EPA/ITAS

044-5-3
September 11, 1989

Agency Automated Document Storage and Retrieval Requirements Analysis

Draft Procurement Analysis

Prepared for

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
NATIONAL DATA PROCESSING DIVISION

INFORMATION TECHNOLOGY ARCHITECTURAL SUPPORT
CONTRACT NO. 68-W8-0083

Prepared by the Viar Team

Viar and Company
300 North Lee Street
Suite 200
Alexandria, Virginia 22314

TABLE OF CONTENTS

SECTION 1 - INTRODUCTION

On July 8, 1988 the EPA issued RFP number W802625-A3 for "Image Processing Systems." A requirements analysis was subsequently performed to identify EPA automated document storage and retrieval requirements.

This procurement analysis will present a number of issues raised repeatedly during the interviews conducted for the Agency Automated Document Storage and Retrieval Requirements Analysis. Section 2 lists these issues and summarizes how the Image Processing System (IPS) RFP addresses each issue. In addition, Section 2 also includes recommended enhancements to the RFP.

SECTION 2 - AUTOMATED SYSTEM ISSUES

2.1 AUTOMATED SYSTEM ISSUES

The interviews conducted as part of the Agency Automated Document Storage and Retrieval Requirements Analysis repeatedly raised a number of issues as concerns with the existing systems and requirements for an automated system. These issues are:

- The ability to perform text searches;
- The desire for a cost-effective records management solution;
- The ability to change documents maintained in an automated storage and retrieval system;
- The ability to support simultaneous multiple access to a single document;
- The ability to provide user support;
- The ability to reduce document retrieval times, particularly to respond to Freedom of Information Act (FOIA) requests;
- The ability to reduce document storage space;
- The legibility of original documents to be maintained in an automated storage and retrieval system;
- The ability to convert current media to a new storage medium;
- The ability to support data exchange; and
- The ability to integrate with existing systems.

2.2 ABILITY OF IPS RFP TO ADDRESS AUTOMATED SYSTEM ISSUES

2.2.1 Items Not Adequately Addressed

2.2.1.1 Ability to Perform Text Searches

The ability to perform full text searches is a requirement which is not adequately addressed by the statement of work. The ability to translate scanned characters into ASCII format is optional; the IPS RFP specifies that documents are to be scanned, compressed, and stored in facsimile format. Therefore, the ability to perform full text searches will not be readily available in the environment defined by the statement of work.

The statement of work requires the ability to define keywords for retrieval. This capability will provide some ability for retrieval based on a program- or project-defined mechanism. The ability to use keywords for retrieval provides as much document retrieval flexibility as any computer-aided paper or microfilm system. However, the statement of work does not address the text search capability required by some program offices and applications.

2.2.1.2 Ability to Change Documents

The ability to modify documents was expressed as a clear requirement by several users. Section C.2.3.2.6 should be a requirement of the RFP and not expressed as "desired". In addition, the optional feature in C.3.3 for an Optical Character Reader should probably be eliminated and instead software which is capable of performing "optical character reader" functions on the already scanned documents should be a requirement of this RFP. This software should be capable of recognizing type as small as eight point and be able to recognize several common fonts.

2.2.1.3 Legibility of Original Documents

Several users expressed a requirement for document legibility and high resolution of document images maintained in any automated storage and retrieval system. In particular, users expressed concern over the legibility of documents scanned since many are not the original documents.

Users want to be able to get the best possible scan of a document by varying contrast and scan resolution, without increasing scanning times. This concern is valid and is addressed in the RFP, although the RFP does not specify any criteria for this capability. For example, scan resolution criteria could be expressed in half tone pattern selections and gray levels.

2.2.1.4 Ability to Convert Current Media

Some users want the ability to convert existing documents, which may be stored on microfilm or microfiche, to a new media, such as WORM. It is uncertain whether this capability exists with current technology. It should be specified in the RFP as a requirement or optional requirement so that respondents must address it.

2.2.2 Items Addressed

2.2.2.1 Ability to Support Simultaneous Multiple Access

The users interviewed did not express a requirement for simultaneous access by more than 20 or 25 people; however, this is probably due to a lack of understanding of the potential of the technology defined in the RFP. The RFP requirement is reasonable and proper, assuming potential users will eventually understand the benefits of remote user file access and as the local area network environment expands at the EPA.

2.2.2.2 Ability to Provide User Support

The RFP requires EPA personnel to be trained in order to provide additional training to other EPA users. It may be desirable to receive price quotes from the respondents to the RFP for additional courses on a task order basis.

2.2.2.3 Ability to Reduce Document Retrieval Times

Many users expressed concern about the time currently required to retrieve documents, particularly to respond to Freedom of Information Act requests. The access times specified by the RFP meet all identified user requirements. The potential for this system to considerably reduce the time required through a careful selection of keywords should provide a considerable time savings to the government and provide a more responsive environment to the public.

2.2.2.4 Ability to Reduce Document Storage Space

Users repeatedly described a lack of storage space as a problem which contributed to inefficiencies in locating and processing. Currently, the WORM technology provides for more efficient storage than paper or microfilm. The RFP can meet all identified user requirements.

2.2.2.5 Ability to Exchange Data

The RFP supports the ability to exchange data within a local area network and among local area networks and remote locations.

2.2.3 Items Requiring Further Analysis

2.2.3.1 Desire for a Cost-Effective Records Management Solution

Some users expressed concerns about the costs of any automated storage and retrieval system. This issue must be addressed on a case by case basis with a cost-benefit analysis and detailed requirements analysis performed in each case.

Potential users should be given an in-depth orientation on the capabilities of the IPS system and possibly be provided with technical assistance in performing the cost-benefit analysis. Often the true costs involved in existing systems are hidden and the potential benefits of a new technology may not be visible to nontechnical personnel.

2.2.3.2 Ability to Integrate Existing Systems

Some users want the ability to integrate a new media, such as WORM, with their existing systems, rather than converting all existing systems to a new media. This could be accomplished by either software or peripheral devices, but in-depth analysis must be performed to determine specific solutions for each application.

2.3 CONCLUSION

The Agency Automated Document Storage and Retrieval Requirements Analysis identified requirements for a number of EPA program offices and functions in terms of numbers of image processing systems, including those which have jukeboxes and those which do not. The required number of these systems was derived based on the number of documents needed to be stored digitally and the requirement for "remote systems" which can access a central storage site.

A separate issue is the number of required workstations for a given image processing system. The number of workstations to be used for scanning, printing, and/or viewing documents must be determined on an application-by-application basis as part of the in-depth detailed requirements analysis to be performed by each program office prior to acquisition of systems, as recommended by Section 7.2 of the Requirements Analysis. The information collected for the current analysis cannot

accurately estimate the number of required workstations on an Agency-wide basis. However, a review of the quantitative requirements as set forth in the RFP as of this writing indicates that 400 workstations and 100 host processors (20 Level I hosts and 80 Level II hosts) are to be acquired, or a 4 to 1 workstation-to-host ratio. It is quite probable that this number of workstations will not be sufficient to satisfy the Agency's overall requirement over the life of the contract, at least for the Level I systems which will feature an optical disk jukebox and storage capacity in the hundreds of gigabytes. For Level II systems, whose storage will be on standalone optical disk drives, a 4 to 1 ratio may be sufficient; however, it is highly recommended that the ratio of workstations to hosts for the Level I systems be increased to a still conservative minimum of 8 to 1 or higher.