34

N90-20658

ONBOARD SHUTTLE ON-LINE SOFTWARE REQUIREMENTS SYSTEM: PROTOTYPE

Barbara Kolkhorst (IBM)

Barry Ogletree (IBM)

INTRODUCTION

Late in 1987, the Spacecraft Software Division (SSD) of the Mission Operations Directorate of NASA's Johnson Space Center (JSC) in Houston asked IBM, as contractor for Onboard Shuttle Software (OBS), to investigate the problem of storing the existing Flight Software (FSW) requirements in an electronic form. These requirements define functions related to vehicle guidance, navigation and flight control and are thus critical to Shuttle missions. These documents, consisting of integrated text and engineering drawings, exist as many different documents residing at several NASA locations and were developed over approximately fifteen years as the Shuttle program evolved. The requirements should be accessible to the NASA community on-line; ultimately, automated requirements to code mapping should be available.

As a result, a small technical team worked in three phases to satisfy the NASA request. In the first phase, the team leader, several software requirements analyst's and a system engineer familiar with commercial product search techniques defined the problem to be attacked; this was documented as a request for information from NASA. In the second phase of the task, a solution for the problem was developed and an engineer experienced in electronic publishing systems was added to the team. Goals were developed to determine which solution would be proposed:

1. The requirements documents should be in electronic form under the central control of the Shuttle Avionics Software Control Board (SASCB) of NASA JSC.
2. Editing and publishing of the requirements should be under strict configuration control of the SASCB. On-line viewing is controlled by system security programs and the publishing system.
3. The solution should be a complete integrated solution which maximized the commercial software content to minimize development and maintenance costs of the system.
4. In addition, the eventual goal would be to provide a solution in which 'what is approved is published'. That is, what was approved by the SASCB had been submitted electronically and incorporated into the requirements data system automatically after proper approval; no rekeying of information would be necessary.

In the third phase of the project, a prototype was developed to prove that the proposed system could indeed be used on the Shuttle FSW requirements; several programmers were added to the team at this time.

This three-phase task was successful and provided a solution with very high commercial content which provided most of the function required. A prototype solution was demonstrated in November of 1989 to the Spacecraft Software Division (SSD) and to the NSTS Engineering Integration Office.

PROBLEM DESCRIPTION

The Shuttle FSW requirements documents consist of approximately 31,000 pages of integrated text and line drawings divided into roughly forty-five books averaging 650 pages each. The documents exist in several word processors and on paper at several NASA and contractor locations. Publication is disjointed across books and there is no consistent document architecture. Drawings are integrated into the documents using manual cut-and-paste methods. Modifications are proposed to these documents on a regular basis by many authors and must pass through an approval process controlled by the SASCB. Until the changes are approved, there is no hardcopy of the requirements documents. Only approved modifications can be added to the baseline document after a requirements writer has certified that all changes are correct. This results in a number of areas of concern.

First, due to the delay between submission and approval of changes and actual publication of a hardcopy version, the software developers are often working with changes plus outdated published requirements. Second, the requirements writer must also have access to the latest version of the baseline document for developing change requests. Since there is a time delay when modifications are being submitted and ultimately approved for publication, the requirements writer must work with outdated versions. Third, the changes are manually integrated into the baseline document for publication and here some transcription errors may occur.

Since requirements definition is critical in the process of maintaining space shuttle avionics software, the proposed system must address the areas of concern and provide ways to compensate for the evolutionary environment in which the software must operate. The needs are best satisfied by a host-based publishing system because these software requirements documents are organized in a book format, created by many authors, composed of information from numerous sources, published for many users, and require centralized configuration control.

Proposed Solution

The proposed solution includes initial document capture, storage, retrieval, hardcopy publishing, electronic distribution, security, change request disposition, and configuration management for the requirements documents.

The initial capture of the documents will be done by either scanning printed pages or through conversion of various electronic word processor formats to the system format. Scanned pages will be converted to text and image files by an intelligent recognition engine and custom software. Finally, the proposed system will provide the foundation for future interfaces to other systems for tracking Space Shuttle components.

As a further enhancement, application bridges to other NASA systems can be developed to connect the requirements document system to other Space Shuttle components and systems. It is also highly desirable that the system be integrated into the existing NASA computer software and hardware base.

Proposed Solution Rationale

It should be noted that alternative solutions were investigated. A solution was considered where the requirements documents were stored as scanned images with no modification capabilities. The existing process for document creation and modification could be used and a configuration control system could be built around this manual process. Neither of these two solutions would provide NASA with as much flexibility to manage and control the entire document process as would a publishing solution.

A host based solution was chosen over a work-station based solution because of the volume of documents to be managed, the security control required to protect access to and integrity of the documents, the greater variety of printers, terminals, and storage devices available for attachments, the ability to connect to the existing information network as a host system, and the capability of supporting a larger number of simultaneous end users. The proposed solution does take advantage of the power and flexibility of intelligent work-stations to download a section of the requirement documents, modify or print selected sections, and submit the modification as a Change Request. This proposed solution allows NASA to build a strategic electronic requirements document system now and for the future.

The hardware for the actual solution consists of a scanner capable of intelligent character recognition and the separation of images from text, all points addressable printers at both the workstation and host, an intelligent workstation processor, disk storage and an IBM compatible host on NASA's JSC Center Information Network (CIN). The software for the proposed solution consists of two parts: a host part (a publishing system with support for viewing and control of documents) and a workstation part (a desktop publishing product and some custom user interface software). In addition, there is software to allow documents on the workstation to be converted and transferred to the host, support for scanner operation, filters which convert documents created on other word processing systems to the host publishing system format, and software used to view the published documents. Security and configuration control are provided by either the publishing system or the operating system. See Figure 1 for a pictorial view of the system. Figure 2 describes the hardware and software defined for the solution which are included in the prototype system.

System Hardware

The publishing host (shown in Figure 3) consists of an IBM System 370 processor, magnetic disk storage, a tape unit, a disk controller, all points addressable (APA) printers, and terminal and communication controllers. (In the future, optical disk storage may be added to allow increased capacity.) It is proposed that NASA use or share an existing host hardware system (tapes, disk, terminal and communications controllers already in place) for this application.

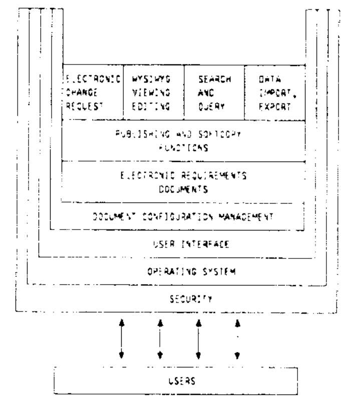

Figure 1. Modular view of the OBS On-line Software Requirements System

The magnetic disk storage will contain the active requirements documents and the application libraries. The application libraries will require approximately 1 megabyte of magnetic storage. The Onboard Shuttle Flight Software requirements documents will use 10 gigabytes of magnetic storage. The 10 gigabytes of storage will allow up to 200 active books (130,000 pages) to be maintained with on-line access. Frozen requirements documents will be archived on the optical storage jukebox. Sizing for the jukebox will be determined after initial implementation.

The page printer would be used to produce camera ready hardcopy documents. The IBM 3820 or 3800 printer is capable of printing complex pages consisting of text, graphics, and images. (However, any all points addressable printer capable of interfacing with the IBM Publishing System could be used to produce cameras ready documents.)

The recommended workstation for the Publishing Specialist, and the SASCB Administrator is an IBM Personal System/2 (PS/2) Model 80 (machine type 8580). The PS/2 has an 80386 microprocessor with MicroChannel architecture and 80 nanosecond memory. The workstation configuration consists of six megabytes of memory, a 115 megabyte fixed disk, a mouse, a 1.44 megabyte diskette drive, and a high resolution IBM 8514 display monitor.

	SOLUTION COMPONENT	PROTOTYPE
H	HOST (IBM COMPATIBLE ON CIN)	Y
A	HOST ALL-POINTS-ADDRESSABLE PRINTERS	Y
R	MAGNETIC STORAGE	Y
D	OPTICAL STORAGE	N
W	WORKSTATION (IBM PS/2 MOD 80 8514 DISPLAY)	Y
A	WORKSTATION (WS) ALL-POINTS-ADDRESSABLE PRINTERS	Y
R	SCANNER (CALERA CDP 9000)	Y
E	NECESSARY INTERFACES	N
S	IBM PUBLISHING SYSTEM	Y
O	DOCUMENT VIEWING SOFTWARE	Y
F	INTERLEAF PUBLISHING SYSTEM	Y
T	SCANNER SUPPORT SOFTWARE	Y
W	DATA EXPORT PROGRAM (TO HOST)	Y
A	WORDPROCESSOR TO HOST FORMAT FILTERS	PARTIAL
R	WS TO HOST IMAGE CONVERSION	Y
E	HOST TO WS PUBLISHING SYSTEM FORMAT CONVERSION	N

Figure 2. Further Comparison of Solution vs. Prototype

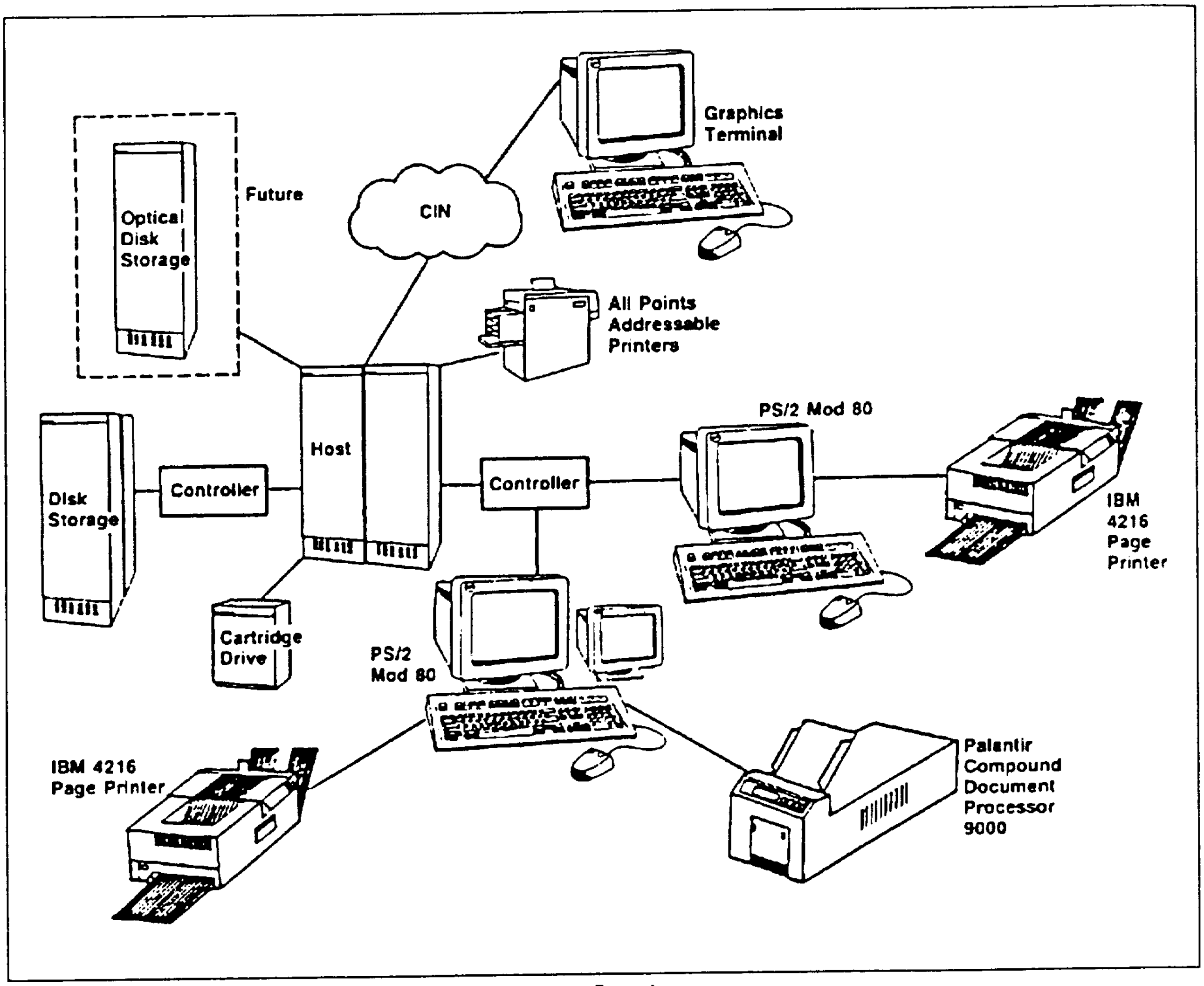

Figure 3. OBS Online Software Requirements Hardware Overview

Some workstations will have a scanner, such as the Palantir Compound Document Processor (CDP) 9000, and a workstation printer, such as the IBM 4216 Page Printer. The Palantir CDP 9000 will provide the capability to scan a document up to 8.5" X 14" in size at a resolution of 300 dots per inch; additional scanners using other resolution densities may also be attached to the Calera recognition engine. In addition, its recognition accuracy improves as more data is scanned. For workstations performing graphics modification, a second monochrome display may be required.

Graphics terminals on the CIN may be used to view documents and to write documentary change requests. These change requests must eventually be keyed into the source files on the host by a publishing specialist.

Scanners and Graphics Concepts

Scanners for graphics work can be categorized on the basis of several characteristics:

- Type of scanning mechanism (flatbed versus page feeder)
- Resolution (low of 75 dots per inch (dpi), high of 1500 dpi)
- Intelligent characteristics:
  - Text recognition (specific fonts versus any font)
  - "Tagging" (output of text to word processor formats)
  - Graphics handling (manual versus automatic)

For graphics work (especially the handling of integrated text and graphics), the ideal solution would be to let the scanner handle all aspects of the conversion process: feed the document into the scanner, separate out text and graphics, perform text recognition, and place the output into text and/or graphics files automatically. In practice, this is difficult to achieve:

- A major factor in this problem is the difficulty involved in identifying the start/end of graphics sections.

  One workaround is to let the user specify the location of graphics sections (this of course requires manual intervention).

- Inability for the computer to understand document "structure" (what figures go where).

  This might be due to the inability of most PC-based word processors to handle graphics (this is becoming less problematic with the advent of new word processors which support graphics manipulation).

Another area of difficulty is in text recognition. This is a "graphics-to-text" conversion where the scanner looks at the pattern of dots produced during scanning and makes a decision about the character represented by that pattern. Some scanners are unable to support this feature (requiring software to do the job); some scanners can only support a limited set of fonts. The most powerful machines perform "intelligent character recognition", recognizing any font style or size; perform spelling correction, marking unrecognized words for later correction; and decipher page layout automatically, distinguishing between text and graphics sections.

The Calera scanner used in this prototype has the features required to support this project. It is a sheet-feed (50 pages maximum) scanner with adjustable resolution (maximum 300 dpi), spelling dictionaries, and intelligent character recognition. In addition, it is represented as a "compound document processor", being able to scan integrated text/graphics documents; however, in its stand-alone mode it requires user intervention to designate graphics areas which are later placed in separate files. There is a board available from the same company which purports to handle integrated text/graphics automatically, but it was not available at the time the prototype was developed.

During the development of our prototype, we encountered several problems relating to graphics/text work:

- Recognition of special characters.

  Special characters (underline, super and subscripts, etc.) are difficult to scan properly.

- Registration of pages in scanner

  Straight lines in the source document become "stair-step" lines in the printed version. The workaround is to use a flat-bed scanner (less problems, but the stair-step effect is still noticeable). This also means production work is more difficult due to the necessity of handling each page separately.

- Loss of image "content"

  The scanning process produces raster files; the original image may have been produced by a vector process. This means that the information about object structure has been lost. There are programs available which can re-vectorize a raster-based image, but the robustness of the conversion is unknown.

  Hardware and software to do image to vector conversions for engineering drawings will be studied later in this project.

- Lack of consistent support for "standard" image formats.

  Specifications are defined for various image formats (TIFF, PCX, etc.) but some programs support only a subset of the available options. If the programs being interfaced do not understand the same set of image data, problems occur. A workaround is to understand exactly what is required by available programs and select those with matching capabilities.

- Storage requirements may be prohibitive for raster format files.

  Scanning an 8 1/2 by 11 inch page at 300 dpi results in about 8.5 Mbits of data (uncompressed). Certain formats (e.g. TIFF) can support various compression schemes to reduce the requirement for storage space. The resulting file may still require about 1 Mbyte of storage; vectorized files require far less storage.

System Software

The proposed software will be distributed between the host and the PS/2 Model 80 workstations. The prototype host software consists of the VM Operating System, IBM Publishing System and the necessary support services and utilities. The workstation software consists of IBM's Disk Operating System (DOS), Interleaf Publisher, and scanner support and image editing software. Custom code in both the host and workstation will facilitate transfer and configuration management of data files.

The host Publishing System software executes in the IBM VM operating system environment and is designed for corporate in-house publishing. An MVS solution is planned for NASA JSC use; it integrates the in-house publishing process from start to finish, including typeset-quality output of documents containing text, graphics and image. IBM's electronic publishing solution uses the host computer, workstations and printers to create, display and print documents.

The host Publishing system is an integrated set of software products (shown in Figure 4) and consists of:

- Publishing Systems ProcessMaster: A set of menus that controls the overall operation of the publishing system and provides a document control library management facility.
- Publishing Systems BookMaster: A powerful document creation application based on IBM's Generalized Markup Language (GML) that provides the tools necessary to create complex document formats.
- Graphical Display and Query Facility (GDQF): A package for viewing and editing CAD/CAM and other graphics data files on the host.
- Publishing Systems BrowseMaster: A series of utilities (provided in GDQF) to:
  - View merged text, graphics and image
  - View and crop GDDM Graphics Data Format (GDF) files and convert them to page segments
  - Import drawings from non-IBM CAD/CAM systems
- Publishing Systems DrawMaster: A menu-driven line art drawing package for creating graphics for use in publications.
- Image Handling Facility: A program to manipulate images for inclusion in documents.
- BookManager: An application for electronically viewing documents stored at the publishing host (SmartBook, an IBM internal product, is used in the prototype).

The workstation publishing software will be the IBM Interleaf Publisher. This standalone product executes under DOS on an IBM Personal System/2 model 80. The IBM Interleaf Publisher is a full-function, integrated publishing program.

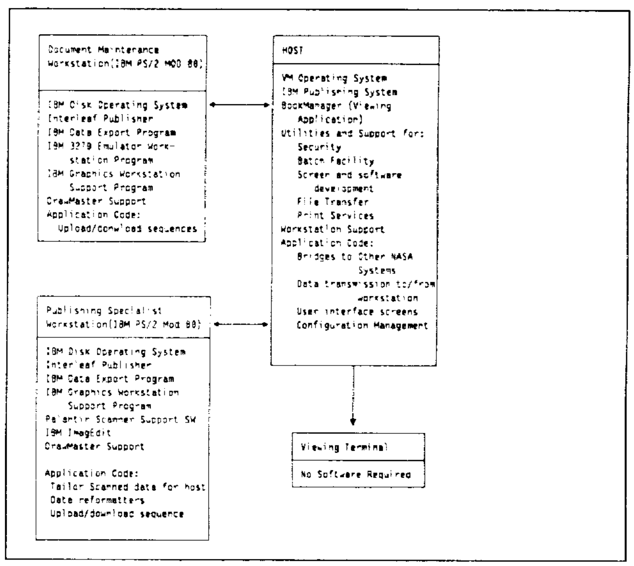

Figure 4. OBS On-line Software Requirements System Software Overview

Workstation-Based Functions

Scanner Support

Calera provides software with their scanner that assists the user in performing scanning chores. This software is divided into two types: applications (PAGEBLD, EDITPRO, TOPSCAN) and utilities (such as PDA2TIFF, DOCBUILD, and others provided by the scanner manufacturer to assist in custom software development).

PAGEBLD is the primary software used for scanning integrated documents. The scanner can be completely controlled from a full-screen (Windows-based) menu; functions to scan and read pages, save results, and work with the document are provided. The document is defined as having "zones" of information. Some zones contain text and are processed using text recognition techniques; other zones are graphics and are processed into CCITT-format files. To automate the process, zones may be predefined in "Zone Format Files"; this is useful when scanning must be automated, but it requires that pages adhere to a consistent format.

After the document has been processed by PAGEBLD, the next step is to use EDITPRO. This is a Windows-based application which helps the user find places where PAGEBLD and the scanner hardware had some difficulty recognizing text. Optional marks can be placed in the processed files; if present, these marks are used to drive the EDITPRO software. Functions are provided to move from mark to mark and the errors found may be changed while in EDITPRO (no need to use a separate word processor). After errors have been removed, files are created with the corrected information.

TOPSCAN is an application that provides scanning functions which understand most popular PC-based word processors and graphics formats. Text recognized by the system can be placed directly into a format understood by the user's word processor; graphics files are placed in TIFF format and can be used by any program understanding this file type.

Utilities

The scanner manufacturer supplies a set of utilities which assist the user in developing customized scanning applications. These utilities include standalone special-purpose programs that can build document files from text or image input, compress and decompress image files using CCITT Group 3 or Group 4 algorithms, modify text files to remove white space, and operate the scanner in a command-line driven (rather than graphics menu-driven) manner.

Graphics Manipulation

Manipulation of image files can be performed on the workstation or on the host system. For workstation-based image editing, IBM's ImageEdit is available. This program understands various file types (including TIFF) and provides editing to a pixel-level as well as the capability to draw lines, circles, and other basic shapes. It can produce TIFF files in both uncompressed and compressed forms.

Host-Based Functions

Because of requirements specified during the prototype definition phase, the major portion of the system resides on IBM mainframe computers. The environment (especially from a software and printer viewpoint) is considerably different from the personal computer environment; graphics formats are unique (GDF is used for vector files, IMG is used for image files). Image Handling Facility (IHF) and other programs in the IBM Publishing System are required to convert images to the format required for printing; this format (Page Segments or PSEG) is used because of the system which prints documents with imbedded images (Document Composition Facility or DCF).

Graphics Manipulation

The primary formats utilized on the host are:

- Vector-based

  GDF, CGM

- Raster-based

  IMG

Host software is available to manipulate both types of format. DrawMaster is a product which produces vector-based files (GDF and others); IHF is available to edit raster-based (IMG) files.

Viewing

Viewing of documentation is provided by two programs: BookManager (for text-based document reference with graphics support) and BrowseMaster (for publishing system specialists required to proof documents before printing). For the prototype, an IBM internal use tool called SmartBook was used to provide BookManager functions; it was the precursor to the BookManager software.

BookManager is the program of choice when users must refer to text and be able to browse figures which are present in the document. It operates by displaying the document in text mode (which means that users without a graphics terminal will be able to read the document) unless the user requests that a figure be displayed; the system then changes to a graphics mode and displays pictures specified by a user command. BrowseMaster is most useful to individuals requiring information about the layout of the document and who must provide error-free printing (as far as layout and appearance are concerned). It is used to provide a preview of the layout (a page image including margins and simulated text) so that those individuals responsible for printing the document can insure there are no major errors before submitting the job to the system printer. This method is similar to some PC-based word processors which allow the user to look at a page before printing it, resulting in savings of time and system resources such as paper.

Printing

Printers available on the host system range from line-based to laser-compatible (IBM's 3820 printer is the printer of choice). The 3820 used in the prototype is a host-connected printer capable of 240 dots per inch and a print speed of 20 pages per minute; since its resolution differs from the resolution available with the Calera scanner, a problem with image degradation occurs. This problem can be avoided in two ways: scan images and reduce them to the required size using the Publishing system, or use an alternate scanner (such as the IBM 3118) which is capable of scanning at the same resolution as the printer (240 dpi).

Summary

The prototype discussed in this paper was developed as proof of a concept for a system which could support high volumes of requirements documents with integrated text and graphics; the solution proposed here could be extended to other projects whose goal is to place paper documents in an electronic system for viewing and printing purposes. The technical problems (such as conversion of documentation between word processors, management of a variety of graphics file formats, and difficulties involved in scanning integrated text and graphics) would be very similar for other systems of this type. Indeed, technological advances in areas such as scanning hardware and software and display terminals insure that some of the problems encountered here will be solved in the near-term (less than five years). Examples of these "solvable" problems include automated input of integrated text and graphics, errors in the recognition process, and the loss of image information which results from the digitization process.

The solution developed for the Online Software Requirements System is modular and allows hardware and software components to be upgraded or replaced as industry solutions mature. The extensive commercial software content allows the NASA customer to apply resources to solving the problem and maintaining documents, rather than spending a large portion of the maintenance resources on custom software.

The actual conversion of scanned text and drawing images to a form which can be stored in a publishing system provides NASA with the capability to transfer any paper documents to editable electronic form for maintenance and update. As the various filters are procured or developed, documents which exist in other word processor formats may be added to the central files. The central repository may consist of magnetic storage for active documents and optical storage for documents which have been frozen in final format. This system may be used for storing and maintaining any documents consisting of integrated text and drawings.

This electronic base of information is suitable for future applications such as hypertext, where specific reference points in the documents are electronically linked to other documents, other parts of the same documents or note information. Additional search and query capability will also provide the NASA community with the ability to obtain information more rapidly than was ever possible with paper-based documents.

Definition of Acronyms

CAD. Computer Aided Design

CAM. Computer Aided Manufacturing

CGM. Computer Graphics Metafile

CIN. Center Information Network

CDP. Compound Document Processor (from Calera)

DOS. Disk Operating System

DPI. Dots Per Inch

FSW. Flight SoftWare

GDDM. Graphical Data Display Manager

GDF. Graphics Data Format

GDQF. Graphical Display and Query Facility

GML. Generalized Markup Language

IHF. Image Handling Facility

IMG. IMaGe Format

JSC. Johnson Space Center

MVS. Multiple Virtual Storage

NSTS. National Space Transportation System

OBS. OnBoard Shuttle software

PC. Personal Computer

PCX. PC Paintbrush Graphics File Format

PPM. Pages Per Minute

SASCB. Software Avionics Software Control Board

SSD. Spacecraft Software Division

TIFF. Tagged Image File Format

VM. Virtual Machine

WS. WorkStation