocNum>1609-PS</DocNum>, as Prosecution Exhibit <ExhibitNum N="92">92</ExhibitNum><HN N="2541"/>, which is on
Page 77 of the English Document Book; and the Court will recall that that was a letter dated 24 October 1942 from
Heinrich Himmler to Rascher, acknowledging receipt of several letters and stating that he had read his report regarding
cooling experiments on humans and that Sievers should arrange for the evaluation
<PB N="309"/>
<RunningHead>12 Dec-M-HM-10-3-Perrin<LB/>
Court No. 1</RunningHead>
of the results of these experiments at institutes which are connected with the SS; and he had further stated that he had
regarded those people as traitors who still today reject these experiments on humans and would instead let sturdy German
soldiers die as a result of these cooling methods. We now go on to Document <DocNum>[NO] 401</DocNum> which will be
Prosecution Exhibit <ExhibitNum N="93">93</ExhibitNum><HN N="1687"/>; and this, if the Tribunal, please, is the report on
the meeting in Nurnberg on the 26 and 27 of October 1942, on medical problems arising from distress at sea and winter
hardship. This is the report on the meeting which your Honors have heard discussed at some length this morning. On the
bottom of the first page you will see that the meeting was sponsored by the Inspector of Medical Service of the
Luftwaffe, and that the chairman of the conference was Stabsarzt Professor Dr. A.J. Anthony; and the Tribunal will recall
that Dr. Anthony was chief of the department for Aviation Medicine in the medical service of the Luftwaffe, and that the
defendant Becker-Freyseng was his assistant in that department from the latter part of 1942 until early in 1944, when the

```

Example of XML markup in transcript text.

Project Budget Summary

Description	Budgeted Amount	Actual Expenditure
Project Editor: Salary	\$XXXX	\$XXXX
Project Editor: Benefits	XXXX	XXXX
Training by Consultant	0	312.50
Digitizing Vendor	28,102.50	33,580.50
Totals	\$XXXX	\$XXXX

Salary and fringe benefits were subject to University increases, and therefore the projections were too low.

Consulting services covered training provided by the former project document analyst to the incoming transcript editor, introducing him to the proofreading and editing work as well as batching procedures for digitization of evidentiary materials and remaining available to answer any questions which arose during the course of the work.

The increased costs of digitization correspond to the additional cost of digitally photographing 904 more document pages than originally anticipated, as well as the much higher per-page cost charged for the processing of 20 large-format items (such as charts)

found within the evidentiary documents -- \$75 per page as opposed to the normal-format charge of \$5 per page. As mentioned previously, it was difficult prior to actually assembling the documents to be digitized to predict precisely the number involved, as well as the possible presence among them of oversized pieces.

Appendix I

Editing a Transcript Day

I. Preliminary steps:

- (i) Have at hand the proofread, keyed text of that day's proceedings.
- (ii) Have at hand the HLSL hard copy of that day's proceedings.
- (iii) Check to see if the proofreader has attached any memos to the front page of the keyed text. If there is a memo indicating missing days, you will need to key in those days. (See **A1** below)
- (iv) Open the day to be worked on in UltraEdit; the files can be found in the NurembergTranscripts file folder under "Edited dailies". Do not use the files in the "Original dailies" folder.

II. Editing:

- (i) Check document and exhibit number citations
 - (a) Open the website to the "Document Search" page; (See **A2** below)
 - (b) In UltraEdit, search that day's proceedings for "DocNum";
 - (c) Double check every document number found against the website and, where appropriate, add the HLSL item number after the exhibit number. (See **A3** below)
 - (d) Run a search on "ExhibitNum" to see if any exhibits were mentioned by themselves. Again search the website and add the HLSL item number where appropriate. (See **A4** below)
- (ii) Automated Corrections
 - (a) Replace <Spkr> with <TP><Spkr> (See **A5** below)
 - (b) Replace </Spkr><TP> with </Spkr># (i.e., add a space)
 - (c) Replace <Spkr>Q.</Spkr> with <Spkr>Q:</Spkr> (see **A6** below)
 - (d) Replace <Spkr>Q</Spkr> with <Spkr>Q:</Spkr>
 - (e) Replace <Spkr>A.</Spkr> with <Spkr>A:</Spkr>
 - (f) Replace <Spkr>A</Spkr> with <Spkr>A:</Spkr>
 - (g) Replace &emdash; with — (see **A7** below)
 - (h) Replace #&emdash;# with — as well as #— and —# with —
 - (i) Replace #-# and #-#-# with —
 - (j) Search for —, checking each instance to make sure that it is properly situated in the text and that it is not misplaced for –. Change to – when necessary. (See **A8** below)
 - (k) Other automatic corrections can be made on a day by day basis. For example, if you find a trial day that involves a regular appearance by or discussion of Herta Oberheuser, it is more convenient to do a replace on

Oberhauser (for it is often misspelled this way) with Oberheuser at the beginning.

- (iii) Incorporate the proofreader's corrections. (See **A9** below)
- (iv) Run spell check. (See **A10** below)

A1

Keying in Missing Text

For the purpose of making the editing process easier, it's important to preserve the formatting of the paper text. So, you should insert hard line breaks (hitting the "Enter" key) whenever the line in the paper text breaks (do not insert the line-break tag -- explained below -- just use the "Enter" key to insert the break here)

Tagging for Page Number and Running Head

1. `<PB N="page number"/>` -- identifies current transcript page; bring up page number as given at bottom of page and include here within the "`<PB ... />`" tag at the very top of the page (do not duplicate at bottom of page)

2. `<RunningHead> ... </RunningHead>` -- identifies trial date, stenographer's initials in parentheses and typist's name; should immediately follow "`<PB ... />`" tag

3. `<LB/>` -- identifies a line break; this is usually the case between the line containing the information above in line 2 and the following line identifying the tribunal no. -- in this case, always "Court No. 1" or "Court 1"

4. example:

```
<PB N="448"/>
```

```
<RunningHead>16 Dec-M-GES-1-1- Daniels (MJ)<LB/>
```

Court 1</RunningHead>

Tagging for Speaker and Speaker's Segment

1. <sp> ... </sp> -- identifies speaker's segment; don't close off with "</sp>" till the speaker finishes his segment

2. <tp> ... </tp> -- identifies a paragraph in the text

3. <spkr> ... </spkr> -- identifies the speaker name given in the text; this can be a person's name or simply "Q" (for "question") or "A" (for "answer")

4. examples:

<Sp><TP><Spkr>MR. HARDY:</Spkr> May it please the Tribunal, our headphones are dead. We couldn't hear the last two or three questions very well. I suggest that we discontinue until the phones are put into condition.</TP></Sp>

<Sp><TP><Spkr>Q:</Spkr> Witness, you spoke about a Luftwaffe Institute in Dachau. I am now asking you did you see some kind of a placard or a letterhead where such a name could be seen?</TP></Sp>

Keyboard Shortcuts for Tags

<PB ... />	Ctrl + P	page number
<RunningHead>	Ctrl + R	running head open
</RunningHead>	Alt + R	running head close
<LB/>	Ctrl + B	line break
<tp>	Ctrl + T	paragraph open
</tp>	Alt + T	paragraph close
<spkr>	Ctrl + S	speaker open

</spkr>	Alt + S	speaker close
<sp>	Ctrl + G	speaker segment open
</sp>	Alt + G	speaker segment close

A2

Prior to editing, you should spend sometime becoming acquainted with the website. It is the most valuable resource you will have in the editing process. Searching on the website can be done in a variety of ways. In the case of the documents, you will most likely search for a specific document or group of documents via the document search engine. The search engine leads to a page that displays the document analysis information for the document(s) searched and provides links to the digitized images of those documents. Through the document search engine, documents can be searched by author, date, literal and descriptive titles, evidence code number, trial date, and transcript page number.

A3

Evidence codes: Evidence file documents (NO, PS, NI etc.) will be searched via the tags for DocNum (document number). It is important to have both the letters and the numbers appear within the tags (<DocNum>NO-41</DocNum>), but often the letters appear outside, especially if NO has been typed as “No.” If an NO document has been typed as “No.,” we have inserted NO between the tags and deleted the No. outside the tags (even though No. might have been intended to mean “number” -- this is something we can’t know for sure and have to interpret in a way that clarifies the important information, i.e., the code number). So, “I’d like to introduce document No. <DocNum>41</DocNum>,” becomes “I’d like to introduce document <DocNum>NO-41</DocNum>.” Other document types, such as PS documents, also need to follow the formula for NO documents. You will notice that PS documents are found in the text with the PS printed after the close document number tag, such as “<DocNum>1398</DocNum>PS”. This

still needs to be corrected to “<DocNum>PS-1398</DocNum>”. Such corrections, however, pertain only to prosecution documents. Defense documents, such as “Brandt No. <DocNum>7</DocNum>”, should contain only the number between the tags (this means that abbreviations, such as Br. or HA etc., will sometimes have to be moved outside of the tags). It is quite easy from context to tell whether a document is a prosecution or a defense document.

Verifying Prosecution Documents and Exhibits

Let’s say you find “<DocNum>NO-440</DocNum>,” in the transcript. To verify this document, you can go to the “Document Search – Evidence Code Number” page on the website, search under the NO symbol for number 440. Usually the document will appear and this will be sufficient verification. If the document does not appear and the search finds “no hits”, then you can search either by the author or by the date of the document. If you find what is unquestionably the document under discussion in the transcript with a number that reasonably could have been meant in the transcript, then correct the number in the transcript. So, if you cannot find NO-440 by searching the evidence code, but do find a document that matches the date, author, and content of the document in question, but is NO-1440, that evidence would be sufficient to change the 440 to 1440. If, however, you cannot find 440, and can find no other document matching the other criteria, then leave the number as is.

Exhibit numbers can change during the trial, and the HLSL collection is also missing some exhibits (even though we have the evidence files pertaining to those exhibits). If you find the document on the website and it exists as a Case File with a Prosecution exhibit number, and if the same exhibit number appears in the transcript, you can then place the HLSL item number after the close exhibit number tag. (See **A4** below) If there is no case file for a particular number, but the exhibit number is listed in the transcript, simply leave it as is. If (what is very rare) there is a case file indicated on the website and the evidence code numbers match but the exhibit numbers differ, correct the exhibit

number in the transcript to match the number stated on the website, but bracket the corrected number.

Verifying Defense Documents and Exhibits

Unlike prosecution documents, defense documents are not automatically searchable by virtue of their code number. Although they often bear a code (BR for Brandt exhibits, HA for Handloser), these codes are, apropos of search-functionality, superfluous. The documents are set apart by the same tagging -- i.e., `<DocNum>...</DocNum>` -- but in the case of a defense document, only the number should appear in between the tags. So, if BR or HA or some other set of initials has been placed by the keyers inside the tags, move it to the outside. It is also important to remember not to exchange the “No.” that appears prior to defense documents for an “NO” within the document number tags.

As is the case with prosecution exhibits, defense exhibits may be followed by the HLSL item number. Finding the correct item number involves searching the website. A helpful tip: if there is a witness under consideration and defense counsel is introducing his exhibits seriatim, you can speed the process along by doing a keyword search for the defendant in question (be sure to set the results per page to list 50 at a time). Scroll down until you find the first defendant document. The documents should then follow in order, with a few pauses here and there. This is an especially helpful procedure when dealing with counsel who present the defense documents but fail to name the exhibit numbers. There is no way to search defense documents on the website, so it becomes a matter of merely tracking them down. If the defendant’s counsel does not provide an exhibit number, but you are able to find the document nonetheless, you may then supply the HLSL item number after the close document tag. Again, just as with prosecution exhibits, if you find the document you are seeking and it is clear that the number listed in the keyed text is incorrect, replace the incorrect number and enclose the correct number in brackets.

Final Notes on Documents

Document books are often mentioned, and as often the keyers placed document tags around the book number. These are superfluous and can be deleted. So, in the phrase “document book number <DocNum>5</DocNum>,” the tags should be removed. In addition, there are often cases when you will have to add tags. Counsel will often mention a document, for example, and say that it was introduced as number, say, 172; it is obvious that the number refers to the exhibit number, but if the word “exhibit” was not used, the keyers tended not to add the exhibit number tags. Thus, in such cases, you will need to add the tags.

A4

HLSL number citations: When the transcript mentions an exhibit that is in the database, add the relevant database item number. To add the item number in the transcript, insert Control-H to create the tag just after the close exhibit number tag, type the item number between the quotation marks, and delete the space between the marks. This information is usually added just after the document is introduced or identified, e.g., following the exhibit number. If a document is referred to repeatedly over the course of a page or a few pages, it is only necessary to add the HLSL item number in the first instance. If mention is made of a defense document, but not an exhibit, and you verify the document in the website, then it is fine in this case to add the HLSL item number after the close document number tag.

A5

Paragraph Breaks: The <TP>...</TP> tags demarcate a paragraph. There are occasions of unwanted breaks, especially above and below page breaks and often above and below unwanted page breaks. These should simply be deleted. For example:

“I want to ask you if this was also the</TP>

<PB N="2777"/>

<RunningHead>

12 Feb-M-PKP-9-1-Meehan-JP<LB/>

Court I

</RunningHead>

<TP>case with regard to the Weigl vaccine?"</TP>

should obviously be:

"I want to ask you if this was also the

<PB N="2777"/>

<RunningHead>

12 Feb-M-PKP-9-1-Meehan-JP<LB/>

Court I

</RunningHead>

case with regard to the Weigl vaccine?"</TP>

Speakers and following text: The keyers have placed these on separate lines, which lengthens the text unnecessarily and is contrary to the format of the original. The <TP> tag must thus be changed relative to the speaker tags. <Sp> indicates a new speaker (and is followed by </Sp> when the speaker is done). The speaker is further specified with his name, title, or by Q (question) or A (answer) marked off by <Spkr>...</Spkr>. Thus, the text looks something like (ignoring paragraph breaks for a moment):

<Sp>

<Spkr>THE PRESIDENT:</Spkr> What the speaker says.

</Sp>

The keyers placed <TP> after </Spkr>, so the text, as it stands, appears as:

<Sp>

<Spkr>THE PRESIDENT:</Spkr> <TP>What the speaker says.</TP>

</Sp>

However, we want the break to appear before <Spkr> to ensure that the speaker's identification and what the speaker says appear on the same line in the final version. Thus <TP> must also be removed after </Spkr>. The text from above thus becomes:

<Sp>

<TP><Spkr>THE PRESIDENT:</Spkr> What the speaker says.</TP>
</Sp>

These changes should be carried out automatically by the search and replace function.

A6

The HLSL transcript has Q (question) and A (answer) sometimes followed by a period, sometimes followed by a space, or sometimes followed by a colon. The convention we adopted was to change each instance to a colon. This is done automatically at the beginning of each day.

A7

Html entities will work properly, though there are not many that have been included in the text. The most common entity you will encounter is for the long-dash or m-dash. The proofreaders have marked this variously as em, 1/m, etc. The html character will remain in the text and the proper entity is —. This has been mistakenly input by the keyers as &emdash;. In addition, the keyers missed many long dashes, and added long dashes where there should be short (or n) dashes. The various replacements above will cover most instances. It is important to remember that there should be no space between the beginning or end of the m-dash.

There is one html entity that has, on occasion, been keyed into the text, but should not have been, namely, the entity for the German umlaut. This entity should, when encountered, be deleted and replaced by an “e”.

A8

The n-dash will most commonly be seen between dates, e.g, 1931–1933. You will need to perform a quick search of — after the appropriate replacements have been carried out to determine if any — needs to be changed to –.

A9

The primary editorial task is to incorporate the corrections noted by the proofreader into the electronic text; you can usually rely on the proofreader to find and note most errors, but do not assume that the proofreader has always interpreted the problem correctly or proposed the best solution. Procedures for revising the text for evidence code and exhibit numbers have been discussed above. Here, it will be enough to mention a few recurring editorial elements.

Page Breaks: The page breaks in the transcript should match the page breaks in the HLSL copy, not those in the NARA copy that the keyers used. Most often this means suppressing an unwanted page break using Control-O where the unwanted page break begins and Control-L where it ends (which is usually at the end of the running head that follows the page number). These unwanted page breaks are usually sub-pages, for example, 5986a or 9875 A.

We decided as a matter of policy not to delete the unwanted material but simply suppress it. So, for example, if there were no page 2895a in the HLSL copy the following:

```
“I want to ask you if this was also the
<PB N="2895 (a)"/>
<RunningHead>
12 Feb-M-PKP-9-1-Meehan-JP<LB/>
Court I
</RunningHead>
case with regard to the Weigl vaccine?</TP>”
```

would become:

“I want to ask you if this was also the
<O><PB N="2895 (a)"/>
<RunningHead>
12 Feb-M-PKP-9-1-Meehan-JP<LB/>
Court I
</RunningHead></O>
case with regard to the Weigl vaccine?</TP>

Spelling: Given the nature of the production of the trial transcript, it is perhaps inevitable that there be differences in spelling as well as differences in spelling conventions.

Differences in spelling occur most often with geographical names. E.g., you will find “Strassburg”, “Strasburg”, “Strassbourg”, and “Strasbourg” in the text. You should aim for some consistency, relying on the spelling provided in M-W. Although all corrections to the transcript should, in general, be bracketed, you do not need to bracket corrections to the spelling of names as they frequently recur. When it comes to spelling conventions – i.e., whether to type “court room” or “courtroom”, “Your Honor” or “your Honor”, or “Prosecution” or “prosecution”, let the transcript remain as is since the meaning in each case is entirely transparent and would be unaffected by any change. It is important to use the proofreader as a guide, but also to test the proofreader when the answer is not clear. For example, if you encounter a medical term that you have never seen before and the proofreader has suggested a correction, look the term up for yourself to determine whether or not the proofreader’s correction is sound. This often involves some investigative online work for more obscure words and – especially – names. The web is a valuable tool in trying to determine the proper spelling of names, titles, etc. The following is a list of helpful websites:

For general proofreader’s marks:

<http://www.m-w.com/mw/table/proofrea.htm>

For German military titles:

<http://www.hpwt.de/2Weltkrieg/Dienstgrade.htm>

For the Ravensbrueck concentration camp:

<http://individual.utoronto.ca/jarekg/Ravensbruck/NaziDoctors.html>

Other helpful websites are linked to on the Resources page of our website.

You do not need to know German in order to do the job, but the German translation of the transcript is available and has proven to be quite useful when, for example, trying to determine the correct spelling of a name. As to the spelling of names, often our website is the most helpful tool, particularly the “Document Search – Personal Author” page.

A10

There has been a fairly substantial list of words pertaining specifically to Case 1 that has been built up in the spell check database. However, the list is not exhaustive, and, as always, when there is any doubt, try to track down (usually by online research) the word or name under consideration for yourself.

Appendix II

Proofreading the Case 1 Transcript

The task is to read the keyed text carefully, word by word, to find and correct errors, including typographical errors, missing words, added characters, and other problems. Read slowly enough so that mistakes in spelling will register. Read the keyed text with the HLS copy of the transcript open to the same page for comparison; part of the task is to ensure that nothing in the HLS text is missing from the keyed text and that the pages begin and end at the same point in the two versions. We expect that, once the initial training period is completed, the work will proceed at the rate of about 5 to 6 transcript pages per hour. This will vary according to the number of problems in the keyed text; relatively clean text may proceed much more quickly, while the worst pages (with 30 or more errors) will require more time.

It is not necessary to know all the details of the case, but knowing the basic elements, including the leading individuals (the defendants, attorneys, and judges) and the subjects at issue (the charges in the indictment), will help make sense of the transcript and identify mistakes (especially in names). Review the “Basic Information” memo and the “Who Was Who” list that are available in the proofreader’s file. Other information is available in the project files and in the introductory text for the Website.

The work will be done in the Special Collections reading room in Langdell. Use the bound HLS transcript, the printout of the keyed transcript, and the guidelines and memos concerning the project. (The printout will be supplied by the project staff; the transcript for each trial day’s proceedings should be kept in a separate folder.) Work with a black pencil. Use standard proofreading marks (see under “proofreading” in *Merriam-Webster’s Collegiate Dictionary* or chapter 3 in *Chicago Manual of Style*). Put proofreading marks in the right margin; use the left margin to note queries for the editor. Draw a line through letters or words to be deleted, with a “delete loop” at the end of the line. Make the pencilled marks fairly dark so that they will be clear if the pages need to be photocopied for others to work with.

Before beginning a day's work, check to see whether the editor has revised the work done on previous days; if any guidelines have been added or revised, note them for reference. At the end of a session, leave the marked pages with notes for the editor either in the left margin or on post-it notes or written on a separate page. This can be done with a folder used to transfer text back and forth between proofreader and editor. Alternatively, if day-to-day supervision is not needed, keep the pages for each trial day's session in its folder and pass it on to the editor when it has been proofread (and contact the editor if a significant question arises in the meantime).

Two different categories of errors will be found: keying errors (errors introduced in the recent keying process), and errors in the original (errors made when the transcript was typed in the 1940s). The two sorts of errors are corrected in different ways. Every keying error that is discovered should be corrected. Keying errors should simply be marked for correction; i.e., the "bad" characters are replaced by "good" characters. If the keying error is an omission of text, add the missing words in the right margin and make a mark indicating where they belong in the transcript.

Errors in the original are corrected on a discretionary basis. We correct them when they cause confusion or present seriously misleading information (such as an incorrect name or exhibit number); simple and minor spelling errors can be noted with a [sic]; trivial typographical errors may be ignored—correcting the original text is not our primary task. Errors in the original should be corrected using []. E.g., "waz" becomes "wa[s]" and "Barndt" becomes "B[ra]ndt." (This is standard practice in legal materials.) Use the standard proofreading marks in the text and in the right margin.

Keying errors in German words should be corrected according to the original. Knowledge of German is not required, but read the words carefully for spelling that matches the original; it will help to become familiar with recurring terms. Errors in German terms in the original (i.e., in the HLS copy) need not be corrected unless the error is striking and the correction is clear. (For example, we need not correct "Fuhrer" to "Fu[e]hrer" but we might well correct "Furrer" if the typist forgot the proper term.) Discretionary items can be noted for the editor to check. Compound letters in German are spelled out. These are the vowels a, o, and u, which sometimes appear with two dots (indicating the letter e) above the letter; spell these out as ae, oe, or ue.

When in doubt, don't guess; make a note for the editor or supervisor. Questions are welcomed.

Compare the page breaks in the keyed text with those in the HLS copy. If the page beginning and ending do not match, mark the keyed text to begin and end where the HLS transcript page begins and ends; i.e., the keyed text should “conform” to the HLS text. Usually this will mean that a fragmentary page (e.g., page 23a) in the keyed transcript should be merged into a page in the HLS text (page 23). In these cases, mark the keyed text as “obs” (for obsolete), and the editor will add the required tags. Occasionally the page number will stay the same but the page break occurs in a different place in the HLS text; in these cases mark where the break should be located. The content of the original running heads does not need to be proofread; these heads will not be used in the final text. New running heads will be entered (automatically) by the program manager when the text for each day is added to the electronic files; these will include a short title, the trial date, and the transcript page. Conforming the page breaks is a crucial task; they need to match so that references to documents will occur on the right page according to our database (which relies on the pagination of the HLS copy).

There is no need for individual lines to begin and end at the same point in the keyed text and the HLS copy; matching the pages is the main point. (Paragraph breaks should also match.) Since we are not trying to replicate the NARA text exactly, there is no need to preserve the line breaks as they are in the film (and hence in the keyed text). We will delete the line breaks that produce short lines and let the lines run together into more compact paragraphs.

If the content is unclear in a way that makes it hard to determine the right word or letter, check the “Basic Information” guide and other reference sources available from the editor. These sorts of problems can be referred to the editor. For spelling, the authoritative dictionary is the *Merriam-Webster Collegiate Dictionary*, 10th edition.

Correcting the language of the original text is not required, but sometimes a serious or confusing mistake can be corrected by the use of bracketed inserts. Examples include accidental mixing of singulars and plurals, mismatched verbs, and obvious factual errors. (In one case, Adolf Hitler’s name was typed in as “Rudolf Hitler.”) The proofreader need not make an effort to spot these problems, since they are not on the

mandatory agenda, but it helps to note them when they are noticed. If the solution is obvious, note it in the margin, using [] to indicate added terms; if not, let the editor figure out what the solution should be (if any).

It is also helpful to notice and mark (in the left margin) two sorts of tags in the text, those for exhibit numbers (exhibit 15) and those for document numbers (document NO 15). These tags have been added by the keyers, inside < > brackets. Since the editor has to double-check these (and often add further information), it helps to have a signal that they are present. For exhibit numbers add (en) in the left margin; for document numbers add (dn).

Small points:

Degree symbol: The keyers have used special codes for the degree symbol, which should print out in the final text (and also for the ampersand). We can also spell out the small circle for degrees (of temperature) as a word in brackets: [degrees].

“Ober”: this word often appears in titles and ranks (Obergruppenfuehrer) (ober = super). The keyers have often mistyped it (e.g., as “Cher”), so titles like this are error-prone and often need checking.

