B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS

1. Description of the Activity

The respondents consist of active, cleared industry facilities under the National Industrial Security Program (NISP) who are approved for classified storage. The collection is deployed to one central point of contact for a facility.

These facilities are stratified using multiple quantifiable identifiers to create a scale of facility complexity with facility labels of AA, A, B, C, and D. These categories are taken from the National Industrial Security System (NISS) and reflect the size, scope, and complexity of a facility's security program with AA being the largest and D being the smallest.

The facility breakdown is constantly fluctuating due to new facilities (and those dropping out of the NISP), but the most recent count within the stratification is as follows:

- AA: 21
- A: 75
- B: 120
- C: 297
- D: 3,375

2. Procedures for the Collection of Information

As described above, the population is prime for stratified sampling. Facilities are set into the stratum by multiple factors directly related to this collection effort. Since facility categories AA through C are relatively small, Defense Counterintelligence and Security Agency (DCSA) will sample the entire population. For category D, DCSA plans to sample approximately 1/3 of that category's population. The five stratified sample groups are then as follows:

- AA: 21
- A: 75
- B: 120
- C: 297
- D: 1,125

Participants will be randomly selected from each stratum using a random sample generator. Based on the 2017 results, expected response rate to the collection is approximately 70% for each stratum (70% of ~1638 facilities=~1150).

NISS maintains many data elements pertaining to a facility's size, scope, and complexity and by using that data DCSA can further stratify most to least complex facilities within each stratum. From this, DCSA will apply a K-nearest-neighbor approach to impute missing values based on neighboring data points of a similar nature.

When missing values are accounted for, DCSA will calculate the mean and standard deviation for each stratum. Removing outlying data points outside of the margin of error (1.96 times the standard error), DCSA will re-calculate the mean for each strata, meeting agency needs. Using the average per stratum and multiplying it out by the total amount of facilities within the stratum, DCSA then has an accurate estimate for the total cost of the category. Taking the sum of total estimated cost for each of the five stratum DCSA will have the total estimated cost for the

population as a whole. This final number is what is reported to the Information Security Oversight Office (ISOO).

3. Maximization of Response Rates, Non-response, and Reliability

DCSA will pre-notify selected participants prior to deployment of the collection effort. During the data collection period, non-respondents will be sent e-mail reminders weekly to ensure timely completion of the collection.

As the collection itself only consists of two clear questions, the simplicity of the form will encourage a high response rate.

DCSA maintains a high working relationship with the respondent field and has consistently experienced high response rates on data collection efforts.

4. Tests of Procedures

This process has been tested over 15 years of deployment of this action. Sample sizes resulting from confidence levels and intervals have been adjusted to ensure maximum accuracy.

5. Statistical Consultation and Information Analysis

Primary POC: Chris Pirch 571-305-6241

Defense Counterintelligence and Security Agency (DCSA), HQ