Progress Report
Office of International Health Programs (EH-63)
U.S. Department of Energy

Title of Project: 1.2a, Data Preservation

Principal Investigator: Dr. Nikolai V. Startsev
Urals Research Center for Radiation Medicine

Dr. Donna L. Cragle
Oak Ridge Institute for Science and Education

Period Covered in this Report: October 1, 1998 through March 15, 1999

I. Summary of Work

The purpose of this project is to implement a document imaging system for the Urals Center for Radiation Medicine (URCRM), Chelyabinsk, Russia. The document imaging system will have the primary purpose of preserving valuable medical records pertaining to individuals living in the Southern Urals region of the Russian Federation that have been exposed to radiation through operations of the Mayak facility and also through releases from accidents. These documents detail physical examinations, individual dose measurement, addresses, causes of death, and many other data items necessary for epidemiologic studies. The records presently reside in hard copy files without any duplicate copies. The document imaging system will provide the capability to: (1) scan documents of varying size, thickness, and condition; (2) provide image enhancement and clean-up capabilities; (3) index patient records and other documents using a minimum of four fields (which accommodate the Russian Cyrillic alphabet); (4) import field data (Russian Cyrillic alphabet) contained in existing computer databases created by URCRM to the document imaging system's data base management system for linkage with scanned document images; (5) store images to CD-ROM, WORM, or optical discs; and (6) perform data search and retrieval over a stand-alone local area network (LAN). The document imaging system will be operated and maintained by URCRM technical staff, who will receive training from technical staff at Oak Ridge Institute for Science and Education (ORISE), who in turn will have received training from the supplier of the system.

II. Milestones and Deliverables Accomplished During the Reporting Period

A. Choosing a vendor: The vendor is Document Access Systems (DAS), in Richmond, Virginia. A contract was signed on May 28. Equipment was to be delivered to Oak Ridge on July 28. All equipment was ordered to be compatible with Russian power requirements.

B. Delivery of imaging system to ORISE by vendor: The equipment was delivered to Oak Ridge on September 28. The system that was originally planned called for the HP 200 FX jukebox, which we learned (on August 23) was discontinued and replaced by the HP 400 EX. The HP 400 EX uses the new 5.2 GB capacity optical drives, instead of the previous 2.6 GB drives, with respective higher density media as well. This equipment change doubled the storage capacity of the imaging system. The higher capacity jukebox should meet the URCRM storage requirements for all four years of scanning and indexing, thus eliminating the need for ORISE to return to URCRM in order to upgrade the jukebox storage capacity.

C. Training of ORISE and URCRM principal investigators (P.I.) On use of the system: URCRM investigators planned a trip to coincide with the anticipated delivery date of August 28. When we learned that the system would not be delivered on this date, it was too late to cancel the trip. URCRM investigators met with the DAS staff in charge of software development and system integration to review progress and discuss software customization. DAS recently added a Russian-speaking software expert to their staff to integrate the system and to interface with the URCRM investigators. Additional software requirements were specified and agreed to during the site visit from the URCRM investigators. It was also discovered that DAS had not interpreted the original specifications correctly. This would have resulted in the development of a non-functional front-end user interface . Fortunately this misunderstanding was corrected at the expense of a four week schedule slippage.

When the equipment was delivered to Oak Ridge in September, it was demonstrated to the ORISE principal investigator. The equipment was set up in a conference room at Lockheed Martin's Data Systems Research and Development Unit, where a 220 power outlet was installed. It is anticipated that the equipment will remain in the conference room through March or early April for completion of testing of the few remaining software issues described below.

System Test Results

The document imaging system performance testing by Lockheed Martin indicated several software deficiencies that needed to be corrected by DAS. These deficiencies

were primarily related to the customized front end software requested by URCRM that added several capabilities not found in the commercial Optika software. The front end software, Imager, was interfaced to the Optika software via customized Visual Basic and C++ programming, which caused some problems initially. The deficiencies included the following:

1. Problems with different techniques for cropping of images and saving them in batch mode.
2. Problems with the front end imaging program, Imager, sometimes crashing for no apparent reason.
2. Inability to import data from the URCRM database.
3. Inability to scan images at 200 or 300 dpi, crop them, and print them with the correct scale.
4. Inability to create CD-ROMs with patient record information for distribution by URCRM to researchers and other clinics.

Lockheed Martin and the imaging vendor, DAS, collaborated on devising solutions for correction of the above problems. Software program revisions were developed by DAS, e-mailed to Lockheed Martin, tested, and results were immediately fed back to DAS. This iterative process has resulted in the resolution of all software issues except for two: 1) 300 dpi scanning and printing, and; 2) CD-ROM distribution. These are currently being addressed by DAS and will be tested by Lockheed Martin as soon as completed.

Lockheed Martin, as part of the testing process, developed a detailed user's manual outlining the important features of the imaging system. The manual contains screen shots of important steps in the scanning, optical disk formatting, backup, and index processes. In addition, the document structure hierarchy for the Patient Folder was transmitted by URCRM for incorporation into the Optika folder and indexing user interface. The preliminary structure (Russian Cyrillic) was input into the Optika folder structure.

D. Acceptance of system by URCRM P.I.: During the August visit to the DAS offices in Richmond, the URCRM P.I. reviewed the software specifications with the DAS software developer. Some modifications were requested and agreed to. The URCRM investigators traveled to Oak Ridge after visiting DAS and after discussions with ORISE investigators and subsequent conference calls with DAS, the URCRM P.I. said that the system was acceptable. This was followed up by several e-mails, in which URCRM transmitted additional guidance and information.

E. Shipment and installation of equipment at URCRM: Shipment was not accomplished in FY98 because of the delay of delivery of the system to Oak Ridge. Resolution of the software problems noted above, combined with political problems associated with the delivery of the equipment to the

Russian Federation, has further delayed equipment delivery. We anticipate that shipment will occur some time between April and June of CY 99. All items has been cataloged and tagged for shipment. In addition, the necessary paperwork for CRDF, the approved shipper, was prepared and approved.

F. Training of URCRM staff: URCRM staff were not trained in FY98 because of the delay of delivery of the system to Oak Ridge and its subsequent delay in being shipped to URCRM. The training may be facilitated by the addition of a Russian-speaking employee to the DAS staff. We are currently negotiating his travel to URCRM for the installation of the equipment and training of URCRM staff. The situation has been complicated by the fact that he is no longer is employed by DAS. Negotiations for his travel and expenses will be conducted with his new employer.

G. Commencement of record scanning and indexing: This milestone was not met in FY98. It is anticipated that this activity will begin approximately two weeks after the delivery and installation of the system at URCRM.

F. Milestones to be accomplished in FY99:

1. Shipment and installation of equipment at URCRM
2. Training of URCRM staff
3. Commencement of record scanning and indexing
4. Completion of protocols for scanning (to include quality control and quota requirements)
5. Six month equipment and progress check by ORISE investigators

III. Other Relevant Information, Including Relevant Trip Reports, Obstacles to Completion of Work Outlined in FY Work Proposal; Unexpected Costs; etc.

It is anticipated that ORISE involvement in this project will not be necessary after the six month progress check following the installation of the equipment. The URCRM P.I. will be responsible for setting and enforcing a weekly quota for scanned records. He will also be responsible for ensuring the quality of the scanned images and for development of procedures to be followed by the scanning and indexing operators.

All equipment will be licensed in Europe thus providing technical support to URCRM from European offices. Some of these offices are in Moscow. The DAS vendor has prepared repair kits with extra parts to be shipped with the equipment. These parts were chosen to cover the most likely failures (i.e., bulbs,

movable parts, etc.). DAS is available for customer support by e-mail.

Cost estimates for FY99 are higher than in the original proposal because of the addition of Al Klein (as requested by the executive committee review) and addition of a technician to travel to URCRM for the installation of the equipment. The technician has installed computer networks for other DOE projects contracted to Lockheed Martin Energy Systems in foreign countries as far away as Saudi Arabia. We feel that his network experience in foreign locations is a valuable asset to the completion of this project.

The ORISE P.I. is included in the cost estimate for the installation trip and the six month evaluation trip in FY99. If these two trips are removed from the budget along with the time allotment for the trips, the FY99 cost estimate will be close to our original estimate.

IV. Publications and Preprints

None. (There may be potential for publication of a lessons-learned paper for an imaging conference in the latter part of the year)