Introduction

This report documents the progress achieved by the Declassification Productivity Research Center (DPRC) during the second quarter of its third year of operations under DOE Grant # DE-FG01-95IS50316. Three quarterly reports and one annual progress report have been submitted per year since the beginning of the Grant in September 1995.

We have made very good progress on our projects this quarter, and have also begun several new projects, each with potential benefits for the declassification community. We experienced significant delays in obtaining the computer equipment necessary to carry out the FIDUL project. A workaround was developed and we are now progressing on that project as well.

The DPRC has apparently become a familiar, respected member of the declassification community because of its efforts and successes in the DPMC’s Automation Working Group. DPRC provided very useful service to DOE and the other agencies of the Government with its lead role and coordinating efforts in delivery of the AWG Electronic Document Interchange Standard (EDIS), and the associated (DPTC developed) (1) Document Interchange System (DIS) and (2) Document Review/Redaction System (DRS). These two PC based application systems are also being offered free of charge, along with several other government-developed computer application programs, to smaller agencies which do not have the resources to purchase or develop such systems themselves to support their declassification activities. We have also moved into a bigger laboratory facility at the GWU Virginia campus, which provides the DPRC with more space for expansion and for setting up the long envisioned DPRC Declassification Test Laboratory.

Finally, we also helped to develop a World Wide Web-based design concept and an operational prototype database system for one-stop coordination of multi-agency equity reviews. Dubbed the Equity Notification Database (or END), this system is being developed to resolve difficulties being experienced by agencies engaged in and responsible for the review of documents with multiple equities. This project, however, represents an over load on our current sponsored personnel and computer resources, as described in this progress report.

Background

The DPRC was established at GWU as an independent, world-class research capability and computer facility to support the DOE Declassification Productivity Initiative (DPI). The original intent of DOE was to provide seed funding for the DPRC, which might be supplemented by other interested agencies as opportunities for research were identified and projects funded. The goal of DPI is to increase the flow of unrestricted government information to the public. To this end, the work at GWU involves both basic and applied research in the areas of (1) system-level declassification process analysis and modeling, (2) development of computer systems to automate declassification processes, including text analysis and interpretation, (3) coordination/integration of new technology among into the processes, and (4) development and promulgation of interoperability and document transfer standards.
DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.
Accomplishments in the first quarter of FY98

I. On-Going Projects

1. The **AWG Electronic Document Interchange Standard (EDIS)** is available for PDF download on the DPRC web site. It represents combined efforts of the AWG members, and provides for the electronic transfer of documents among Federal agencies for coordinate reviews. The success of EDIS is an essential element for reaching the goals of EO 12958. Also, this project represents the most significant result of the AWG after over three years working together on community-wide issues and problems.

2. The **EDIS Tests**, which were planned last quarter, were begun this quarter. A comprehensive (Draft) EDIS Test Plan was developed by the DPRC and submitted to DOE for comments. The first phase of the test entailed a training session on the EDIS standard, and on the use of the DIS and DRS computer systems. This training, which was carried out by the DPRC, was hosted by the USIA using its NT networked computer system. Following user training, an exercise of the EDIS and computer systems for the case of unclassified document transfer was initiated. Several member organizations of the AWG volunteered as either Organizations of Primary Responsibility OPR’s and/or Organizations of Secondary Responsibility OSI’s. A review and lesson-learned session was held at the March AWG meeting. Several changes have been recommended for the DIS and DRS, as well as for the documentation. These will be carried out and the unclassified tests completed before moving on to the classified test situation. As reported last quarter, testing the EDIS is a critical activity for the AWG this year. The EDIS is intended to support necessary transfers of equity documents to meet the EO12958 schedule for April 2000.

3. Our **Voice Activated Guidance System (VOGUE)** project for the Federal Intelligent Document Understanding Laboratory (FIDUL) got off to a slow start this quarter when organizational and financial changes at FIDUL precluded the delivery of the computer equipment promised us by the Government in the SOW for VOGUE development. After several months of unsuccessful attempts to deliver the computer equipment, FIDUL and The DPRC agreed to simply convert some of the FIDUL funds for salaries to equipment. This apparently simple initiative took over 60 days to implement. Meanwhile, we continued as best we could using much slower equipment lent us by FIDUL. Voice activation makes heavy demands on the computer CPU and available RAM, so that we were not able to do very much until the new equipment was finally delivered towards the end of March. We are moving ahead at full speed now and hope to complete the effort described in our SOW under FIDUL before the end of the next quarter.

4. The DPRC continued to host and attend meetings of the **Automation Working Group (AWG)** of the multi-agency Declassification Productivity Management Council (DPMC) at the GWU Virginia Campus this past year. Our Virginia facility is adequately equipped and conveniently located for this service. Two AWG meetings were held at GWU/VA this winter.

5. The DPRC maintained and improved its **DPRC Internet Web Site** at (http://128.164.20.200/) this quarter. It serves as a communication site for the declassification community. We intend to study the site traffic this coming quarter to be able to assess the utility of the site and to be able to propose improvements.
6. We also continued to maintain and operate the DPI Program Web Site at (http://128.164.25.200/dpi) to serve as a point source for all DPI information. This web site is being well received by the community, as it contains most of the DPI program materials in readily accessible form and format. Many favorable comments have been coming in to describe its look, feel and overall effectiveness for information dissemination.

7. Good progress was made on our OCR Corpus Development project. We acquired copies of (1) the Characterization System---QUAL software system from Mike Cannon, and (2) the Document Research CD ROM on OCR projects, databases and tools from the University of Washington. Both systems have been installed and booted up. Operations to understand and use their capabilities are underway. Our plans for the coming months are to build a catalogue of sample pages representing the "score" developed by the Cannon’s QUAL program, and to get familiar with the OCR scoring system used on the UW CD ROM, prior to attempting to use these systems for building corpora or evaluating OCR performance.

8. A Winter issue of the DPRC newsletter, The Communicator, was published and distributed widely, both in paper and over the DPRC web site. The Communicator is now reads over 400 persons. We will be conducting a reader survey of the Communicator during the Spring to measure its utility and use in the declassification community as well as to solicit recommendations for improvements.

II. New-Start Projects and DPRC-Related Developments

9. Research was begun on the development of an Ultra Structure Theory (UST) Intelligent Engine for Web-based distribution of previously released government documentation. Dubbed the Intelligent Information Management and Retrieval (IIMR) Project, it is being carried out by one of the DPRC graduate fellows, Youngsuck Oh, as his D.Sc. dissertation research project. There are several reasons why this project has meaning for DOE. First of all, we believe it will provide new understanding for the application of UST systems for automation of OD’s declassification operations. Secondly, the system addresses the longer term problem of distribution of released information to the public, a very significant problem which is mostly ignored these days in light of the more pressing demands for meeting the schedule mandated by EO 12958.

10. The DPRC also participated in the development of a web-based design concept for an Equity Notification Database (END). A prototype version of the system is being developed as time permits, to provide a clear understanding of what END is about, how it will look, and operate, and how it will help resolve difficulties associated with the management of multiple agency equities. The DPRC contributions to and involvement in END to date naturally lead to a great role in the areas of (1) prototype system development and test, (2) maintenance of END’s data quality and (3) development of intelligent capabilities for self maintenance of data quality at some future time. Unfortunately, we do not have sufficient funds to be able to carry on with this important effort, in light of the (above) long list of projects being carried out by the current DPRC staff.

11. We also developed and proposed a research program into three important aspects of the Duplicate Document Detection (D3) problem, which is considered to be causing
significant difficulties in some of agencies. Our proposal included (1) development of a
detailed specification describing the characteristics of the D3 problem in the agencies, (2) An evaluation and assessment of the technologies available for dealing with D3 problems and, by comparisons to the specifications defined in (1) a series of recommendations for future developments for improvements, and (3) Development of a comprehensive D3 Test Process, including a test corpus to standardize the anticipated tests to verify commercial and developmental claims for D3 systems. The CIA has expressed interest in funding the proposed research efforts this year.

12. The DRPC continued its involvement in the doctoral level (D.Sc.) research effort by one of our uncompensated doctoral fellows, Sameer Hussian, to develop a dynamic simulation model of equity interactions among multiple agencies engaged in FOIA and EO 12598 document reviews. Sameer is investigating the implications of the recommendations made by Senator Moynihan’s Commission on Secrecy in Government by modeling and contrasting decentralized and centralized declassification operations. Dr Scotti is the research director for this study, which is expected to be completed by the end of calendar 1998.

13. The DPRC Declassification Laboratory concept moved closer to becoming a reality this quarter with our move into new space at the GWU Virginia campus. DPRC’s new office and laboratory space occupies more than over 450 square feet, and has excellent network wiring and web access facilities to meet future research expansions. An article with photos will be written for the Spring edition of the Communicator to describe the new space and our future plans to the entire community.

Participants

The above efforts and accomplishments involved the joint efforts of the following group of seven (7) individuals at GWU, most of whom are only paid part-time (or not paid at all) for involvement in the DPRC.

Professor Richard Scotti, Principal Investigator (PI) and DPRC Director (1/3 full time)

Mr. Jeffery Long, Senior Research Associate and DPRC Deputy Director (1/10 full time)

Ms Carol Lilly, Mr. Youngsuck Oh, and Mr. Nadir Khan, (Full time) Graduate Fellows of the DPRC

Mr. Sameer Hussain, (uncompensated) Graduate Research Fellow of the DPRC

Note: Dr. Paul Prueitt was recently made an (uncompensated) Executive Research Fellow of the DPRC. Dr. Prueitt brings to the DPRC a new wave of creativity in the area of machine/computer intelligence. His interest and long-time involvement in advanced intelligent computer systems is already stimulating a lot of new ideas at the DPRC. Some of the projects being looked into by Dr. Prueitt include: (1) development of a solid theoretical foundation for the Ultra Structure Theory, (2) Computer-based induction, deduction and intelligence, (3) development of a solid design methodology for UST systems.
Report Number (14) DOE/IS/50.316--T7

Publ. Date (11) [1998.04]
Sponsor Code (18) DOE/NN, XF
UC Category (19) UC-900, DOE/ER

1332.15 not included in folder

DOE