

Supporting Statement for the Census of Users of the National Plant Germplasm System

OMBC Control No. 0536-XXXX

Part B: Collection of Information Employing Statistical Methods

March 9, 2017

Contents

Part B: Statistical Methods for Data Collection

Subsections:

- *Sample Selection and Respondent Recruitment*
- *Estimation and Mapping*
- *Degree of Accuracy Needed*
- *Adequacy of information collection for intended purposes*

Attachment G: Census Invitation for NPGS Users and Email Reminders

Attachment H: Invitation to Online Survey for Non-response Bias Study

Attachment I: Electronic Questionnaire for Non-Response Bias Study

Attachment J: Non-Response Telephone Questionnaire Script

Supporting Statement – Part B

PART B. Collections of Information Employing Statistical Methods

1. Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection method to be used. Data on the number of entities (e.g., establishments, State and local government units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection had been conducted previously, include the actual response rate achieved during the last collection.

The Census of Users of the National Plant Germplasm System (NPGS) will solicit data from all 6,009 individuals who requested germplasm (i.e. living tissue from which plants can be grown) for any of ten crops including beans, barley, cotton, maize, sorghum, squash, soybeans, potato, rice, and wheat from the NPGS during 2009-2013.¹ (These are the most economically important row crops for U.S. agricultural production, along with certain crops that originated in the Americas.) Of these institutional representatives, 78% percent are in the U.S. Twenty-two percent are from individuals outside the U.S. U.S. users come from four different groups: 1) commercial companies; 2) federal agencies; 3) state agencies, universities, and non-profit organizations; and 4) individuals without stated affiliation.

Individuals who request germplasm represent a wide variety of institutions, and the numbers of accessions they request vary widely both within and across institutional types. Two-thousand-one-hundred-eighty-three (46.67 percent) of U.S. users represent non-profit institutions, either federal agencies, state agencies, universities, or other non-profit institutions, and 824 (17.64 percent) represent commercial companies. An additional 1,667 users (35.69 percent) are unaffiliated individuals. However, U.S. non-profit institutions represent 73 percent of the *total numbers of accessions* (or samples) distributed to U.S. users, and U.S. commercial companies receive 22 percent of accessions distributed to U.S. users. Unaffiliated U.S. individuals represent 5 percent of accessions distributed to U.S. users. While the NPGS is an important source of germplasm for these individuals, they only represent a small fraction of the demand for NPGS germplasm.

Since a major purpose of the study is to analyze demand for NPGS germplasm, the accuracy of the results will be determined more by the proportion of total demand represented by the respondents who complete a questionnaire than by the proportion of users who complete a questionnaire. For the U.S., the expected response rate for users is 73 percent; however the expected percentage of total demand represented by responses to the survey is 80 percent (attachment D). For the entire Census (U.S. and non-U.S.), the expected response rate for users is 73 percent, but the demand-weighted response rate is 77 percent.

¹ Four percent of requestors did not supply email addresses, and searches could not locate them. Those requestors were eliminated from the census universe.

Should the demand-weighted response rate for the U.S. fall below 80 percent, a demand-weighted sample of 100 non-respondents will be contacted first by email and directed to an electronic survey to analyze the potential for non-response bias (attachment I). If there is no response, non-respondents will be contacted by telephone and asked the same set of questions. The expected response rate for the follow-up contact, defined as the proportion of the follow-up sample that complete a shorter set of questions by phone, is 97 percent (as per Frey 1996, *National Plant Breeding Study 1*, Iowa State University).

2. Describe the procedures for the collection of information including:

- **Statistical methodology for stratification and sample selection,**
- **Estimation procedure,**
- **Degree of accuracy needed for the purpose described in the justification,**
- **Unusual problems requiring specialized sampling procedures, and**
- **Any use of periodic (less frequent than annual) data collection cycles to reduce burden.**

Sample Selection and respondent recruitment

The Census of Users of the National Plant Germplasm System is a universe survey of all requestors who requested germplasm for ten crops over a five year period during 2009-2013. The crops are beans, barley, cotton, maize, potato, rice, sorghum, squash, soybeans, and wheat (consisting of economically important row crops in the U.S. and crops originating in the Americas). All representatives who requested at least one accession for at least one of these ten crops over the five year period will receive an invitation via email to participate in the survey from the National Germplasm Resources Laboratory of USDA. The National Germplasm Resources Laboratory is familiar to the recipients because it is the agency that provided the requested germplasm.

The census will be delivered to users of the NPGS both in and outside of the U.S. via an email invitation with an individualized link to the questionnaire (attachment G). SurveyMonkey (a survey design and collection software) will provide the electronic link to the questionnaire. The invitation will also provide a telephone number that respondents can call for assistance in completing the questionnaire online.

In the event of a demand-weighted U.S. response rate that falls below 80 percent, a modified stratified sample of 100 U.S. non-respondents will be contacted first by email, then if needed, by telephone, and asked to respond to a smaller set of questions (attachments H, I, and J). Four strata will be constructed based on classification of NPGS users by institutional type:

- U.S. commercial companies
- U.S. federal agencies
- U.S. state agencies, universities, and non-profit organizations
- U.S. individuals without stated affiliation

The sample will be allocated to each stratum in proportion to the number of demand-weighted non-responding users in the stratum. The non-response bias study will focus on U.S. users for several reasons. First, the geospatial analysis of germplasm use is based on U.S. requestors only. Second, U.S. users receive the majority of the samples distributed and meeting the needs of U.S. users (now and in the future) is the primary policy objective for the NPGS. Thus, the non-response bias study will focus on U.S. users and will provide additional information about whether non-response bias exists for this group. Nevertheless, the NPGS is interested in the ways germplasm is being used by international requestors, in part because it not only supplies genetic resources, it also receives them from international sources. Asking questions of international users will allow comparisons to be made with results of the earlier study of respondents who demanded germplasm during 1995-1999 that was mentioned in Statement A.

To determine which U.S. users are non-respondents, completed questionnaires will be compared to the administrative list of germplasm requestors for the ten crops, maintained by the National Germplasm Resources Laboratory. The National Germplasm Resources Laboratory maintains information on the institution associated with the requestor, as well as the location of that institution

Estimation and Mapping

Two data products are planned:

- Estimates of NPGS users' evaluation of their experience with NPGS germplasm distributions, including:
 - Usefulness of the germplasm received
 - Expected future use of germplasm
 - Most important traits sought in the germplasm received from NPGS
- U.S. maps indicating:
 - Geographic distribution of relative importance of traits sought in NPGS germplasm, in particular traits that respond to abiotic stress (that is, stresses such as temperature or rainfall extremes from the environment) and traits that respond to biotic stresses (that is, stresses due to the deleterious effects of pests or pathogens)

Degree of Accuracy Needed

This study is a census of all requestors who requested germplasm for ten crops over the five year period of 2009-2013. Information collected will be used by USDA to assess future demand for germplasm and likely trends in the traits sought in that germplasm. This will include information by specific crops (e.g., the use of landraces in soybean breeding, the search for biotic tolerance in wheat); the quantity of germplasm by type and purpose; institutional needs for germplasm (both public and private); and requestors' anticipated future use of the system. This information will also assess biological traits that are needed for adaptation to changing growing conditions.

Agriculture is highly geography-specific, given that growing regions vary by rainfall and temperature conditions, pest and disease pressures, and soil types. Accordingly, plant breeders work to develop unique varieties for different geographic locations. As a result, each requestor of NPGS germplasm is likely to have one characteristic--geographic location--which is unique and important to that institution's use of this germplasm, thus the use of a census. Moreover, it would be difficult to get adequate representation of the matrix of crops, germplasm types, and locations for some smaller crops (e.g., squash) without conducting a census of all germplasm requestors to the NPGS for the ten crops. Finally, a census is needed to have results comparable with the study of demand for NPGS resources during 1995-1999, which will allow us to assess new sources of demand.

This information request does not use sampling; it is a one-time collection.

3. Describe methods to maximize response rates and to deal with issues of non-response. The accuracy and reliability of information collected must be shown to be adequate for intended uses. For collections based on sampling, a special justification must be provided for any collection that will not yield "reliable" data that can be generalized to the universe studied.

To maximize response rates, the web-based instrument will be kept as simple and respondent-friendly as possible. The study instrument is based on a mailed paper instrument used in the 2000 study, "Demand for Genetic Resources from the National Plant Germplasm System." This earlier instrument was jointly developed by the International Food Policy Research Institute (IFPRI), the Department of Agricultural Economics and Rural Sociology of Auburn University, the National Germplasm Resources Laboratory of the NPGS, and the Economic Research Service. The mail instrument used in the 2000 study was administered by IFPRI and Auburn University and had a response rate of 35%.

Several measures will be taken to increase the response rate. Information will be collected via the internet rather than by mail. This data collection mode is more convenient for intended respondents and will allow for rapid follow-up with initial non-respondents. It is also consistent with the means by which NPGS users request germplasm from the system, i.e., via a similar web-based form.

The information collection will be cosponsored by the National Germplasm Resources Laboratory. The Laboratory and its senior staff are familiar to the recipients as it is the agency that provided the requested germplasm.

Recruitment protocol will include sending the instrument with an email from a senior staff member of the National Germplasm Resources Laboratory that solicits completion of the instrument. It also includes up to three reminder emails to non-respondents (attachment G).

Should the demand-weighted response rate for the U.S. fall below 80 percent, a non-response bias study will be conducted. A sample of 100 non-respondents will be contacted by email, and

if needed, by telephone, in order to analyze non-response bias. To provide further information about the nature of non-response, U. S. NPGS users have been stratified by institutional types:

- U.S. commercial companies
- U.S. federal agencies
- U.S. state agencies, universities, and non-profit organizations
- U.S. individuals without stated affiliation

This sample will be selected by stratified random sampling of the four groups of U.S. users. The probability of selection for non-respondents in each group will be proportional to the percentage each group contributes to U.S. demand for NPGS germplasm. These NPGS users will be asked to complete an abridged set of questions, focusing on whether they expect their institutions use of germplasm from the NPGS to increase, stay the same, or decrease. Responses from these individuals could be used to analyze whether there is any difference in their responses from that of those who completed the full questionnaire.

Adequacy of information collection for intended purposes

The adequacy of the information collection for the purpose of assessing demand for NPGS germplasm will be dependent on minimizing non-response bias, should it exist. Response imputation can be used to effectively remove non-response bias when unit non-response is random within subpopulations with the same values of known variables common to respondents and non-respondents (Little, 1988, *J. Bus. Econ. Stat.*, Vol. 6, No. 3). If non-response bias is sufficiently small, a response rate lower than 80% will still provide adequate coverage for the maps of relative importance of traits sought in NPGS germplasm, for an assessment of the usefulness of NPGS germplasm, and whether or not future use is expected to change. Should the demand-weighted response rate for the U.S. fall below 80 percent, the sample of 100 U.S. non-respondents, will provide additional information on the degree of non-response bias. If non-response bias is found to be an issue, caution will be noted in the reporting of results, specifically that the results from the original study may apply only to respondents (see below). This will be especially pertinent if non-response bias could lead to an overestimation of demand for NPGS germplasm.

As we wish to guard against the possibility of overestimation of future demand for NPGS germplasm, the basis of comparison in testing for non-response bias will be the results of the 2000 study, which asked respondents whether they expected their future use of NPGS germplasm to increase, stay the same, or decrease. In that study, 34 percent of U.S. users expected their future use of NPGS germplasm would increase, and 50 percent expected their future use would stay the same. Only 16 percent thought their use would decrease. Recoding this variable into a binary variable, 84 percent of respondents expected similar or increasing use; 16 percent expected use to decrease. If we assume that respondents to this census of NPGS users today are similar to those in the 2000 study, the priority would be to detect a difference of 15 percentage points or more between responders and non-responders in the likelihood of decreasing use, with 80 percent certainty. Since we are concerned that non-response bias might cause an overestimation of future demand for NPGS germplasm, the appropriate comparison is a one-tailed test of the differences between respondents and non-respondents for those who expect

use to decrease their use and those who do not. In other words, we wish to test the hypotheses that in reality 31 percent or more of non-respondents (16 + 15 = 31) are likely to expect decreasing use of NPGS germplasm. For large samples, the binomial distribution (decreasing use, non-decreasing use) can be approximated by the normal distribution. The statistical test is an independent sample, one-tailed z-test based on the normal approximation to the binomial distribution.

Given a stratified sample of 100 non-respondents, the sample size will result in a minimum detectable difference of 15 percentage points, for $\alpha = 0.05$, and 81 percent power for the hypothesis that non-respondents are more likely to expect decreasing use. If the difference is found to be above 15 percentage points, suggesting non-response bias, this will be noted as a caution in the reporting of results. In particular, results may apply only to census respondents. In that case, results from the population of *respondents* can still be compared with the population of respondents who demanded germplasm during 1995-1999. (We have no reason to suppose that the structure of non-response bias would be any different in these two groups.) Other information from these respondents (e.g. regarding traits sought in NPGS germplasm) will still be valuable to policy makers and NPGS managers; it is not available from any other source. Furthermore, information from respondents will contribute to the assessment, by USDA's Economic Research Service, of the role played by genetic resources in agricultural research and development, and will allow for better understanding of the availability of resources for the research enterprise.

4. Describe any tests of procedures or methods to be undertaken. Testing is encouraged as an effective means of refining collections of information to minimize burden and improve utility. Tests must be approved if they call for answers to identical questions from 10 or more respondents. A proposed test or set of tests may be submitted for approval separately or in combination with the main collection of information.

The questionnaire was reviewed by seven USDA Agricultural Research Service (ARS) scientists:

1. Dr. Todd Campbell, Research Geneticist, ARS, Florence, SC
2. Dr. Tommy Carter, Research Geneticist, ARS, Raleigh, NC
3. Dr. Michael Dowd, Chemical Engineer, ARS, New Orleans, LA
4. Dr. Sherry Flint-Garcia, Research Geneticist, ARS, Columbia, MO
5. Dr. Bob Graybosch, Supervisory Research Geneticist/Research Leader, ARS, Lincoln, NE
6. Dr. Dolores Mornhinweg, Research Geneticist, ARS, Stillwater, OK
7. Dr. Roy Navarre, Research Geneticist, Pullman, WA

Results from the pretesting and actions taken in response are summarized below:

1. *You might want to consider making all percentage estimates as pull down options. Some are and some are not. Some will type in a % sign and others will not.*

Response: Added percent signs after answers for all relevant questions.

2. *Differences between germplasm categories is a bit vague.....cultivars vs. advanced breeding materials? Maybe a quick definition is in order at first occurrence.*

Response: The categories were set by the NPGS. We have expanded the titles of the categories in the instrument to be more explanatory.

3. *The order of the germplasm groups changes during the survey.....specifically weedy relatives and genetic stocks. It's easier if always in the same order.*

Response: For questions involving germplasm types, the order is now always the same:

- Cultivars
- Advanced breeding materials
- Landraces
- Wild and weedy relatives
- Genetic stocks
- Material type unknown

4. *Choice of traits seem random. I know you can't list all traits for lack of space, but seems a little odd.*

Response: National Germplasm Resources Laboratory NGRL staff agreed that the choice of traits was suitable, especially because we want to capture resistance to temperature and moisture changes.

5. *"over next 10 years, expect use of accessions to change?".... do you mean how often I will be requesting germplasm, or if I ordered it for one purpose that I will use in another way in the future? I assumed the latter.*

Response: The former is correct. The instrument was reworded to “... do you expect your requests for NPGS accessions to change?”

6. *Primary professions are a bit odd.....These are not really professions. These are activities, which is what you asked in an earlier question. I would add "Geneticist" to the list as a default. And I would allow multiple options to be chosen.*

Response: We reworded the question to “What best describes your primary professions?” A single option was needed so responses can be compared with 2000 results. The categories were edited to read as follows:

- Acquisition/curatorial activities
- Pre-breeding/evaluation activities
- Breeding

- Genetics/molecular biology
- Education
- Farming
- Other (please specify)

We asked nine NPGS users about time required to complete the questionnaire:

1. David Weaver, Auburn University
2. Charles Brown. Brownseed Genetics, Inc.
3. Angela Davis, H.M. Clause, Inc.
4. Walter De Jong, Cornell University
5. Stephen Kresovich, Clemson University
6. Craig Morris, USDA-ARS
7. Jamie Sherman, Montana State University
8. Emmalea Garver Ernest, University of Delaware
9. Karen Moldenhauer, University of Arkansas

The mean time required to complete the questionnaire was 19.8 minutes. The assumed time for looking at emails without responding was assumed to be 1 minute, thus a total of four minutes for each non-respondent.

5. **Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.**

Statistical reviewer:

Mingshan Zheng (Mingshan.Zheng@NASS.USDA.GOV, 202-720-0830), USDA National Agriculture Statistics Service (attachment F: NASS review of Census of Users of the National Plant Germplasm System and ERS response).

Advisors for questionnaire design and data collection protocol

Quinn Sinnott (Quinn.Sinnott@ARS.USDA.GOV, 301-504-6072) and Gary Kinard (Gary.Kinard@ARS.USDA.GOV, 301-504-5951) of USDA's Agricultural Research Service/National Germplasm Resources Laboratory advised ERS the design of the questionnaire and methodology.

Data collection:

Quinn Sinnott of the National Germplasm Resources Laboratory will oversee the web-based data collection.

All other data management and analysis activities will be conducted by ERS in-house.

ERS project officer:

Kelly Day Rubenstein

kday@ers.usda.gov, 202-694-5515