SUPPORTING STATEMENT
UNITED STATES INTERNATIONAL TRADE COMMISSION QUESTIONNAIRE

Inv. Nos. 332-562 and 332-563

*Global Digital Trade 2: The Business-to-Business Market, Key Foreign Trade Restrictions, and U.S. Competitiveness;* and *Global Digital Trade 3: The Business-to-Consumer Market, Key Foreign Trade Restrictions, and U.S. Competitiveness*

Part B-Collection of Information Employing Statistical Methods

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

### Survey objectives

In a letter dated January 13, 2017, the U.S. Trade Representative (USTR) directed the Commission to produce two reports that analyze the impact of regulatory and policy measures in key foreign markets on (1) the ability of U.S. firms to develop and/or supply digital products and services abroad, and (2) the competitiveness of U.S. firms engaged in the sale of digital products and services, as well as on international trade and investment flows associated with digital products and services. The USTR specified that the Commission's report be based on a review of available data, including a survey of U.S. firms in selected industries particularly involved in digital trade.

### Respondent universe

The respondent universe includes all companies in industries that are particularly involved in the development and supply of digitals products and services. The USITC has identified 52,863 such firms for its sampling frame, of which 13,000 will be sampled. The sampling unit is the firm rather than the establishment.

To examine the effect of policy and regulatory measures, it is necessary to include a broad list of firms that are currently doing business in foreign markets, as well as those that are not, but are potentially interested in doing so. As a result, the response universe will include firms with a potential for digital trade and not only firms that are currently facing these measures in certain markets.

The potential respondent universe represents the sum of firms, net of duplicative records, identified in the Bureau van Dijk's Orbis database in industries likely to engage in digital trade, membership lists from relevant associations, and firms that are registered with the EU-US Privacy Shield. These industries include selected sectors based on NAICS as discussed below.

Because the focus of this study is on firms with international presence and these tend to be larger than the average firm, the response universe was restricted to firms with 20 or more employees or more than $7.5 million in revenue, with some exceptions.

**Sample design**

Survey respondents will be selected through a stratified random sampling methodology that stratifies firms through a combination of industry and size. There will be 33 strata by industry/size. Size cutoffs were determined by the Small Business Association's size standards by NAICS. Firms with neither employment nor revenue data have been combined into their own strata by industry.

Industries comprise the following 10 sectors: (1) Online commercial and retail services, and payment services, (2) telecommunications, (3) web, software, and hardware services, (4) transport and delivery services, (5) media and publishing, (6) finance, (7) insurance, (8) travel and hospitality, (9) other selected services, (10) manufacturing.[1]

Table 1 presents the sampling frame—the population of firms in each stratum. Table 2 presents the sample size for each stratum. Thirteen thousand (13,000) firms will be sampled.

Based on results of past surveys conducted by the Commission for other investigations, we expect the response rate to range from 40–60 percent,[2] which would result in 5,200–7,800 surveys received from the sampled companies (assuming 13,000 surveys sent out). Responses in previous and ongoing USITC surveys have not differed significantly by firm size or across industries. Thus, a uniform response rate has been assumed for all strata.

[1] Although the agriculture and natural resources industries are not explicitly covered, they are selectable as primary industries in the questionnaire for the purposes of complete industry coverage in question 1.3.

[2] The Commission anticipates a higher response rate than in previous surveys based on the use of a web-based questionnaire, which should reduce burden and be easier to respond to. However, since this collection method has not been used by the Commission in the past, a more conservative response rate was used in the calculation of sample size to avoid under sampling.

**TABLE 1** Sampling frame – Population per stratum

| Industry | Small | Medium/ Large | Very large | Unknown size | Total |
|---|---|---|---|---|---|
| Online commercial and retail services, and payment services | 244 | 161 | 44 | 197 | 646 |
| Telecommunications | 935 | 75 | 121 | 557 | 1,688 |
| Web, software, and hardware services | 5,154 | 1,220 | 708 | 5,314 | 12,396 |
| Transport and delivery services | 1,555 | 1,359 | 943 | 3,453 | 7,310 |
| Media and Publishing | 2,099 | 462 | 372 | 1,410 | 4,343 |
| Finance[a] | — | 1,357 | 145 | — | 1,502 |
| Insurance[a] | — | 1,362 | 45 | — | 1,407 |
| Travel and hospitality | 2,069 | 1,607 | 507 | 1,828 | 6,011 |
| Other selected services | 3,104 | 4,754 | 849 | 4,037 | 12,744 |
| Manufacturing[b] | 4,816 | | | | 4,816 |
| Total | | | | | 52,863 |

[a] Within the finance and insurance industries, small firms and those with an unknown size will not be surveyed as it is unlikely that they will be within the scope of the request.
[b] Within the manufacturing industry, only firms that have an indication of having foreign affiliates, and those with 10,000 or more employees, are included in the population.

**TABLE 2** Number of firms in the sample by stratum

| Industry | Small | Medium/ Large | Very large | Unknown size | Total |
|---|---|---|---|---|---|
| Online commercial and retail services, and payment services | 70 | 161 | 44 | 197 | 472 |
| Telecommunications | 100 | 69 | 121 | 366 | 656 |
| Web, software, and hardware services | 2,113 | 1,220 | 708 | 531 | 4,572 |
| Transport and delivery services | 75 | 204 | 802 | 345 | 1,426 |
| Media and Publishing | 100 | 85 | 372 | 456 | 1,013 |
| Finance[a] | — | 355 | 145 | — | 500 |
| Insurance[a] | — | 400 | 45 | — | 445 |
| Travel and hospitality | 139 | 241 | 431 | 366 | 1177 |
| Other selected services | 110 | 713 | 722 | 404 | 1,949 |
| Manufacturing[b] | 790 | | | | 790 |
| Total | | | | | 13,000 |

[a] Within the finance and insurance industries, small firms and those with an unknown size will not be surveyed as it is unlikely that they will be within the scope of the request.
[b] Within the manufacturing industry, only firms that have an indication of having foreign affiliates, and those with 10,000 or more employees, are included in the population.

## 2. Collection of information employing statistical methods

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

A stratified sample is being implemented for this collection. The goal of the stratification scheme is to develop a set of strata that minimizes the variance of responses (such as level of employment and type of activities) within each stratum. Because no pro-forma reliable data exist on the size and scope of the number of firms that are engaged in digital trade, the minimum size of firms included in the population and the classification of size within industries for the stratification scheme were based on information gathered from industry representatives and the best judgment of USITC experts.

Because some firms had no data for both the number of employees and revenue, the approach to sampling could not rely on either of these variables. Thus, the sample size for each stratum were calculated using Cochran's sample size formula for categorical data along with an oversampling correction with an anticipated return rate of 50 percent.[3] The sampling rate by stratum was then adjusted to sample at least 10 percent of each stratum.

The sample size for stratum h is calculated by the formula

$$n_{h0} = \frac{t^2 pq}{d^2}$$

where $t$ is the t-value at $\alpha = 0.05$ for two tails, $pq = 0.25$ as an estimate for variance, and $d = 0.05$ as an acceptable margin of error for categorical data.

The formula for Cochran's oversampling correction is

$$n_{h1} = \frac{n_{h0}}{\left(1 + \frac{n_{h0}}{N_h}\right)}$$

where N is the size of stratum *h*.

### 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. There is an equal probability of selection within each stratum.

- *Sample selection weighting*: Because the sampling rates are based on two criteria, as discussed above, the selection weight factor will account for both the probability of selection within a particular industry and size, and any oversampling of firms.
- *Nonresponse adjustment*: The nonresponse adjustment factor is designed to attenuate bias due to differential response rates. This adjustment will be calculated using firm characteristics, if

[3] J. Bartlett, J. Kotrlik, and C. Higgins. "Organizational research: Determining appropriate sample size in survey research," *Information Technology, Learning, and Performance Journal* 19, no. 1, 2001, 43.

warranted. See the section below on accuracy and reliability of information collected for further discussion.

- *Poststratification weighting*: If necessary, a poststratification weighting factor will be used to attenuate bias due to sample frame noncoverage, overcoverage, or omissions. Population information from Census data, such as the number of firms in each NAICS industry and in each size category (organized by number of employees), may be used to conduct poststratification. Although the best effort has been made to obtain a representative sample, the distribution of firms across industries is not known with certainty in advance.

The general weighting formula can be represented as

$$W_h = S_h \times NR_h \times PS_h \,, \tag{1}$$

where $S_h$ is the sample selection weight for stratum h, $NR_h$ is the nonresponse adjustment factor for stratum h, and $PS_h$ is the poststratification weight of stratum h. $W_h$is the weight applied to all observations in stratum h. This formula may be adjusted to include a firm-specific weighting component if non-response is determined to be related to factors aside from the factors used to design the strata.

Standard estimation procedures will be used as in Heeringa et al (2010).[4] For example, the formula used to estimate the population attribute of interest is found in equation 2. Per standard notation, the total estimate for industry $k$, $\tau_k$, from a stratified random sample, is given by

$$\tau_k = \sum_{h=1}^{L} N_h \overline{y}_h, \tag{2}$$

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

The variance estimate for sampling without replacement is given by

$$var(\tau_k) = \sum_{h=1}^{L} N_h (N_h - n_h¿) \frac{s^2}{n_h} ¿ \tag{3}$$

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*.

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

A sample size of 13,000 is needed to achieve estimates of +/- 5 percent at 90 percent confidence. 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 continuous and categorical variables. For example, table 3 provides the maximum margin of error for a binary question, given alternative response rates.

[4] S. Heeringa, B. West, and P. Berglund, *Applied Survey Data Analysis*, CRC Press, 2010.

**TABLE 3** Margin of error for 90 percent confidence interval

| | Response rates, percent | | |
|---|---|---|---|
| Measure | 40 | 50 | 60 |
| Number of respondents | 5,200 | 6,500 | 7,800 |
| Standard error, percent | 0.69 | 0.62 | 0.57 |
| Margin of error, percent | 1.14 | 1.02 | 0.93 |

Note: This assumes a maximum margin of error of 50 percent for a binary question.

**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. As such, the total recurring annual cost burden is zero.

**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 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 as clear and succinct as possible to gather the specific material requested by the USTR. (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 rates.

**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 providing and selling digital products and services in foreign markets.

Response rates in similarly scoped USITC surveys have recently approached 60 percent. 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. These may include variables such as firm size, industry, NAICS code, or location. Any differences in response rates can be further investigated through logistic regression analysis, using firm characteristics 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. 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.

The Commission expects that all sampled information will yield reliable data that can be generalized to the universe studied.

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

The Commission sought comments on the questionnaire with industry representatives of several relevant industries through field testing. These representatives provided feedback in areas such as availability of data, reporting burden, product coverage and definitions, clarity of instructions, disclosure, and reporting format. The Commission also went through a period of cognitive testing to make sure the questions, and the intent behind them, are clear. See part A for the comments field and cognitive testers made 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. Industry analysts and economists have reviewed the questionnaire to ensure it requests information needed to adequately answer questions posed in the study while imposing a minimum burden on the responding businesses.

The sampling methodology and procedures in this survey are quite similar to those in prior USITC survey work. Prior studies, for example, also have had populations drawn from Orbis; have also stratified by industry and size; and have used similar methods of survey distribution and data collection. Although the USITC has not specifically tested the methodology and procedures of the current Global Digital Trade survey, prior surveys have provided implicit tests of its practicability and utility.

**5. Contact information**

Collection and analysis of the data will be the responsibility of the Office of Analysis and Research Services, the Office of Economics, and the Office of Industries within the Commission. The project leaders are Dan Kim and Alissa Tafti (Inv. No. 332-562) and Ricky Ubee and Christopher Robinson (Inv. No. 332-563). The survey team can be reached by email at globaldigitaltrade@usitc.gov. If you prefer to contact them by phone, please call 202-205-3342 or 202-205-3225.

Industries and their NAICS code composition

Online commercial and retail services, and payment services

- 454110 Business to business electronic markets
- 454111 Electronic shopping
- 454112 Electronic auctions

Telecommunications

- 517110 Wired telecommunications carriers
- 517210 Wireless telecommunications carriers (except satellite)
- 517911 Telecommunications resellers
- 517919 All other telecommunications

Web, software, and hardware services

- 511210 Software publishers
- 518210 Data processing, hosting, and related services
- 519130 Internet publishing and broadcasting and web search portals
- 541511 Custom computer programming services
- 541512 Computer systems design services
- 541513 Computer facilities management services
- 541519 Other computer related services

Transport and delivery services

- 481111 Scheduled passenger air transportation
- 488320 Marine cargo handling
- 488330 Navigational services to shipping
- 488390 Other support activities for water transportation
- 488510 Freight transportation arrangement
- 492110 Couriers and express delivery services

Media and publishing

- 511110 Newspaper publishers
- 511120 Periodical publishers
- 511130 Book publishers
- 511199 All other publishers
- 512110 Motion picture and video production
- 512120 Motion picture and video distribution
- 512191 Teleproduction and other postproduction services
- 512199 Other motion picture and video industries
- 515120 Television broadcasting
- 515210 Cable and other subscription programming
- 519110 News syndicates
- 519190 All other information services

Finance

| | |
|---|---|
| 522110 | Commercial banking |
| 522120 | Savings institutions |
| 522190 | Other depository credit intermediation |
| 522210 | Credit card issuing |
| 522220 | Sales financing |
| 522293 | International trade financing |
| 522294 | Secondary market financing |
| 522298 | All other nondepository credit intermediation |
| 522320 | Financial transactions processing, reserve, and clearinghouse activities |
| 523110 | Investment banking and securities dealing |
| 523120 | Securities brokerage |
| 523130 | Commodity contracts dealing |
| 523210 | Securities and commodity exchanges |
| 523910 | Miscellaneous intermediation |
| 523930 | Investment advice |
| 523991 | Trust, fiduciary, and custody activities |
| 523999 | Miscellaneous financial investment activities |
| 561450 | Credit bureaus |

Insurance

| | |
|---|---|
| 524113 | Direct life insurance carriers |
| 524114 | Direct health and medical insurance carriers |
| 524126 | Direct property and casualty insurance carriers |
| 524128 | Other direct insurance (except life, health, and medical) carriers |
| 524210 | Insurance agencies and brokerages |

Travel and hospitality

| | |
|---|---|
| 532111 | Passenger car rental |
| 561510 | Travel agencies |
| 721110 | Hotels (except casino hotels) and motels |

Other selected services

| | |
|---|---|
| 532230 | Video tape and disc rental |
| 541310 | Architectural services |
| 541330 | Engineering services |
| 541380 | Testing laboratories |
| 541430 | Graphic design services |
| 541611 | Administrative management and general management consulting services |
| 541618 | Other management consulting services |
| 541712 | Research and development in the physical, engineering, and life sciences (except biotechnology) |
| 541810 | Advertising agencies |
| 611420 | Computer training |

Manufacturers with a presence in a foreign market, and those with 10,000+ employees

31-33 Manufacturing