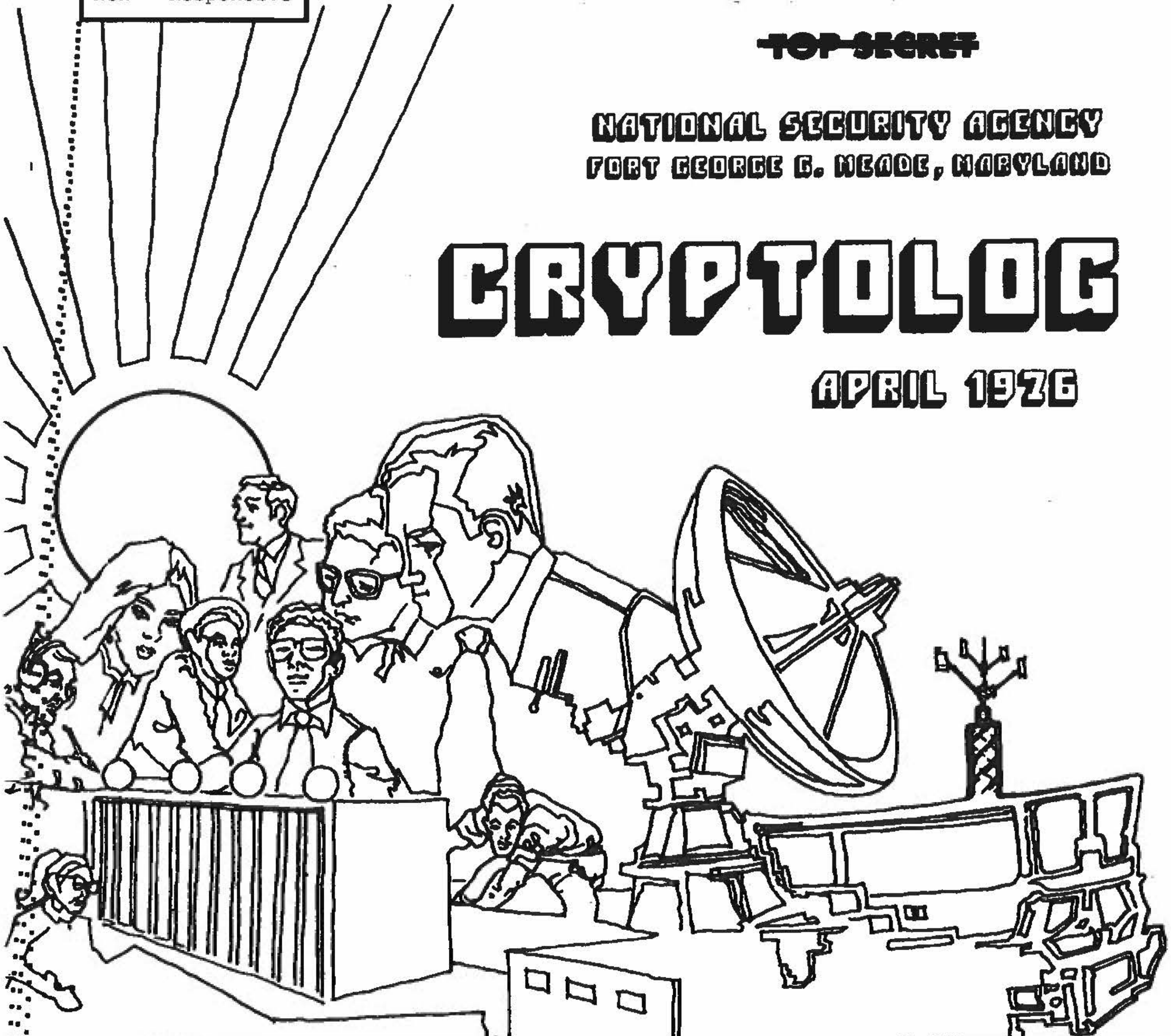
Non - Responsive

~~TOP SECRET~~

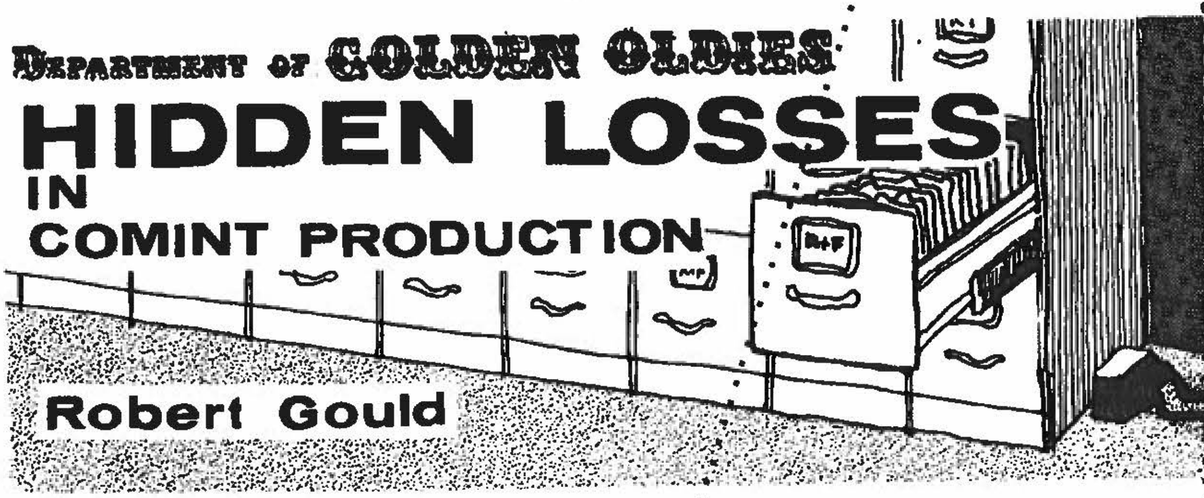
NATIONAL SECURITY AGENCY
FORT GEORGE G. MEADE, MARYLAND

CRYPTOLOG

APRIL 1976



EO 3.3b(3)
PL 86-36/50 USC 3605



Department of GOLDEN OLDIES

HIDDEN LOSSES IN COMINT PRODUCTION

Robert Gould

Most of the Golden Oldies published in CRYPTOLOG so far have been light in tone, but the Editor feels that this department should also reprint certain serious works that continue to remain "golden" and that readers may have missed when they were first published. One such work is the following article by retired NSA-er Robert Gould, which was originally published in KEYWORD, June 1971.

In discussing the function of language in the production of COMINT, much emphasis has been placed on the harm that can be done by misleading translations. Such errors may be potentially the most damaging, but there are other errors that can and do result in consistent and long-term losses that may never be made up. Some of these, which are discussed in the following paragraphs, are obviously ticklish questions, and no recommendations for solution are made here. Only the people directly responsible for the individual problems are in a position to recognize and correct the causes.

Failure to Recognize Intelligence Information

This failure may mean not recognizing the intelligence content at all, or not recognizing that it has a very limited life span. Its results may be an irretrievable loss of intelligence because of lack of timeliness or of an outright failure to publish. Commonly this error originates in unawareness of requirements at the level of selection, either because the analysts have not been properly instructed or because the customer's information needs have outstripped his expressed requirements. A review of processing methods may reveal still other hazards, though.

In most organizations the linguist who is [redacted] submits those messages which he decides to "read and file" (that is, file rather than process) to his checker for confirmation of the decision. Assuming the linguist has not delayed his decision in favor of processing other messages first and that the checker clears his read-and-file box promptly, the time lost may be slight, but if neither of these conditions exists the effectiveness of the unit is jeopardized. If the unit contains a relatively large proportion of new personnel, inexperienced and uninstructed in requirements, the hazards are magnified.

The simplest example is that of the new employee [redacted] The translator's mistake was understandable in light of his newness. It is a little harder to understand why he had the responsibility of making a decision on the message [redacted] Intelligence that is not recognized by the linguist will probably be lost permanently, and nobody will know that anything is going wrong.

If the traffic is a miscellany and processing is limited to issuing translations of individual messages, the losses will probably not be great, since the intelligence content per individual message is low anyway. If the material is the kind that is exploited in depth for the preparation of reports or if it can only be exploited in that fashion, the loss may be more significant, unless the linguist-analysts know their requirements and the various methods for satisfying them.

In one of our areas, a requirement was received for information on the procedures employed [redacted]

EO 3.3b(3)
PL 86-36/50 USC 3605

[redacted] The fact that the requirement could be satisfied was discovered only by chance, since the response necessarily depends upon the opinion of those who read the traffic.

Failure to Recognize Essential Technical Information

Probably the most easily documented losses are those caused by a failure to recognize the importance of information [redacted] There is no way of estimating what these losses may have been over a period of years.

[redacted]

Misplaced Effort

Circumstances may require that the decision to translate a message be left to the line linguist. In such cases people have sometimes spent hours on a message that doesn't contain a shred of intelligence, that cannot be published as COMINT, or that is too late by weeks. In the worst instances the production of real COMINT is delayed by the translation of trivia, or the product in general is degraded. The time lost can also become serious if an analyst is directed to devote research to a term report for which there is neither need nor requirement.

There are several reasons for such errors. Unfamiliarity with requirements is often one of them; another is a failure to understand the value and use of COMINT. A third one, which affects the newcomer, primarily, is simply inadequate exposure to the language. No linguist has at his disposal the entire vocabulary of a language, and the inexperienced, if confronted with volumes of new and varied material, frequently cannot really determine its value by reading it, and they play safe by translating.

The Checking Bottleneck

Language errors exact a regular toll. One simple mistranslation, if caught, may cost very little; the checker merely has to make a change and, if time permits, instruct the erring translator orally. But one checker supervising a large number of translators may find that an accumulation of such corrections measurably lowers his productivity and affects the timeliness of end product. A good many errors are not simple mistranslations of single words but fundamental misconceptions of the meaning of an entire message. These may require extensive or even complete rewriting and a consequently greater loss of time.

Where there are large numbers of inexperienced personnel, there sometimes arises a situation in which the entire production effort hangs upon the work of a very few checkers. It is not simply that the quality would suffer were it not for them, but that there would be no reliable product at all. This is theoretically only a temporary situation, but "temporary" in this case may mean a couple of years, and if such a situation exists the appropriate managers should know about it.

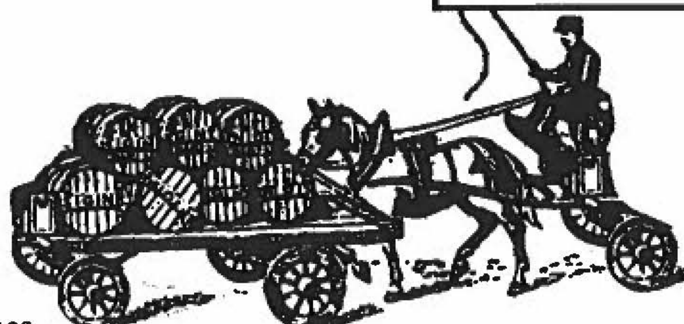
Scraping By

The worst error is attempting to operate with inadequate resources. It has not been uncommon to find areas in which the entire lexical and grammatical resources consisted of a couple of bilingual dictionaries and an elementary introduction to the grammar. No language is satisfactorily represented in such scanty references, and the result of this inadequacy has been that really crucial messages could not be issued in time or with sufficient clarity to be useful. Much of the blame belongs to the COMINT professional who often does not recognize how much he depends on expert knowledge, acquired imperceptibly over a period of 10 years or more. He consequently is inclined to rely on the obvious native intelligence of his juniors to guard them against errors from which they can actually be protected only by knowledge, or, lacking this, by adequate working aids.

Mistakes will always occur, and some of them will be serious. The object of these comments is not to eradicate errors, but to signal certain conditions or environments that regularly engender them. Although only a couple of examples were given, none of the errors cited is in any sense hypothetical, but all were drawn from personal experiences, observations, or specific studies.

EO 3.3b(3)
PL 86-36/50 USC 3605

LETTERS TO THE EDITOR



To the Editor, CRYPTOLOG:

After reading the December 1975 CRYPTOLOG article "What Are We About?", by LCDR James T. Westwood, I am compelled to comment.

Overall, the article was good food for thought. But one sentence in one paragraph prompted the following "tirade."

LCDR Westwood stated, "It follows that, if the processing and reporting effort ever catches up with the collection effort, we would be in real trouble because we would certainly have the cart before the horse."

I have shown the article to several people who were puzzled by that statement. Personally, I disagree with his philosophy although I sometimes think that the NSA/CSS and the Intelligence Community do not disagree. The philosophy that I would prefer to operate under is to *attempt* to keep pace with the collector in processing and analysis. Reporting of intelligence in a *selective* manner should follow, thus *not* inundating the user with reports he doesn't want or need.

To illustrate the frustrations and potential danger of that kind of thinking, I will tell you of an experience I had in the world of ELINT in 1966-1967.

I was one of three analysts assigned to a processing and reporting division. We had a new collection system feeding us tape-recorded data at a rate of between 5 and 10 tapes per day, 7 days a week. We had no automatic processing equipment and our job was to find out what signals were recorded on those hour-long tapes and analyze the important signals. The analysis results were eventually turned into a hardcopy intelligence report and distributed to users.

Since the data bandwidth and format were different from previous collectors, we had no experience to rely upon and promptly fell behind in our processing and reporting. Some of our fellow analysts used to stop by daily and jokingly ask us how many pulses we had processed (versus signals or tapes).

When our collection system went bad, it was replaced with one that automatically sensed when important signals were active and tape-recorded them. The good new was that we didn't have to sort through the garbage to find the hot signals. The bad news was that now *every* tape would require analysis in detail! We were swamped, deluged, absolutely and hopelessly buried with good tapes!

After a short huddle to figure out some way to cope with our collection system, we attacked our problem. (I don't know who said it first, but our motto was "The difficult we'll do right now, the impossible will take a little time.") Our objective was to determine which of the hot signals was the hottest and analyze that one first. So, two of us turned to the task of previewing tapes by running oscillograph records and observing the signal activity on an oscilloscope. The other worker did the analysis. I sure would like to have had stock in 3M, Kodak, and Polaroid.

One day, hundreds of rolls of oscillograph paper later, as the two of us were playing back a tape previewing [redacted] We both checked the paper record to confirm what I saw. It was there! We rewound to that part of the tape and watched the scope. There it was again and we both saw it! We excitedly showed the change in the radar signal to the senior analyst, who said that when we got done with the number one priority signal we would analyze that tape first. We noted the anomaly on the log, filed the oscillograph record, and went on to the next tape. We continued to preview tapes and almost hated to see "number one" signals because that meant we were getting farther and farther away from that hot tape.

Months later we ran out of local tape storage and moved all but a few dozen tapes to the IRC building. Still months later, we ran out of paper storage and were forced to get rid of those oscillograph runs. We were preparing for a new collection system because the second system had quit operating. The new one went operational and the tape poured in. There were new signals, new recognizers, and "old #2 signal" fell lower in priority. We never got to it!

The airman analyst who worked with me got out of the Air Force and tried to get hired as a civilian analyst. He was told that analysts were not needed, so he went back to Ohio and found work there. Eventually I moved to a different branch; new analysts were doing what I had done in the early days of the program. The "old #2" fell lower in priority and reports were published that said we knew virtually nothing about it.

I haven't forgotten "old #2." [redacted]

I did make one more try to get to that tape. Along about 1969 I was talking to someone about "old #2" and they got excited about the possible implications of a change like I was describing. The excitement got to me. I took the day off and located the tapes still in storage at the IRC building. The only way to find the tape was to go through each tape and check the intercept log for the comment that the airman and I made on the log. I spent the day looking through tapes in that hot, humid storage room. (A far cry from SAB-3 where tapes are stored today.) I kept thinking, "It has to be the next one," but no luck. The tape must have been reused or destroyed, as was the fate of tapes in those days. I did find that most of the logs had no analyst comments, which meant they were never analyzed, even where SOIs (signals of interest) were recorded. That's really sad.

But how do you measure the value of data unprocessed and not analyzed? If nothing bad happens to you, does that mean it wasn't important? [redacted]

Things are pretty much the same today. Collectors can bury the analysts and computers with data. We put filters and faucets on them in an attempt to control them. We must restrain them because we don't have the manpower, money, and may never have the technology, to process and report properly all that can be collected.

All through my NSA career I've been trying to catch up with the collector. I've tried to do my part to improve processing, analysis, and reporting. What LCDR Westwood says is that if I succeed, we will be in trouble. I don't believe that.

Remember the old motto, ". . . the impossible will take a little time."

Eugene D. Greiner,
F91

LCDR Westwood was asked whether he would care to comment on Mr. Greiner's letter to the editor. He replied as follows.

To the Editor, CRYPTOLOG:

Thank you for the opportunity to comment on Mr. Greiner's comment on my short item which appeared in the December 1975 issue.

In the light of Mr. Greiner's comment, I regret that my statement about the relationship between collection and processing was puzzling. That statement was, of course, a generalization. Accordingly, my first obligation is that, as a generalization, the statement must be accurate and non-exclusive in accordance with the "laws of logic." If my general statement is discouraging to processors and analysts, I regret that as well, though I can offer no alternative to the generalization that the sheer volume of collected SIGINT intercept has got to be greater, in the aggregate, than the volume of the same intercept that can be processed *per unit/time*. I emphasize "per unit/time" because it strikes me that what we face in this context is what Operations Research (OR) people refer to as the "basic inventory model," i.e.

$$X^* = \left(\frac{2C_s D}{Ch}\right)^{\frac{1}{2}}$$

(I must emphasize that I have no OR competency and will not be able to defend my point much beyond this "model.") From the OR standpoint, we are wont to consider an incoming collection "inventory" that produces an optimum volume of collection in terms of storage and production costs, and it has to be a volume which will *guarantee* (I'm still sorry!) that processors never catch up with collectors! I realize that this also is a generalization, but the converse is even more disastrous to comtemplate than is the generalization -- if only to managers!

The Chief of P1, Mr. William Lutwiniak, has spoken often and eloquently to the point that we have to collect a rather large volume to be in a position to isolate and exploit the relatively few "gems." The rationale seems abundantly clear. I believe this condition exists because we are the *unintended* recipients of the signals we collect. Understandably, the *intended* recipients do not face the problem illustrated by Mr. Greiner, inasmuch as their information-handling systems are inherently, and by necessity, self-serving. In conclusion, I appreciate the "old #2" problem. It is worrisome, but it steps beyond inventory and volume considerations into the realm of prioritization, and that approaches one of the theses of my article, i. e. inundation. From the maverick point of view, it seems to me that the "old #2" might well have been elevated to #1, if only temporarily, inasmuch as there is no substitute for the on-scene "judgment calls" of the person who happens to be attending the signal -- admittedly another generalization.

LCDR James T. Westwood, D/Chief, A732
("a processing and reporting shop")

COMPUTER-AIDED TRANSCRIPTION OF RUSSIAN VOICE MATERIAL

VIRGINIA SITTLER, A633

Introduction

Skeptics have long ago, and very convincingly, put down the notion that the computer is going to replace the linguist at NSA, or indeed anywhere else. However, a great deal of thought is going into designing ways in which the computer can *aid* the linguist. In this article I would like to describe a project currently underway in A Group which is designed to aid transcribers of Russian voice material. The project, as its name, STEPSTONE I, implies, is destined to be replaced by a more comprehensive system, entitled, naturally, STEPSTONE II. Both are part of what A Group regards as REDSTAR Phase I. However, since "interim systems" have the tiresome habit of hanging around long after their predicted demise, I shall not even speak of the future or of greater glories yet to come. Instead I shall deal only with the very real present.

Some time ago it was noticed that highly-trained Russian transcribers were spending at least a measurable portion of their time performing functions that were really clerical in nature -- typing, proofreading, indexing 5 x 8 card files, manually searching those 5 x 8 card files, etc. A wish was born to free transcribers from as many clerical or nonlinguistic functions as possible and at the same time make the fruits of the transcription effort available to other transcribers and to analysts on an on-line basis. The wish eventually became father to the deed in the form of LAYAWAY I, a computer-aided transcription system. The need to generalize the software for "exporting" to other transcription problems, and the overriding need to provide the intelligence data gleaned from transcription to the intelligence community as a whole via regular Prod Data Base (PDB) methods fathered STEPSTONE I.

STEPSTONE I, accordingly, is a generalized, on-line terminal subsystem designed to aid the voice transcriber in preparing transcripts. The software is generalized in the sense that it can be adapted for use on other transcription problems with minimal effort. STEPSTONE I is currently operational in A646 and equipment for six more systems in A Group, one in B Group, and one in G Group is on order. Interest has been expressed in acquiring STEPSTONE software both by [redacted] and by field sites. The term "terminal subsystem" implies that data is staged locally for relay to a large data bank (in this case the IBM 370). The large data bank computer provided manipulative services and a storage capacity that cannot be expected of a mini-computer.

Although the main thrust of the STEPSTONE system is data entry, temporary storage and retrieval capabilities are provided, as well as editing of input and file management. Access to the system is controlled by the user's Social Security Number. Transcripts prepared on-line are forwarded via magnetic tape to the PDB and thus made generally available. The PDB is the primary all-source file for all of A Group. Regular procedures exist for extracting information from the PDB. Briefly, they include the following:

- SELLERS, a program which daily selects and distributes data for each organization and for each analyst within that organization;
- TAFBUILD, an ad-hoc file extract, sort, and list procedure; and
- SPECOL, a sequential retrieval program that permits retrieval of individual data items using Boolean functions. Recently, a Cyrillic SPECOL package has been made available to permit retrieval of Cyrillic data.

A direct link to the IBM-370 via a typed command on the STEPSTONE terminal enables on-line access to the Model 204 retrival system Answer Files. The Model 204 system provides the fastest and most complete file-search capability in the building.

One of the desirable features of the STEPSTONE system is that, for the first time, Russian linguists are able to work on a computer in the natural language. STEPSTONE terminals are provided with Cyrillic keyboards and Cyrillic display. Cyrillic lists can be obtained locally or from the 370, and in general the linguist can remain in the environment where he is most comfortable -- the natural-language environment.

Advantages and Disadvantages

Lest I sound too much like Pollyanna, I want to hasten to add that there are some disadvantages to a computer system of this type:

- First, there is the matter of *computer down time*. If you have 48 transcribers, say, on one computer, and the computer goes down, you have 48 idle employees. Various schemes have been proposed and tried -- such as mirroring the hardware, so that when one piece of equipment goes down, its twin steps in and takes over. In addition, diagnostic tests are run daily to attempt to isolate potential hardware problems. But down time has been a problem in the past, is a problem now, and will undoubtedly be a problem in the future, despite our best efforts.

- Secondly, there is a matter of *system training time*. Fortunately, this has been minimal -- one day on-line has pretty well sufficed to make the transcriber an effective STEPSTONE terminal user.

- The third problem, for want of a better name, can be called *"Old Habits Die Hard."* Confidence is a fragile thing, and slowly built. The transcriber may not trust the computer not to do away with that critical piece of transcription, or he may fear that the computer will be down when he needs it most. In the extreme this could cause some duplication -- both entering data into the computer and sticking it in the right-hand drawer. On occasion, transcribers have been known to retrieve a transcript just after entering it, just to be sure it was there. This phenomenon tends to disappear with experience on the system.

Counterbalancing the disadvantages of such a system are the advantages, which I see to be primarily the following:

- The *ability to enter transcripts directly into the PDB*, thus making voice processing a regularlized component of A Group intelligence production.
- *Instantaneous accessibility*. Information that Jones has entered is available to Smith in its entirety immediately. There is no need to wait for a grapevine process to get the information to other transcribers and analysts.
- *Perfect readability of entered copy*. A system editor checks to make sure that dates are valid and that fields are filled in appropriately and that no necessary ones are left blank. Erasures and strike-overs do not exist on terminal-prepared copy -- the copy is always clean and readable.
- *Reduced noise level*, by virtue of the removal of all those clacking typewriters.
- *Economy*. It is hoped that improved transcription efficiency will lead to an improved ratio in the number of pages transcribed per transcriber. Paper costs should be reduced, since hard-copy is no longer the medium of data exchange.
- *Research capability*. With the local system retrieval, the transcriber can readily check identifications as he goes along. With the Model 204 retrieval capability, the transcriber has an immediate research-in-depth capability not available previously.
- *Automatic generation* of a good deal of what we call *"management information,"* instead of via the laborious manual procedure now used. This includes such information as the dates on which the input tapes were received, the number of transcription hours required, the lag time between transcription and linguist review, etc.

Now that we have considered the *why* of STEPSTONE (including the good and the bad sides) let us turn our attention to the *how*.

Hardware

STEPSTONE I is implemented on a Digital Equipment Corporation (DEC) PDP11-40 with 96K words of memory. It uses the RSX 11-D operating system. The peripherals for each system include

- an RP03 disk, with 21.5 million words of storage;
- an RF-11 disk with 246,000 words of storage; and
- a line printer.

In addition, two TU-16 magnetic-tape units, a card reader, and a paper-tape reader and punch will be shared between two systems by means of a unibus switch. An additional RP03 disk for each system has been requested, to facilitate system backup and recovery in case of disk failure.

Each PDP-11 can support up to 48 Incoterm SPD 20/20 CRT (cathode ray tube) "intelligent" terminals. The 20/20 terminal system is a cluster configuration, with one controller and 16K of core driving eight terminals. The terminal software is the Incoterm 3270 Emulator, a software package that makes the terminals act like IBM 3270 terminals. The emulator has been modified in-house to support the Cyrillic alphabet and to translate between ASCII and EBCDIC character sets. Fig. 1 diagrams the hardware configuration.

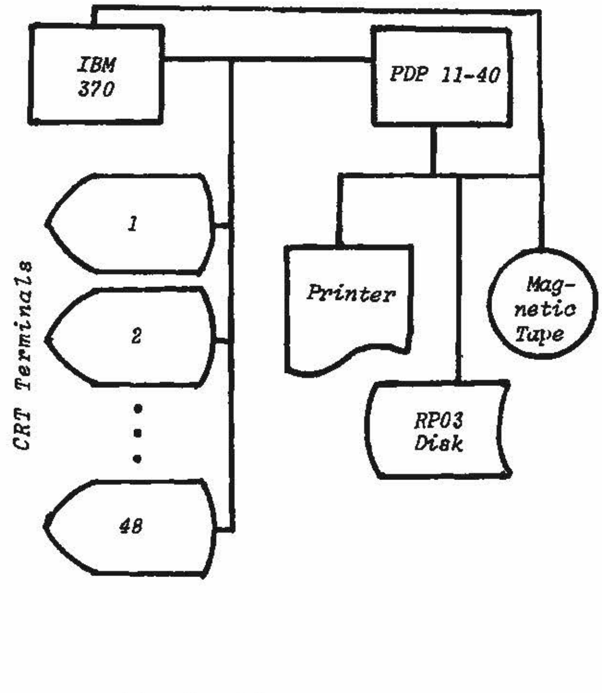


Fig. 1. STEPSTONE I Hardware

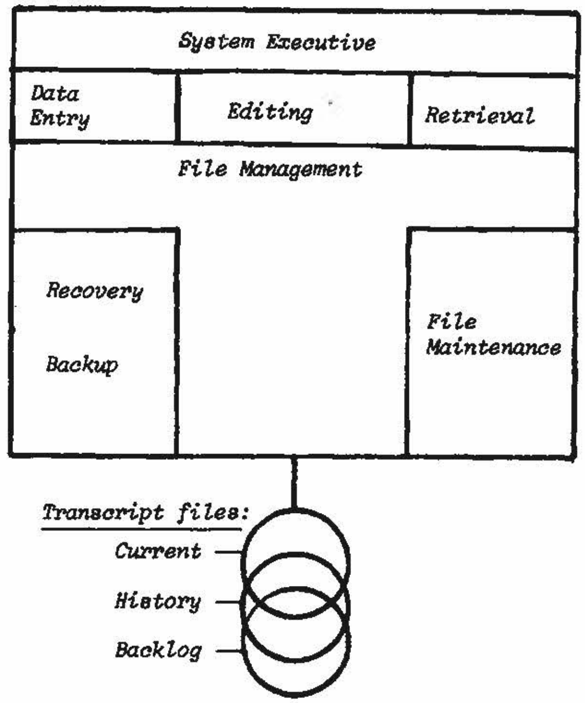


Fig. 2. Relationship of files to major software components

System Flow

The system is designed to support three files, which we call the Current File, the Backlog File, and the History File. Briefly, all data entered into the system goes into the Current File. When the transcript has been reviewed linguistically and analytically, it is both sent to the PDB and to the on-line History File. It remains in the on-line History File 15 days for local access and is then purged. In this manner the quantity of data in the STEPSTONE system remains manageable. Of course, the data is retrievable from the PDB after it is purged from the local STEPSTONE system. The Backlog File is for data that has been in the system for 5 days but has not been quality-controlled. The data is transferred to the Backlog File to free up space in the Current File. As time permits, linguists and analysts can review this data and release it back to the Current File. It then follows the normal path to the History File and PDB. Fig. 2 diagrams the files and their relationship to major software components.

The user's screen is divided into three parts:

- the system message line for messages from system programs and from the operator;
- the system command line to input typed requests to the system; and
- the data entry area for display of system templates, retrieval of data, and composition of transcripts.

Fig. 3 is a diagram of the user's screen and keyboard. The user's keyboard is identical to a Cyrillic/English typewriter, but has an additional 24 function keys which are used to transmit system commands and to control the screen's cursor.

Data Entry

Each terminal is assigned a workspace in the system. All functions -- data entry, editing, storing, retrieving, updating, and printing -- are applied to data in the user's workspace. Each user also has available upon request an alternative workspace. If he were composing a transcript, for example, and wanted to interrupt that process to retrieve information on a certain personality, he could SAVE the workspace with the partially completed transcript, do his retrieval, then RESTORE the transcript workspace and proceed from there.

In order to prepare a transcript, the user requests a blank header from the system, and fills in the header data and as many screens of text as are required. During the process he may page back and forth through his workspace, reviewing or correcting the transcript. When he is satisfied with the accuracy and completeness of the transcript, he uses a function key to call in the system editor.

The system editor has several important functions. It checks to be sure that the required header fields are present. It checks each field to be sure that numerics are not

EO 3.3b(3)
PL 86-36/50 USC 3605

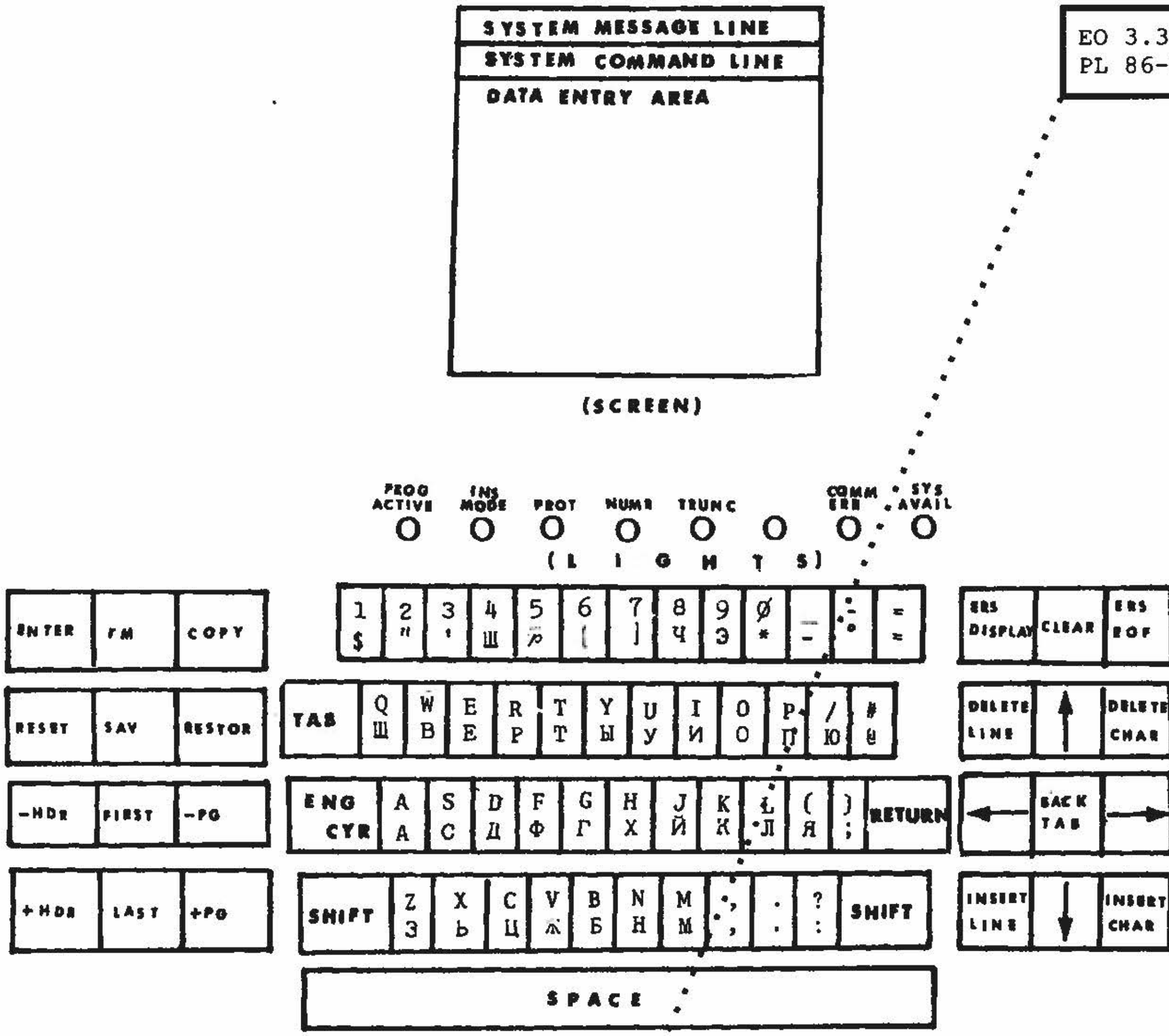


Fig. 3. User's screen and keyboard

present where alphabetics should be, or vice versa. In addition, it checks certain fields against each other. [redacted]

There are eight different types of text modules, which have been designed to be compatible with the STRUM field reporting format and the PDB. [redacted] The entire transcript is edited and if any errors are encountered, the transcript is returned to the screen with erroneous parts intensified for easy identification.

When the transcript is correct, the editor releases the record to be added to the on-line Current File. At this point, if the transcript were printed via system command, it might look like the one in Fig. 4. Any header field or text line may be in either Cyrillic or English, as the transcriber chooses.

[redacted]

Fig. 4. Sample transcript

Retrieval

As an aid to the preparation of transcripts, a sequential retrieval is available which operates against certain of the transcript header fields. To use it the transcriber requests a template, which is a list of allowable retrieval fields, from the system. He then fills in the fields and the values he is searching for. A hit record is returned to the screen immediately, rather than waiting for the entire file search to be completed.

Management Information

As mentioned previously, STEPSTONE I also provides certain management information to the sponsoring organizations. Information extracted from each of the on-line files is sorted and listed on a daily basis for file managers. A daily magnetic tape is also created of management information records for historical purposes. Follow-on processing of this informatior has not yet been specified. It is intended, however, to dovetail with the A Group CSR (Consolidated Scan Record), a tape management and accounting information record which is intended to monitor intelligence throughput.

PDB Processing

Once a transcript has been released from the STEPSTONE on-line system for PDB processing, it must pass through a STRUM-formatting program. STRUM is a highly-structured field-reporting format, consisting of many different types of lines, each with its own individual kinds of data. Many of the text lines in the STEPSTONE transcript are almost identical in STRUM, but other lines, [redacted] are generated from STEPSTONE header fields.

Once records are in STRUM format, they are passed through GENED and FILE GEN, a pair of programs that edit STRUM data for PDB standards and generate the PDB.

IBM Interface

The STEPSTONE terminals are interfaced to the IBM 370 via a typed system command in the terminal command line. The user will then be able to address Answer Files on the Model 204 retrieval system. While in 370 mode, the user's terminal will act just like a Model 204 terminal and the user must follow 204 protocol. Features local to STEPSTONE I will not be available in 370 mode. However, since the Incoterm is configured with Cyrillic, the STEPSTONE terminals will have a Cyrillic capability which the standard 204 terminals do not have. Information obtained during STEPSTONE processing will be incorporated into Model 204 Answer Files.

Summation

To sum up, I will simply restate the features of the STEPSTONE transcription support system:

- On-line data entry and editing of transcripts;
- Local retrieval and file management;
- Cyrillic data base;
- Direct access to Model 204 Answer Files on IBM 370; and
- Automatic entry of data into PDB.