

1SUPPORTING STATEMENT  
 UNITED STATES INTERNATIONAL TRADE COMMISSION QUESTIONNAIRE  
 Digital Trade in the U.S. and Global Economies, Part 2

Part B—Collection of Information Employing Statistical Methods

**1. Response universe, sample sources, and sampling strata**

The potential respondent universe includes all companies that conduct business in the United States and have primary NAICS categories corresponding to industries that USITC staff and industry experts identified as containing the highest concentration of firms that produce and consume digital goods and services. The sampling unit is the firm, which represents the highest level of an organization, rather than establishments or individual franchise locations. The universe has been determined to encompass 140,710 firms, of which, 10,000 will be sampled.

The potential respondent universe represents the sum of firms, net of duplicative records, identified in these data sources:

- A database derived from industry associations’ membership lists and industry-expert knowledge.
- Firms obtained from Bureau van Dijk’s Orbis database belonging to selected NAICS categories from seven industries most likely to take part in digital trade.<sup>1</sup> These industries and their selected NAICS groupings are:

**TABLE 1** Digitally intensive industries and selected NAICS codes

<b>Industry</b>	<b>Selected NAICS codes</b>
Manufacturing	323, 325, 3332, 3335, 3336, 334, 335311, 335314, 336, 3391
Wholesale trade	423110, 423120, 423430, 423610, 423620, 423690, 424320, 424330, 425110
Retail trade	441, 442, 443, 448, 454
Content industries	511 (except 5112), 512, 515, 51911, 51912, 51919, 711130
Digital communications industries	5112, 518, 51913
Finance and insurance	52
Services	4921, 5412, 541310, 541330, 541430, 541511, 541512, 541613, 54183, 5615

Firms entering the population via the industry association lists and/or analyst research will be stratified by size. Firms that are sourced from Orbis will be stratified by industry and size.

1. Industries include the seven sectors given above, such as manufacturing, wholesale trade, etc.

---

<sup>1</sup> Orbis is a proprietary global database with information on public and private companies.

2. Size is defined by employment. Although the main variable of interest is exports, there are no firm-level export data available prior to sampling. Employment is the most available measure of firm size in the Orbis data and is known to be highly correlated with exports.
  - a. For firms from the Orbis list, the smallest firms in each stratum are not sampled to reduce respondent burden and to improve the statistical properties of the remaining estimates.<sup>2</sup>
  - b. Small firms are defined as firms with fewer than 100 employees.
  - c. Large firms are defined as firms with 100 or more employees.
  - d. The 25 firms that did not have employee data were assigned “small firm” status, as we have found it very unlikely in previous surveys that large firms do not report employee statistics.
  - e. Three industries had very heterogeneous large firms, so a separate stratum for very large firms with 1,000 employees or more was introduced in the manufacturing, finance and insurance, and services sectors.

Table 2 presents the number of firms and the sample size in each of the 19 strata, following the methodology described below.

**TABLE 2** Number of firms in the population and in the sample

Data source and industry	Number of firms in the population				Number of firms in the sample			
	Small	Large	Very large	Total	Small	Large	Very large	Total
Industry list <sup>a</sup>	160	141	— <sup>b</sup>	301	160	141	— <sup>b</sup>	301
Orbis								
Manufacturing	19,069	4,202	499	23,770	611	326	227	1,164
Wholesalers	13,599	1,068	— <sup>b</sup>	14,667	578	1,068	— <sup>b</sup>	1,646
Retail	25,082	2,881	— <sup>b</sup>	27,963	696	1,195	— <sup>b</sup>	1,891
Information content	10,292	1,444	— <sup>b</sup>	11,736	385	1,444	— <sup>b</sup>	1,829
Digital communications	2,860	375	— <sup>b</sup>	3,235	253	375	— <sup>b</sup>	628
Finance and Insurance	17,302	4,028	624	21,954	445	320	490	1,255
Services	34,575	2,323	186	37,084	915	236	135	1,286
Total	122,939	16,462	1,309	140,710	4,043	5,105	852	10,000

Note: All figures subject to revision.

<sup>a</sup>The industry list covers all seven industries and is only stratified by size.

<sup>b</sup>Not all sectors required a separate stratum for very large firms.

<sup>2</sup> In all industries, the smallest firms are defined as those with fewer than 10 employees. Within each industry, cutoffs varied by the NAICS codes listed in table 1, and ranged from 10 to 50 employees. For example, within manufacturing, the cutoff for printing and related support services (NAICS 323) was 10 employees, whereas the cutoff for chemical manufacturing (NAICS 325) was 50 employees. The cutoffs were higher for some groups because small firms accounted for a much smaller share of employment and revenue in some codes than others, according to Orbis data. One caveat: firm size is not known for all firms in the industry association list, so sampling some very small firms from this list will be unavoidable.



Since this is the first large-scale survey of trade in digitally intensive sectors<sup>3</sup>, and no definitive list of all relevant firms exists, some coverage error is unavoidable. This error has been minimized by incorporating data from industry association lists and industry analysts. If a firm was found to be in both the association list frame and the Orbis frame, the observation in the Orbis frame was removed from the population in order to maintain uniqueness.

The sample size of 10,000 is the number of surveys that will be sent out and is based on what is needed for a statistically significant response, given historical response rates. Similar past surveys on remanufactured goods and used electronic goods have yielded response rates that ranged from 55–60 percent. Since the target populations for each of those surveys was rare, and the population of firms engaging in digital trade is not, we believe that the response rates will be higher in this case.

## **2. Collection of information employing statistical methods**

### **a. Statistical methodology for stratification and sample selection**

A stratified sample based on a simple stratification process is being implemented for this project. The goal of the stratification scheme is to develop a set of strata such that the variance of responses (such as level of employment, type of activities, and likelihood of undertaking digitally intensive trade) within each stratum is minimized to the extent possible. Stratification is also being used to include rare observations. Because no pro-forma reliable data exist on the size and scope of digital trade, the stratification scheme was based on the best judgment of industry and USITC experts.

The approach to stratification in this survey is based on a procedure designed to maximize efficiency of the resulting estimates, and hence reduce the total number of firms sampled. First, due to the size of the lists, the industry association and industry analyst lists are stratified by firm size alone and sampled at 100%. These lists allowed for the inclusion of firms that were engaged in digitally intensive activities, but unlikely to show up in the Orbis list. Second, firms identified by the Orbis database are allocated across size and industry strata using Neyman allocation based on the coefficient of variation of employment within each strata. Because this procedure may lead to sample sizes that are greater than the population per strata, the procedure is iterated after sample limits are placed on such strata until the number of firms sampled from each strata sums to the desired total, in this case 10,000.

### **b. Estimation Procedure**

Survey estimates will be based on weighted data. The weighting procedure will incorporate a sample selection weight, a nonresponse adjustment factor, and if necessary, a poststratification weighting factor.

---

<sup>3</sup> The closest analogue to this survey is Census' E-Stats report which measures e-commerce for four sectors (manufacturing, wholesale, retail, and services) based on results from multiple surveys that target each industry. Further information may be found at: <http://www.census.gov/econ/estats/>.

- *Sample selection weighting*: The selection weight factor will account for the probability of selection within a particular industry and size.
- *Nonresponse adjustment*: The nonresponse adjustment factor is designed to attenuate bias due to differential response rates across strata. See the section below on response rates for further discussion.
- *Poststratification weighting*: If necessary, a poststratification weighting factor will be used to attenuate bias due to sample frame noncoverage or omissions. Although the best effort has been made to obtain a representative sample of digitally intensive firms, this survey represents the first broad-based survey with the purpose of measuring the effects of digital trade barriers, so the distribution of firms across industries cannot be known with certainty in advance.

In order to produce population estimates and precision statistics about the estimate, the following equations will be used:

The formula used to estimate the population attribute of interest is found in equation 1. The precision statistics about the estimate are found in equations 2-3. Per standard notation, the total estimate from a stratified random sample  $\tau_{st}$  is given by

$$\tau_{st} = \sum_{h=1}^L N_h \bar{y}_h, \quad (1)$$

where  $h$  denotes an individual stratum,  $N_h$  equals the population of stratum  $h$ , and  $\bar{y}_h$  equals the average of the attribute of interest of the sampled items in stratum  $h$ . For example,  $\bar{y}_h$  could represent the average amount of revenue within each stratum.

The variance estimate for sampling without replacement is given by

$$\text{var}(\tau_{st}) = \sum_{h=1}^L N_h (N_h - n_h) \frac{s_h^2}{n_h} \quad (2)$$

where  $s^2$  equals the standard deviation of the attribute of interest within stratum  $h$ , and  $n_h$  is the sample size for stratum  $h$ .

Its standard error is given by

$$\text{Standard error} = \sqrt{\text{var}(\tau_{st})} \quad (3)$$

### c. Degree of accuracy needed for the purpose described in the justification

It is expected that it will be feasible to produce statistically significant results for the majority of survey items at the aggregate level at a 90 percent confidence level, both for the binary questions and for questions requiring responses in U.S. dollars. For example table 3 provides the maximum margin of error for a binary question, given alternative response rates. (These values are

probably conservative, given that response rates to recent USITC surveys have ranged between 55 and 60 percent.).

**Table 3** Margin of error for a 90% confidence interval<sup>ab</sup>

	Response Rates			
	30%	40%	50%	60%
Sample size	3,000	4,000	5,000	6,000
Standard error	0.91%	0.79%	0.71%	0.65%
Margin of error	1.50%	1.30%	1.16%	1.06%

<sup>a</sup>Assuming a standard deviation of 50% for a binary question.

<sup>b</sup>Sample size of 10,000.

Given the sample size per stratum, it is assumed that it will also be feasible to distinguish the responses across the largest industries within a 90 percent confidence interval. This degree of confidence is sufficient for the purposes described in the justification.

**d. Unusual problems requiring specialized sampling procedures**

No unusual problems were encountered.

**e. Any use of periodic (less frequent than annual) data collection cycles to reduce burden.**

This data collection is currently only intended to occur once, and therefore will not be repeated on a periodic basis.

**3. Methods to maximize response rates and deal with non-response**

**a. Maximizing response rates**

Commission staff will employ several techniques to increase the response rates of questionnaire recipient firms. Recipients will receive separate notices in the mail that (1) notify them that their firm was selected for the survey, (2) direct them to complete the survey, and (3) remind them, if necessary, to complete the survey before the deadline. Once the submission deadline has passed, firms that still have not responded will receive an additional reminder. Each of these communications will include a phone number and email address of a person who can help firms with filling out the questionnaire or answer their questions regarding the survey and/or study. Commission staff may also contact firms directly, via phone or email, to urge them to complete the survey and to answer any questions they may have regarding this information collection or study in general. Commission staff may also contact firms, via phone or email, to correct information or fill in incomplete responses, or solicit additional information about a response. The burden associated with follow up calls or emails is included in the total response burden amount.

In addition to pre-contact and follow-up, the questionnaire itself has been designed to be clear and succinct as possible to gather the specific material requested by Senate Finance Committee. (See discussion of testing below.) This clarity and brevity should reduce burden and improve response rates. The questionnaire will clearly point out that firms are obligated by law to respond. Finally, the ability to access, fill out, and submit the survey electronically may also increase response.

## **b. Accuracy and reliability of information collected**

The sample methodology has been designed to be as accurate and reliable as possible, based on Commission experience in past surveys. The sampling frame has been chosen to include firms in industries that are most directly involved in using and producing digital goods and services, and hence are also the most reliable reporters of trade in digital goods.

The size of firms included in the survey has also been carefully considered to improve accuracy and reliability. For each NAICS based industry included in the sampling frame, the Orbis data were used to ensure that the population included the large majority (generally over 90 percent) of total industry revenue.<sup>4</sup> Thus, the survey should capture nearly all significant digital trade, while excluding firms of the smallest size, for which a reliable population cannot be determined from the Orbis database.<sup>5</sup> On the other end of the size distribution, strata of very large firms were included when necessary (the manufacturing, finance and insurance, and services industries) to improve homogeneity of firms within strata, and hence to improve the reliability of resulting estimates.

Response rates in USITC surveys have recently been near 60%. The USITC will examine survey responses to detect and correct for any non-response bias. The team will first examine conditional response rates for groups of firms based on characteristics available in the data frame that are hypothesized to impact outcomes of interest. Any differences in response rates can be further investigated through logistic regression analysis, using firm characteristics (such as firm size, industry, location, etc.) as predictors, and whether or not a recipient responded to the survey as a binary outcome. If the results of the logistic regression indicate that one or more of the characteristics investigated above affects the propensity of a survey recipient to respond to the survey, then those characteristics will be examined to determine whether they are associated with differences in the outcome variables under study across the dataset of survey responses collected. As in previous surveys, if any sources of non-response bias are found, they can be controlled for by the development of weights, which can then be used in concert with weighting based on population stratification, in the extrapolation of results to the entire population.

## **4. Tests of procedures or methods to minimize burden or improve utility**

The Commission field-tested the questionnaire with companies from several industries. These testers provided feedback in areas such as availability of data, product coverage, definitions, and

---

<sup>4</sup> The portion of revenue contributed by firms with 0–4 employees was not included in the calculation.

<sup>5</sup> Previous studies have shown that Orbis data on the number and revenue of firms of this size are unreliable. For example, Orbis generally reports several times more firms of this size than are reported by the U.S. Census Bureau.

clarity of instructions. See the table in part A for the field-testers comments and the subsequent changes made to the questionnaire.

In addition to field testing, the questionnaire has been made available for public comment. Notice of the draft questionnaire was published in the Federal Register. It has also been extensively reviewed within the Commission and outside by survey consultants hired by the Commission. Industry analysts and economists have reviewed the document to ensure it contains information needed to adequately answer questions posed in the study while imposing the minimum burden on the responding businesses. The burden on the smallest companies (generally, those with fewer than 10 employees) has been eliminated, as these firms have been excluded from the survey.

The sampling methodology and procedures in this survey are quite similar to those in the previous USITC surveys of remanufacturing activities and used electronic goods. Each study, for example, has populations drawn from Orbis and an industry association list and uses similar methods of survey distribution and data collection. Although the USITC has not specifically tested the methodology and procedures of the trade in digital goods survey, the used electronics and remanufacturing surveys have provided an implicit test of its practicability and utility. In addition, results from these two surveys have helped the digital trade survey team refine several questions in order to increase usability of the survey instrument by the respondent and reliability of the responses.

## **5. Contact information**

Collection and analysis of the data will be the responsibility of the Office of Economics and the Office of Industries within the Commission. Project leader James Stamps can be contacted at 202-205-3227, deputy project leader David Coffin can be contacted at 202-205-2232, and lead economist for the survey portion of this study Ravinder Ubee can be contacted at 202-205-3493. Commission staff also worked with Corey West and his colleagues at Summit Consulting, a survey design and data analysis consulting firm. Mr. West may be contacted at 202-407-8300 or at [corey.west@summitllc.us](mailto:corey.west@summitllc.us).