Initial Information Services Group Requirements
For Machine-Based
Information Storage and Retrieval System

Assumptions

1. ISG area divisions will input 360,000 to 400,000 records annually into its machine information system.

2. Request experience, both during normal times and in crisis situations, has demonstrated the value of storing and maintaining machine files of information for at least ten years for most series of indexed items.

3. The CRS generalized information system will be required for contingency support of production offices.

4. Date is the most effective purge criteria for these files.

5. Some change in file order will be necessary.

6. It should also be anticipated that CRS will continue to take on special projects in support of the various DDI production offices and/or will periodically acquire already existing machine language files.

Immediate Requirements (up to one year)

Input

1. Free current limitations for the following data elements:

    (a) Title length: Present - 252 characters
                                     Increase to - 504

    (b) Keywords: Present limit - 10
                                     Desired limit - 20

    (c) Twenty-character
          document number   Present limit of 3rd field - 15 characters
                                     Desired limit of 3rd field - 25 characters

2. Eliminate the EDPSD technique of subject/area code reversal to cut down the number of machine records created. Elimination will also provide the capability of 20 subject and 20 area codes to replace the current limitation of 10 each. ISG's original requirement to EDPSD was for 20 subject and 20 area codes.

3. Linkage

    (a)   Present periodic sets appear to be a detriment to satisfactory linkage and efficient retrieval.

    (b)   Devise method to relate periodic set 2 (subject/area/place name) and set 3 (keyword) with periodic set 1 (subject/area) to improve linkage and simplify retrieval.

4. Capability for periodic print-outs of keywords and area place names recorded in index records to provide dictionaries for input standardization.

Output

5. Machine print-outs of only those subjects, keywords or areas used in the request statement (or conversely provide the ability to suppress in any record those subjects, areas and keywords not used in the request statement).

6. Flexible output, i.e., the ability to call for specified fields in the record to printout with those not specified being suppressed; in other words release from the present rigid formats.

7. Expansion of clear text equivalents for data elements stored in codes (subject, area, security, dissem and CW control, occupation, organization, source of document, etc.).

8. Continuation of computer listing of the document numbers in sequence as an ordering device.

9. Charge cards for document references on listing.

10. Choice of output product such as interpreted cards.

11. Listings of "direct hit" references after analytical screenings to provide the user with a bibliography.

12. Continuation of the printing of appropriate security and dissemination and CW caveats on top and bottom of all pages of listings. *+ pagination*

13. Capability to sort 25-character document numbers through entire nomenclature (SRT & Fin Intell) to provide listings of documents arranged by subseries and other elements.

14. Capability to sort 10-digit document numbers so that all numbers are ordered by year and not interspersed.

15. An additional print space between fields of the 25-character document number used for Fin Intell and some information reports.

16. A simple query language as close to natural language as possible.

17. A simple query syntax.

18. A simple means of ordering an output format. (At present a separate query is required to designate the format.)

19. Table look-ups to derive request terms rather than rewriting terms for recurring requests, e.g., matching geographic dictionary entries to geographic coordinates to be searched.

20. An acceptable turn-around time as the data base grows and the ability to interrupt the activity for "crash" requests.

21. "Canned queries" for recurring requests; such queries could be kept in a query catalog maintained by SSG, and ISG would call for the query by its catalog call number; "canned" queries would require updating so flexibility for automatic up-dating would need to be programmed.

Maintenance

22. Capability to make corrections, deletions, etc., in blocs rather than record by record.

23. Continuation of daily error lists of subject file input.

24. Ability to correct daily input errors in natural language, without writing maintenance orders in machine language.

25. Periodic counts of subject codes by area and by code to assist ISG in combining and opening subject code schedules.

Long-Term Requirements (over one year)

1. Remote Consoles -- CRS's on-line requirements are not firm at present. There has been no demonstrated need for querying existing files. ISG will experiment with specialized files through the installation of remote terminals in USSR and FEPAC Divisions. It is anticipated that CRS will always have the need for a basic batch processing file management system regardless of what is undertaken in the on-line area. Specialized files involved in trial applications would include Bloc Travel; Personality Affiliation; Soviet Elite, and Soviet and Chinese Leader Appearances. Some experimentation will be undertaken to test the use of remote terminals as input and maintenance devices to the AEGIS subject file, the Soviet and Chinese Leader Appearances File and to the Bloc Travel File.

2. Continue capability to assist ISG in preparing machine indexes to various ISG biographic directories.

3. Experimentation with full text document storage and text searching.

4. Conversion of MTST products for computer processing and searching.

5. Acquisition of tapes and programs from non-CIA machine files for ISG conversion, searching and manipulation.

6. Experimentation with automatic book and periodical charge-out systems.

7. Synonym tables for selected words.

8. Variable record length for title and area place name fields.

9. Upper and lower case printing.

10. Direct input to computer records to replace page reader or cardpunch.

Distribution:
    Orig. & 1 - Addressee
          1 - DD/CRS
          1 - C/SSG
          1 - C/EDPSD
          1 - SA/IO
          1 - ISG Circulation
          2 - file