

SUPPORTING STATEMENT

PREDICT System Information Collection

(OMB No. 1640-0012)

A. Justification

1. The PREDICT initiative of the Department of Homeland Security (DHS) Science and Technology (S&T) directorate facilitates cyber defense research and development through the establishment of distributed repositories of security-relevant computer and network operations data and making such data available for use by researchers. The PREDICT Coordinating Center (PCC) has established application procedures, protection policies, and review processes necessary to make this data available to the cyber defense research community. PREDICT has been operational since Fall 2008.

The purpose of this system is to:

- Provide a central repository, accessible through a Web-based portal that catalogs current computer network operational data.
- Provide secure access to multiple sources of data collected as a result of use and traffic on the Internet.
- Facilitate data flow among PREDICT participants for the purpose of developing new models, technologies and products that support effective threat assessment and increase cyber security capabilities.

The Homeland Security Act of 2002 [Public Law 1007-296, §302(4)] authorizes the Science and Technology Directorate to conduct “basic and applied research, development, demonstration, testing, and evaluation activities that are relevant to any or all elements of the Department, through both intramural and extramural programs.” In exercising its responsibility under the Homeland Security Act, S&T is authorized to collect information, as appropriate, to support R&D related to improving the security of the homeland.

2. The content of PREDICT is data gathered from researchers and persons associated with

PREDICT, such as data providers, data hosts, and PREDICT application review board, publication review board, and advisory board members. It also includes metadata regarding the datasets that are made available to researchers through the PREDICT program in its efforts to build products and technologies that will better protect America's computing infrastructure. The PREDICT program uses the data to track usage of the data, ensure compliance with operational policies and procedures, and to evaluate the effectiveness of the PREDICT program. This use of the data enables the PREDICT program to provide researchers with access to various types of datasets to use in their efforts to develop solutions to provide improved security to networks, applications, and data, which will benefit all computer users and help protect the homeland.

3. Using a secure Web portal, accessible through <https://www.predict.org/>, the PREDICT Coordinating Center manages a centralized repository that identifies the datasets and their sources and location, and acts as the clearinghouse and operational authority for access to and release of the data. All data input to the system is either keyed in by users (Data Providers) or migrated (via upload of an XML file). The interactive nature of the PREDICT portal and the manner in which it is programmed to generate communications with researchers and/or other documentation for PREDICT eliminates the need for additional documentation, forms, and communications to researchers using PREDICT.
4. DHS S&T has coordinated with other DHS S&T divisions and is aware of research activities sponsored by other agencies and has found no duplication of efforts in the collection of the requested information, and there are no similar forms currently available that can be used for this system. DHS is not aware of any other duplications outside of this agency.
5. This collection assists small businesses or other small entities or individual researchers because it streamlines the information collection process for persons interacting with PREDICT, and it helps them electronically track interactions with PREDICT and history

of usage of PREDICT datasets.

6. If the information is not collected, DHS S&T will be unable to fulfill the objectives of the PREDICT system to provide secure access to multiple sources of data collected as a result of use and traffic on the Internet or facilitate data flow among PREDICT participants.
7. The special circumstances contained in item seven of the supporting statement are not applicable to this information collection.
8. By notice in the **Federal Register** on June 14, 2010 (75 FR 33629), DHS S&T notified the public that it was requesting comments on this information collection. The notice allowed for a 60-day public comment period. No comments were received. DHS S&T then by notice in the **Federal Register** on August 30, 2010 (75 FR 53705) notified the public during a 30-day public comment period. No comments were received from the public during either period.
9. DHS S&T does not provide payments or gifts to respondents in exchange for a benefit sought.
10. The security safeguards for the system shall meet the policy requirements set forth in the PREDICT System Security Plan (SSP) and its implementation manual and/or regulation. All systems are subject to monitoring consistent with applicable laws, regulations, agency policies, procedures and practices. PREDICT follows the Privacy Act of 1974 (Public Law 93-589), which mandates that personal information solicited from individuals completing Federal records and forms be kept confidential. PREDICT's Privacy Threshold Analysis (PTA) was approved April 2010 by the DHS Privacy Office and was determined the system was privacy sensitive. In accordance with the privacy ruling, a Privacy Impact Assessment (PIA) was drafted/approved to cover the system as well as an existing System of Records Notice (SORN) – DHS/ALL-002. In addition, PREDICT

will be operated in accordance with the E-Government Act (P.L. 107-347), December 2002 and the Federal Information Security Management Act (P.L. 107-347, Title III), December 2002.

11. There are no questions of a sensitive nature in this information collection.

12. Estimated Annualized Burden Hours and Costs

Form Name / Number 100XX	Account Request Form 10029 (12/07)	Request a Dataset Form 10032 (12/07)	My Datasets Page 10033 (12/07)
No. of Respondents	45	15	30
No. of Responses per Respondent	1	1	1
Avg. Burden per Response (in hours)	.25(15 minutes)	.25 (15 minutes)	.75 (45 minutes)
Total Annual Burden (in hours)	11.25	3.75	22.5
Average Hourly Wage Rate	\$100	\$100	\$100
Total Annual Respondent Cost	\$1125	\$375	\$2250

Form Name / Number 100XX	Memorandum of Agreement PREDICT (PCC) Coordinating Center and Researcher/User 10035 (12/07)	Memorandum of Agreement PCC and Data Provider (DP) 10036 (12/07)	Memorandum of Agreement PCC and Data Host (DH) 10037 (12/07)
No. of Respondents	15	2	1
No. of Responses per Respondent	1	1	1
Avg. Burden per Response (in hours)	1 (60 minutes)	.75 (45 minutes)	.75 (45 minutes)
Total Annual Burden (in hours)	15	1.5	.75
Average Hourly Wage Rate	\$100	\$100	\$100
Total Annual Respondent Cost	\$1500	\$150	\$75

Form Name / Number 100XX	Authorization Letter for Data Host 10038 (12/07)	Authorization Letter for Data Provider 10039 (12/07)	Sponsorship Letter 10040 (12/07)
No. of Respondents	1	2	45
No. of Responses per Respondent	1	1	1
Avg. Burden per Response (in hours)	1 (60 minutes)	1 (60 minutes)	1 (60 minutes)
Total Annual Burden (in hours)	1	1	45
Average Hourly Wage Rate	\$100	\$100	\$100
Total Annual Respondent Cost	\$100	\$200	\$4500

Form Name / Number 100XX	Notice of Dataset Access/Application Expiration 10041 (12/07)	Notice for Certificate of Data Destruction 10042 (12/07)	Amendment to Research/User Agreement 10060 (04/10)
No. of Respondents	15	15	5
No. of Responses per Respondent	1	1	1
Avg. Burden per Response (in hours)	.50 (30 minutes)	.25 (15 minutes)	.25 (15 minutes)
Total Annual Burden (in hours)	7.5	3.75	1.25
Average Hourly Wage Rate	\$100	\$100	\$100
Total Annual Respondent Cost	\$750	\$375	\$125

Form Name / Number 100XX	Notice of Data Access Expiration 10061 (04/10)		
No. of Respondents	15		
No. of Responses per Respondent	1		
Avg. Burden per Response (in hours)	.25 (15 minutes)		
Total Annual Burden (in hours)	3.75		
Average Hourly Wage Rate	\$100		

Total Annual Respondent Cost	\$375		
------------------------------	-------	--	--

Annual Reporting Burden and Respondent Cost: The total estimated ICR Public Burden in hours is 118*. This figure was derived by summing the total annual burden hours from all forms. The total annual number of respondents is 206. This figure was derived by summing the number of respondents to each form.

Public Cost: The total estimated annual public reporting cost is \$11,900. This figure was derived by summing the estimated annual respondent costs for all forms.

*Note – As stated in Question #12, the total estimated ICR Public Burden in hours is 118, however ROCIS reflects this number as 121. The reason for the difference is that ROCIS either rounds up/down the annual burden in hours for any number that is not a whole number (i.e. 3.5 would be rounded up to 4; 1.25 would be rounded to 1).

13. There are no capital or start-up costs associated with this information collection. There is no fee charge of for filing and of the information collection forms. Any cost burdens to respondents as a result of this information collection are identified in Item 14.
14. **Government Cost: Estimate annual cost to the federal government in relation to this information collection is \$2.5 million.** This cost includes about \$1.25 million for staffing costs related to the collection of this information for senior level directors, computer programmers, system administrators and mid level technical/engineers. Costs also include the purchase of servers and software for collecting/retaining/storing the collected information at about \$625k.
15. Since the launch of the PREDICT portal in March 2008 we have had three major refinements of functionality that automated data collection processes and increased efficiency of the work flow. The main enhancement to data collection via web pages is that the three pages originally used to add a dataset (DHS Form 10030) and update data provider and host information (DHS forms 10034 and 10033, respectively) were

consolidated into one form, the My Datasets Page (DHS Form 10033 – kept one of the form numbers). The information collected on these pages will not change but will be consolidated onto one Web page, greatly increasing the ease of use. Also, the proposed Annotate Dataset form (DHS Form 10031), which was never implemented, was officially deprecated from the requirements. There are no plans to implement this form. There was a need to add forms to the collection to support extension of the time Researchers have to apply for or use the data, or confirm they no longer need the data and to destroy them. Researchers are allowed access to data for 12 months, based on a detailed request that lists datasets they want. That list becomes part of their MOA, which includes a provision to either ask for an extension to continue using the data or prove they have destroyed it at the end of the 12 months. The new Amendment to Research/User Agreement (DHS Form 10060) and Notice of Access Expiration (DHS Form 10061) provide PREDICT with a way to follow up on this policy to ensure compliance. Workflows are planned to be implemented into the portal that will send Researchers expiration reminders beginning 90 days in advance. On the expiration date we send DHS Form 10061; if we do not receive an extension request, the Researcher has 30 days to prove he has destroyed the data. The PCC will contact a Researcher for 5 days after that to obtain the certificate. Failing that, we refer the matter to the PREDICT legal consultant for action. If a Researcher opts to destroy the data, that concludes a specific dataset request. They continue as users of the system and can apply for other datasets at any time. Their information remains active as long as they hold an account on PREDICT. If they close the account (or if our annual account audit determines they no longer qualify for an account) their information remains in our database but the account is deactivated so they can no longer access the portal. With the exception of forms 10030, 10033, and 10034 being consolidated into one and the creation of forms 10060 and 10061, no other program changes are being done at this time. Reporting adjustments in Items 13 and 14 reflect these changes as well as results of increased traffic to the PREDICT portal due to greater public awareness.

16. DHS S&T does not intend to employ the use of statistics or the publication thereof for this information collection.
17. DHS S&T will display the expiration date of OMB approval for this information collection. The current OMB number and expiration date is displayed in the upper right corner of the Web pages and PDF documents. The appropriate disclaimer and privacy notice are displayed in the footers of the Web pages and on the first pages of the PDF documents.
18. DHS S&T does not request an exception to the certification of this information collection.

B. Collection of Information Employment Statistical Methods

Not Applicable.