FBIS MODERNIZATION REQUIREMENTS

DOC. NO.: SDS-002

30 May 1984

Revised version submitted

15 August 1984

by

Xerox Special Information Systems

Declassified and Approved For Release 2012/06/05 : CIA-RDP88-00218R000300180001-4

1.0 HEADQUARTERS REQUIREMENTS

1.1 General Requirements

Headquarters must support a range of simultaneous communication, data base management, and publication production functions.

Input and output functions supported 24 hours a day, 7 days per week include:

- automatically process incoming and outgoing messages to and from the bureaus
- automatically process outgoing wire service
- electronically accept incoming translations from independent contractors

Daily but not continuous processes supported include:

- electronic processing (edit, format, compose, and typeset text and graphics with the eventual full electronic implementation of graphics handling) of material for any FBIS printed or electronic output. Products include the Daily Reports, JPRS publications, wire service, Analysis Group reports, and miscellaneous newsletters, memoranda, and ad hoc translations
- storage and retrieval of data and products from files by FBIS users
- electronic processing of instructions and contract service orders to independent contractors

Periodic processes to be supported include:

- compilation of operations, system management, and administrative statistics

System characteristics include:

- ease of new user training including such features as on-line "Help"
- support 350 workstations with 300 simultaneous logged on users
- complete file access control
- interruptible user tasks
- simultaneous manipulation and display of English and foreign language text

1.2 Processing

The system is intended to support, expedite and automate the processing and analysis of open source material. This includes tasks of selecting material to be processed, managing its translation, editing and publishing. In addition it assumes storage of the material for retrieval and processing by analysts.

Display goals are:

- the display of files at every stage in the processing flow
- the simultaneous terminal display of more than two (and perhaps as many as five or six) files

Intercommunication goals are:

- communication between terminals of messages, data files, and instructions
- multiple incoming message or traffic streams from different sources, e.g., wire traffic, English Press Agency, input from monitors, etc.

Access control goals are:

- control total file access as well as specific read and write access (control user permissions)
- more specifically, limit the authority to alter or append files and fields
- prevent simultaneous alteration to the same record

File manipulation goals are:

- accept annotations and comments both to text and to header (i.e. tracking and indexing information) areas
- some facility (commonly referred to by the trade name "Edit Trace") must exist to track and display changes made to text at sequential processing steps. Implicit in this is the maintenance of several versions of a file reflecting at least three editing processes and including the identity of the individual making the change or comment. The ability to control the quality of FBIS products, which relies in the paper world on review of penciled editorial changes and comments, is deemed critical to both field and Headquarters.

Processing revolves around:

Text processing (See 1.7) for the creation and editing of FBIS products

- Data bases for support to processing and for support to analysis (See 1.4, 1.5, and 1.6)
- Message handling (See 1.3)
  - = communication between field and headquarters and independent contractors and headquarters
  - = communication with consumers (Wire Service, soft-copy "publishing", output to printing plant)
  - = communication among headquarters elements
- Composition (See 1.8)
  - = requirement to proof output

1.3 Message Processing Dissemination

Headquarters requirements fall into three areas:

- field traffic
- Wire Service
- independent contractor traffic

Headquarters use of field traffic is the converse of message traffic support in the bureaus (See 2.6). In addition, there are the following constraints:

- all outgoing messages are released by the Wire Service Slot Editor
- all "Immediate" and "Flash" messages - incoming and outgoing - are delivered to the Chief, Wire Services
- there is considerably more parallel dissemination in Headquarters, for example all publishable messages go to the Wire Service, the Daily Report and Analysis Group

Wire Service support includes:

- interface, currently in standard teletype format, to 20+ telephone circuits which deliver the wire to its consumers
- 24 hours a day, 7 days per week access
- announcement (standard messages) transmitted
  - = beginning of day
  - = after 60 minutes of idle-time

= end of day

- transmission in queues
- logging
- pre-emption
  = interruption is permitted
  = change order of queue
  = interrupted item is re-inserted in top of queue
  = deletion of any item from transmission queue

Independent Contractors send in completed translations. Some, but by no means all, independent contractors have personal computers.

- Incoming material should be receivable by such means as:
  = telephone/MODEM
  = reading in floppy disks or cassettes in agreed upon standards
  = OCR or rekeying
- Outgoing material, items to be translated, glossaries, newsletters, etc., will be
  = in hard copy (bulk of material)
  = floppy disk or cassette in agreed upon standard
  = potentially facsimile for some time critical materials

1.4 Data Base Management Requirements

The system is intended to support the creation, storage, maintenance, retrieval, and report generation operations for a large number of files that may be grouped into one or more data bases. The nature of the files ranges from large text files to small highly structured files. Composition and mark-up characters will be ignored during text searches.

- Data value types include: text (alphanumeric), alpha, numeric, date, time. Nothing in the system should preclude the later inclusion of graphics and raster images as data types.
- Data element (field) attributes include:

= length "unlimited, an entire four hour speech looks like a single field to a user. (This description of the user view should not be construed to be an implementation requirement.)

= groups, a related set of different fields within a data record, are considered to be logically bonded and can repeat within a single record just as records repeat within a file

= a record can have more than one type of repeating group

= mandatory elements (fields); i.e. a group and/or record cannot exist unless the field is present

File relationships within a data base are such that data elements (fields) in one file have logical relationships with data elements in another file. Data base management systems should be able deal with such relationships during file maintenance, retrieval, and report operations.

A file maintenance capability is needed that understands update transactions with identification of input fields (by name or position of occurrence) and validates the individual field values input by a user or in some cases automatically. The file system must also preserve the logical relationships of groups and records such that distortions will not occur during retrieval and reporting.

File maintenance must be able to be performed in batch and interactive modes. There is a requirement to create new files and modify and add logical structure, data types, and data attributes within existing files.

Specific file maintenance operations include:

add new records

delete records

add fields to existing records/groups

change exist fields in records/groups

delete exist fields in records/groups

add a new occurrence of a group

Other file maintenance processing include:

- multiple changes to a file based on a conditional statement

- insuring the integrity of files, records and data structures with automatic checks of updated records versus the files definition

The files listed in Exhibits A and B include the major file currently kept by FBIS. They do not include the myriad of working files that, temporary in nature, may support the various bureau and headquarters processing.

1.5 Data Base Management Retrieval Requirements

A general retrieval capability is needed to operate against the types of data values, attributes, and relationships described in section 1.4. There must be an ability to search with or without regard to the logical relationships inherent in a particular data base. These conditions include:

- within a single document or message
- between or among documents, messages, and data bases
- within a single occurrence of a group
- among groups

Search operators include:

- equal, not equal, greater than, less than
- the greater and less than combined (into a single range statement)

Data values are stated according to the definitions discussed in section 1.4. Special conditions and expressions for character string and text searching are required as follows:

- for any data element defined as text, the system will use spaces, commas, periods, etc. to denote word boundaries
- don't-care character indicates whatever position the character occurs in the string is correspondingly ignored in the search. It can occur anywhere in the string
- prefix character, indicates that the first n characters of a word are to be searched against the value in the search argument, remaining characters of the word are ignored
- suffix character, indicates that the last n characters of a word are to be searched against the value in the search argument, beginning characters of the word are ignored

- a don't care character can be used in conjunction with either a prefix or suffix character search argument

- synonym table, a set of data values, user-controlled, that may contain prefix, suffix and don't care characters; the use of a synonym table in a query results in all of the values in the table being used in that query as ORed search arguments

- word proximity, specify relative position of search arguments; within n words before or after, within a sentence, within a paragraph (when applicable), within a record

Boolean expressions include:

- and, or, m of n

- at least two levels of nesting and multiple nests per query statement

The file records that satisfy the truth condition of a search will be saved as a hit file, which can then be subsequently searched either automatically as in a multi-file search or simply user invoked.

- a capability is required for creating, storing and revising search queries. The system will permit a user the facility to call-up a stored query and cause it to be executed

- the system will permit a user to display the number of "hits" encountered while the search is being executed

- a user will be permitted to cancel a search while it is being executed

1.6 Data Base Management Report Generation Requirements

A general report generation capability is needed to operate against the types of data values, attributes, and file relationships described in section 1.4. There must be the ability to disply data in either soft or hard copy mode and to arrange data elements in such a way that their logical relationships are apparent to a viewer. The following operations apply equally against files and "hit" files, it is assumed that users invoke the report system directly and indirectly when using the search or browse operations. Report generation features include:

- prior to display/print perform major and minor sorts according to user specified data elements

- a capability is required for creating, storing, and revising output formats

- output format definitions include: specific data elements, placement and width of data values, headers, footers, automatic page breaks and page numbering
- print/display entire file, entire records, parts of records
- permit a user to specify the device output destination
- highlighting: allows a user to specify, which data element(s) within retrieved records the user whishes to see highlighted. The term(s) specified may be either one or more of the query search terms, or may be terms added to the query statement that affect only highlighting
- print/display a window (displacement) around data values that matched arguments in the query statement
- apply mathematical manipulations to both file stored and derived/calculated data values, incorporate resulting data in output reports graphics such as bar charts, piecharts, etc.

After a search has been executed the system will provide the capability to "browse" the hit file(s) via softcopy; capabilities desired include:

= page forward and page backward

= scroll up and scroll down

= skip pages and records

= go to first/last/next record

A capability called Hold to allow a user, while browsing, to save records for subsequent browsing.

1.7 Text Processing

Text processing requirements vary between field applications and headquarters. While the need for high-quality editing wordprocessing functionality is required universally, headquarters requires the ability to control composition functions. Conversely, bureaus only need to be concerned with a simpler set of composition functions (that is, column width, word wrap, etc.) on those non-publishable material intended for hard-copy output on local line or page printers (correspondence, reports, memos, etc.).

Basic assumptions:

- A test processing capability does not sacrifice any of the file management, directory, access, and control features which characterize the other processing functions (collection, selection, translation).

- Test processing in the FBIS is inseparable from the file management capability. That is to say, the overall efficiency of text processing is as much determined by the simplicity of the procedures to find, review, and route the copy as in the actual ease of inserting, deleting and altering text characters (the conventional definition for the concept of text processing).

- Useful, flexible directories are essential to the process whereby source material is chosen for editing, styling, composition, pagination and ultimate inclusion in reports. Directory requirements are:

  = incoming files sorted, for example, by geographical topic, precedence, origin, time, length

  = reviewed files sorted by required processing status (hold, work, reject, supervisory review, to be composed, composition completed)

  = all files sorted dynamically FIFO, LIFO, alphabetically

  = directories can search selectively and globally on text, keyword, editor, time, date, status, precedence, and length

  = directories of versions of individual files (audit trail for edit trace)

Text manipulation features required:

- state-of-the-art word processsing functionality. Including, but certainly not restricted to full implementations of insertion, deletion, search and replace, block, sentence, paragraph, line, word and character manipulation, memory and burst keys, text copy and merging, etc., and support for multi-lingual word processing

- support for line or page printer formatting (including margins; line spacing; quadding left, right and center; page numbering, tab stops; running heads; justification; etc.) to the extent of the printer capability (access underlining, bold, sub- and super-scripts). The printed output shall match the displayed copy in format to facilitate locating a passage in the displayed file.

- support for generic composition coding

- support for input of full typographic composition coding (markup) (Headquarters only)

- support for page, chapter, and book assignment (Headquarters only)

- support for automatic or semi-automatic generation of indices and tables of contents (Headquarters only)

- support word/character counting with alert to AUTODIN take length limitations

Editing tools required (significantly in excess of the norm for text processing):

- efficient, full-featured spelling checking (specifics to be determined)

- syntactical checking for unusual terminology and generally acceptable grammar (along the lines of Writer's Workbench test for vulgarity, slang, syntax and capitalization)

- the creation and use of on-line dictionaries, glossaries and references: both personal, departmental, and system-wide (Note: The foreign language text should be handled in the vernacular representations)

- support for the integration of text and graphics

1.8 Composition Requirements

Ultimately the product of the system is information in both electronic and printed form. Composition addresses the aesthetics and legibility of the printed product.

Composition requirements include:

- high-quality hyphenation and justification

- quality book typography (a reasonable subset of kerning, tracking, ligature generation, hung punctuation)

- support for extensive collection of accents, diacriticals, etc.

- efficient, full-featured batch book pagination with interactive review and alteration (full-features to be determined but to imply footnoting, floating illustrations, widow and orphan control, etc.)

- code translation (support for "exploding" generic coding)

- potential support for input, manipulation, and output of line art and halftone graphics. This might include a what-you-see-is-what-you-get composition device and "soft typesetters" for the integration of text and graphics

- generate APS-compatible output

2.0 BUREAU REQUIREMENTS

2.1 General Requirements

Field bureaus must support a range of simultaneous communication, data base management, file management, and text processing functions.

- Number of Workstations per Exhibit C.

2.2 FBIS: Collection Requirements

The system is intended to collect and store original information from four major sources: radio, press agencies, television broadcasts, and publications. Each source presents different specific challenges, but the overriding objective is to expedite, manage, and automate the accumulation of source information regardless of the input method.

1. Radio Broadcast

Exhibit D summarizes the scope of radio broadcast collection for all the field bureaus. Information includes numbers of simultaneous transmissions which must be accepted, the number of monitoring stations, the number of monitors per broadcast and the total number of minutes per week of broadcasts that are received. Other values include the number of sites remote from the actual bureaus at which radio reception takes place, and the numbers of simultaneous broadcasts to be received at these remote reception sites.

Automation of the reception of radio broadcasting would include:

- provision for display of date and time-of-day (GMT) on the workstation screen
- provision for displaying the monitor's assigned schedule in a window on the workstation screen
- provision for similar display of the entire bureau schedule, if desired, for viewing by editors, the chief monitor, the shift supervisor, etc.
- provision for easily copying a schedule item (source and language), as well as date and time of day, to a Program Summary or Translation window on the workstation screen

Broadcast items are recorded on recording tape, and do not exist as conventional source text files until they are translated and manually keyboarded. Thus, they require a unique set of processing controls to support their indexing and retrieval while

on this sequential recording media. While alternative approaches are welcome, the goal is to be able to retrieve both broadcasts and individual items within larger broadcasts. (It is presumed that specific item retrieval will be based on a cross-index between a manually generated broadcast summary list with specific start and stop positions on a recording tape. However, any approach which satisfies the goal is welcome.)

2. Press Agency Sources

Exhibit E summarizes the scope of the collection of press agency transmissions at all the field bureaus. It should be noted that while the large majority of information is transmitted using conventional character-oriented, digital (i.e. teletype either by land line or radioteletype) means, a significant number of services use non-character-oriented "graphic" modes: Hellschreiber and facsimile. For both these modes, exhibit E includes numbers of press agencies, types of transmission (i.e., transmission technology, not content), and weekly volume of transmission.

A system should be able to accept and store and recall all digitally transmitted (i.e., forms of teletype) English language transmissions without manual intervention. Non-English transmissions should be handled automatically if the process of reviewing and selecting material for translation can be accomplished as efficiently on-line as from printed paper copies. (Note: Some non-English transmissions use non-Roman alphabets, e.g. Arabic) There is no "preselection" of incoming press agency news beause selection, by definition, must be done after the transmissions have been received. Alternatively, this is to say that press agency transmissions is done virtually constantly, not at rigidly-scheduled broadcast times as in radio or television reception.

Graphic-oriented services (Hellschreiber and facsimile) would be recorded on conventional hard-copy, imaging machines.

3. Television Broadcast

Exhibit F summarizes the scope of television broadcast collection for all the field bureaus. Information includes weekly volume of transmission and the input source technology: either conventional "local" vhf/uhf broadcast, satellite reception, audio reception without accompanying video, or input of video tapes provided by an outside source.

The automation of television broadcast reception, recording and retrieval will consist of provision for workstation aids similar to those described under Sec. 1, Radio Broadcast.

4. Publications

Exhibit G summarizes the number of distinct publication titles received and the number of individual issues received weekly.

Automation of the publications collection follows that of any good reference library. It should include:

- Logging the receipt of each publication
- Alerting to late and missing issues
- Alerting to subscription expiration in time to assure reception continuity

There is a parallel requirement for Headquarters

The search for new or improved quality radio and TV transmissions (cruising) is an important function of the bureau. The cruiser searches for stations identifying the ID, language broadcast in, and any necessary program details.

Automated support of this function would include:

- independence from collection operations, since it often occurs at the same time
- coverage of the radio and television spectrum per Exhibits D and F
- keyboard entry into a file of
  - = current time and date
  - = frequency to which receivers are tuned
  - = antenna configuration/pointing
  - = signal strength
- automated distribution of the above data to monitors and to a file for distribution to MOD

2.3 Processing

The system is intended to support, expedite and automate the processing of source material. This includes the tasks of monitoring, selecting, translating, and editing source material. In general this is perceived to be implemented through a multi-terminal system to support the selective access, manual manipulation, and subsequent refiling and routing of individual files.

Display goals are:

- the display of files at every stage in the processing flow
- the simultaneous terminal display of more than two (and perhaps as many as five or six) files

Intercommunication goals are:

- communication between terminals of messages, data files, and instructions
- quasi real-time viewing of a running summary from terminals other than the one it is being created on; this will facilitate the production of FYIs and translation of important quotes during a live broadcast speech
- multiple incoming message or traffic streams from different sources, e.g., wire traffic, English Press Agency, input from monitors, etc.

Access control goals are:

- control total file access as well as specific read and write access (control user permissions)
- more specifically, limit the authority to alter or append files and fields
- prevent simultaneous alteration to the same record

File manipulation goals are:

- accept annotations and comments both to text and to header (i.e. tracking and indexing information) areas
- some facility (commonly referred to by the trade name "Edit Trace") must exist to track and display changes made to files at sequential processing steps. Implicit in this is the maintenance of several versions of a file reflecting at least three editing processes, including the identity of the individual making the change or comment. The ability to control the quality of FBIS products, which relies in the paper world on review of penciled editorial comments, is deemed critical to both field and Headquarters.

Location and retrieval goals:

Press Agency --

Transmissions should be logged automatically upon entry into the system with retrieval keys exploiting the standard subject and source codes already embedded in the files by the press agencies. In addition to these keys, access should also be by system generated source and time stamps.

Data Files --

There is an implicit and fundamental assumption that each processing step will be reflected in the database so that both data and status information is updated. Upon this assumption, files can be accessed by their processing status attributes. Illustrative examples, (not to be taken as specific requirements) would be the generation of directories of files entered on a specific date, files originating from one specific source, or files presently translated but not passed through editorial review. Commercial text processing systems call this concept selective and global directories.

2.4 Publishable Traffic Dissemination

Field bureaus distribute their publishable product--English-language translations of source material primarily--to FBIS Headquarters in Washington, to other field bureaus and to lateral consumers. The dissemination process requires the ability to access a file, append appropriate precedence and routing instructions, and to enter those files into a communication network(s).

Exhibit H summarizes the total volume of products to be disseminated, the number of unique subject packages into which all products fit, and the number of consumers subscribing to at least one subject package.

Address selection goals:

- correlate the file's content with the appropriate recipients interested in that subject. That is to say, each file may go to a unique combination of destinations dependent on the subject and importance of the file contents. (Alternatively, not every subscriber wishes to receive every file)

- cross-check to prevent multiple versions of a file being sent to a single consumer (automatic deletion of redundant addresses)

- on-screen display, review, addition, and deletion of subscribers for a particular package or for a particular file

Precedence Management:

- support four different priorities for first in-first out transmission of files in sequence

- support dynamic alteration of a file's position within a queue, including its immediate transmission

- support dynamic alteration of a file's queue, thus permitting a change in precedence

2.5 Data Bases & User Support

The data base at the field bureaus falls into four categories:

1. original source magnetic tapes, received periodicals and other non-digitized foreign language input to the field bureaus. This is manually archived as needed and will not be automated.

2. press service source material received digitally from wire services. This material will be read by the automated system, then archived and distributed electronically.

3. translations and summaries as composed into outgoing messages.

4. translation aids.

5. administrative files (for such things as payroll, etc.)

The goal is to automate the last four of these categories.

Press agency data base goals are:

- input of press agency messages directly into the system
- automatic short-term archiving of press agency messages
- no re-keying of English language press agency material
- ensuring that press agency messages that have not yet been purged from the data base will be available for recall by editors and monitors. Recall will be based on date and time of original source and press agency name

Translation and summary data base goals:

- automatic addition of translations and summaries to the short term archive data base after they are entered into the system
- no re-keying of translations and summaries
- provision for translations and summaries that have not yet been purged from the data base to be available for recall by editors and monitors. Recall will be based on date and time of original source and source name

Translation aids data base goals:

- automatic addition of translation aids to the data base after they are entered into the system
- provision for multi-lingual translation aids
- provision for translation aids to be edited at monitor and editor workstations

- ensuring that translation aids will be available for recall by editors and monitors

Administrative files DBMS goals

- maintaining administrative files (payroll, subscription, etc.) on a data base at the field bureaus
- providing a reasonable complement of DBMS reporting and retrieval commands

2.6 Message Traffic Support

The system must support a two-way messaging facility between all bureaus, lateral customers, and headquarters. Addressing, formatting, and technical standards and conventions are rigidly standardized facilitating the highest degree of automation in the addressing (both outgoing and incoming), formatting, header validation, and technical interfacing to a range of alternative communication links. Exhibit H summarizes the volume daily message traffic.

A system must support the conventions required to use these alternative communication links:

- AUTODIN Mode I and Mode V
- Diplomatic Telecommunications Service (DTS)
- Defense Communication Service (DCS) dedicated channel (London only)
- Telex
- Department of State COMSAT circuit

.....as well as being posed to exploit alternative technologies under consideration.

Fundamental assumptions include:

- message creation and transmission implies the retention of both the text as well as the transmission status information of that message at the originating point (messages are not automatically deleted by the act of transmission)
- no unreasonable system overhead
- conformity to existing FBIS message formats as outlined in the FBIS Editorial Handbook
- all incoming and outgoing messages should be automatically logged by both the sending and receiving systems

Precedence/priority handling goals:

- messages handled in order of precedence (four levels)
- within each precedence level, messages handled FIFO sequence

- support dynamic alteration of a message precedence to allow its transmission even if it had originally been assigned to a lower precedence queue

- automatic alert upon the receipt of the highest precedence message

Message integrity goals:

- in addition to error checking inherent in the transmission medium (e.g. parity checking, checksum verification, etc.) a secondary level of error checking of header information will examine message designation, format, precedence, and administrative check numbers

- messages which fail the error-checking pass will be routed to an intercept queue for manual remedial attention

- the originator of failed messages is to be alerted and offered alternatives such as aborting the message, retransmitting the message, or displaying, re-editing, and retransmission of the message

Message addressing goals:

- incoming messages to individual bureau terminals should pass to the user's mailbox automatically

- incoming messages should be logged and copied into the target bureau's files in addition to the addressee's mail box

- outgoing messages should be automatically formatted for transmission according to the header information contained in the addressee list

- header information should be automatically generated for continuation takes of a multiple take message. Similarly, an operator should not have to specifically flag first and last pages of a long file when it is passed to the message processor for transmission

- outgoing messages should be logged and stored in the bureau's files

3.0 Interfaces

3.1 AUTODIN per Exhibit J

3.2 Diplomatic Telecommunication Service per Exhibit J

3.3 Commercial Telephone (TBD)

3.4 Consumers (TBD)

3.5 Printing and Photography Division (TBD)

3.6 Commercial Data Bases (TBD)

3.7 Agency Computer Systems (TBD)

3.8 Signal Interfaces (TBD)

3.9 Antennas (TBD)

4.0 Security

All FBIS Headquarters users need access to both unclassified and classified systems. Some FBIS Headquarters users, including all AG analysts and TBD % of Production Group users (a total of approximately TBD users), spend most of their time in a classified environment. The rest of FBIS Headquarters users spend most of their time in an unclassified environment.

All classified material must be kept in a classified system. This includes all files in 1.4 which are classified. All two-way communications between other Agency computer systems and FBIS must be via a classified system. All hardware in a classified system must be Tempest approved or kept in shielded enclosures. Unclassified hardcopy products can be produced from a classified system. Outputs of the classified system must be labelled with the proper classification. One-way-only computer to computer communications are possible from an unclassified system to a classified system.

All unclassified material can be kept in either an unclassified system or a classified system. However, certain functions must be in an unclassified system, including:

- all bureau functions
- all computer--as opposed to hardcopy--communications outside of the Agency, e.g., the FBIS Wire Service, field bureau traffic, etc.
- all JPRS functions