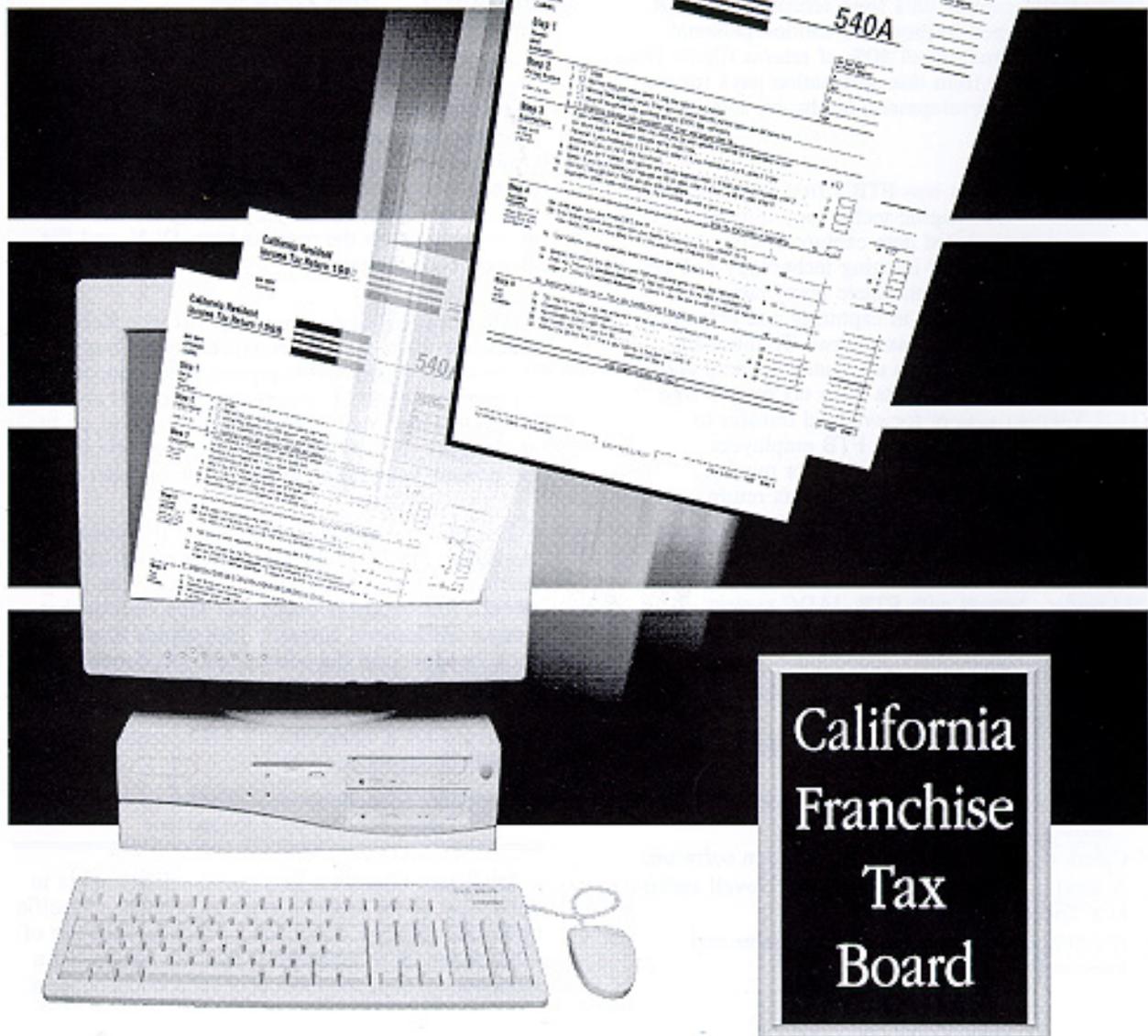


IMAGE ASSISTED



California
Franchise
Tax
Board

DATA CAPTURE

IMAGING

This guide describes the Franchise Tax Board's (FTB) Image Assisted Data Capture (IADC) system.

The IADC system uses imaging technology to electronically capture data from returns. The IADC system is designed to process 5 million personal income tax returns (about 40% of returns filed). The savings resulting from this automation pays for the IADC system's development, hardware, and software.

The IADC system was FTB's first step in a plan to eventually use imaging technology throughout the department. Automated data capture offers a big productivity payoff for imaging technology at FTB. IADC enables FTB to eliminate many labor intensive tasks involved in capturing and processing information from paper tax returns. Traditionally, key data operators using a computer terminal and keyboard manually key data from tax returns into FTB's Tandem system for eventual transfer to FTB's mainframe computer. FTB employees manually stamp a "document locator number" (DLN) and file date on each paper tax return. Imaging technology eliminates these tasks.

The IADC system was developed in 1993 by IBM under contract with FTB. IADC system components include:

- Kodak scanners;
- Compaq Deskpro 4000 computers using OS/2 operating system software;
- Compaq "pluggable" Ultra-Wide SCSI Raid Drives for data and image storage;
- Southern Computer Systems data entry software "IMAGEKEY";
- Calera intelligent character recognition software;
- A local area network (LAN) using Novell servers in a Token Ring environment; and
- A combination of customized programs and integrated software.

The IADC System Consists of the Following Processes

Station 1

Scanning and Data Recognition

All paper returns are received by FTB via mail. Those returns that are "scannable" are assembled and prepared for IADC processing. An FTB operator then feeds returns into one of FTB's scanners, which digitizes (takes an electronic picture of) each return and stores the imaged file for further processing. Using an ink jet printing mechanism, the scanner prints the process year, DLN, and file date on each tax return.

The scanning workstations consist of a Kodak Imagelink 900D scanner connected to a Compaq Deskpro 4000 personal computer. The scanners support resolution of images at 100 to 400 dots per inch (DPI) and operate typically at a rate of up to 3600 pages per hour. This rate includes printing of the process year, DLN, and file date on each return.

The personal computer monitor at the scanning workstation displays the image of a paper return for selected returns being scanned. Typically an image of each fifth return is displayed. This ensures that returns are scanned correctly (not upside down, etc.). Images from the scanner are transferred to and stored on the Compaq "pluggable" Ultra-Wide SCSI Raid Drives. All hardware items are connected via a Token Ring LAN operating on a Novell file server.

Intelligent Character Recognition

Intelligent Character Recognition (ICR) refers to the ability of the IADC system to recognize specific alpha and numeric characters in the digitized file of a tax return. ICR, although it can't be "seen" at a workstation, is at the heart of FTB's IADC system.

ICR uses line geometry (the size and layout) to identify the type of return (scannable Form 540A or Form 540 tax returns) from the scanned image, locates specific fields on the scanned image and converts the imaged data to character data. ICR recognizes the imaged data, performs various edits, and formats data for eventual transfer to FTB's mainframe computer. Data that is not recognized causes the image file to be referred automatically for correction within the IADC's Reject/Reentry or Exception Processing workstations.

The ICR subsystem operates as an "engine" that uses 9 Compaq Deskpro 4000 computers operating Calera recognition software.

Station 2 *Reject/Re-entry and Exception Processing*

If the ICR process does not recognize a few characters on a tax return, these unrecognized characters are sent to a key data operator for a quick, out-of-context correction. This consists of an operator viewing a "snippet" of imaged data at a Reject/Re-Entry workstation. The "snippet" shows the area on the tax return which needs to be interpreted by the operator to correctly define a character. Returns are automatically sent to exception processing if:

- ICR does not recognize the return type;
- ICR does not recognize a pre-defined percentage of characters;
- Pre-defined edit conditions are identified.

If needed, a Reject/Re-Entry operator can refer a return to Exception Processing.

The Exception Processing workstation monitor shows the entire image of the tax return, not just "snippets." An operator reviews the entire page and enters data from the image into a "data entry shell" designed to maximize speed and accuracy. The Reject/Re-Entry and Exception Processing workstations consist of 14 Compaq Deskpro 4000 computers. Reject/Re-Entry and Exception Processing use Southern Computer System's Image Key software, which is an industry standard data entry application that has been customized for FTB applications.

Station 3 *Administration*

The administrative functions allow an authorized operator to manage the workflow of the IADC system by controlling trucks, batches, individual documents; obtain workflow reports; and perform quality review activities. The administrative function is also used to transfer the return data to FTB's mainframe computer for further processing. The final administrative function is the purge of IADC return images.

The IADC system's administrative functions are provided from menus presented within an OS/2 Presentation Manager software program. Although the administrative functions (with the exception of "transferring data to the mainframe computer") could be performed from any workstation, 3 Compaq Deskpro 4000 personal computers have been designated specifically for this purpose.

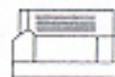
One of FTB's goals is to move towards paperless tax return filing. We are committed to using technology to meet this goal. While imaged returns are not truly paperless, we are working on expanding the Imaging Program to image all returns (computer prepared and handwritten) and store only the image. This will increase the speed and efficiency of document processing and reduce the need to store the paper copy.

IADC System Overview

Scan Stations: (2)



Line Printer



Database Server



27 GB of
RAID 5
Storage



Image Server



45 GB of RAID 5
Storage



IRM List
Manager



Admin Stations: (3)



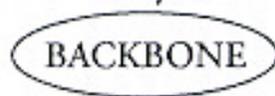
KDE Stations: (14)



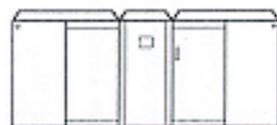
ICR
Stations: (9)



BACKBONE



MVS Mainframe



IADC LAN Segment