Federal Committee on Statistical Methodology

Transparent Quality Reporting in the Integration of Multiple Data Sources: A Progress Report, 2017-2018

Prepared by the Working Group on Transparent Quality Reporting in the Integration of Multiple Data Sources

October 2018

Transparent Quality Reporting in the Integration of Multiple Data Sources: A Progress Report, 2017-2018

Prepared by the Working Group on Transparent Quality Reporting in the Integration of Multiple Data Sources of the Federal Committee on Statistical Methodology

October 2018

Federal Committee on Statistical Methodology

Members

Nancy Bates
U.S. Census Bureau

Chris Chapman
National Center for Education Statistics

Jennifer Edgar
Bureau of Labor Statistics

John Eltinge (Co-Chair)
U.S. Census Bureau

Dennis Fixler
Bureau of Economic Analysis

Barry Graubard
National Cancer Institute

Jennifer Madans
National Center for Health Statistics

Rochelle (Shelly) Martinez (Co-Chair)
Office of Management and Budget

Wendy Martinez
Bureau of Labor Statistics

Jaki McCarthy
National Agriculture Statistic Services

Peter Miller
U.S. Census Bureau

Jeri Mulrow
Bureau of Justice Statistics

Jennifer Nielsen (Secretary)
National Center for Education Statistics

Jennifer Parker
National Center for Health Statistics

Polly Phipps
Bureau of Labor Statistics

Nancy Potok
Office of Management and Budget

Mark Prell
Economic Research Service

Joseph Schafer
U.S. Census Bureau

Rolf Schmitt
Bureau of Transportation Statistics

Marilyn Seastrom
National Center for Education Statistics

Joy Sharp
Bureau of Transportation Statistics

G. David Williamson
Agency for Toxic Substances and Disease Registry

Linda Young
National Agriculture Statistic Services

Working Group on Transparent Quality Reporting in the Integration of Multiple Data Sources

Members

Chair
John Eltinge, U.S. Census Bureau

Bureau of Justice Statistics
John Popham

Bureau of Labor Statistics
Wendy Martinez
Nicole Nestoriak
Polly Phipps

Bureau of Transportation Statistics
Rolf Schmitt

Census Bureau
Paul Marck
Joseph Schafer

Economic Research Service
Mark Prell

National Agricultural Statistics Service
Linda Young

National Center for Education Statistics
Chris Chapman
Richard Reeves

National Center for Health Statistics
Irma Arispe
Renee Gindi
Lisa Mirel
Carolyn Neal
Jennifer Parker

Statistics of Income Division
Brian Balkovic
Tamara Rib

Veterans Health Administration
Alon Ben-Ari
Ned Confer
Katherine Hoggatt
David Maron
Brian Sauer
Elani Streja

Table of Contents

List of Acronyms

BEA	Bureau of Economic Analysis
BJS	Bureau of Justice Statistics
BLS	Bureau of Labor Statistics
BTS	Bureau of Transportation Statistics
CNSTAT	Committee on National Statistics
CPS	Central Processing System
ERS	Economic Research Service
ESS	European Statistical System
FBI	Federal Bureau of Investigation
FCSM	Federal Committee on Statistical Methodology
FSS	Federal Statistical System
ICSP	Interagency Council on Statistical Policy
IMF	International Monetary Fund
IPEDS	Integrated Postsecondary Education Data System
JPSM	Joint Program in Survey Methodology
LEHD	Longitudinal Employer-Household Dynamics
MPR	Mathematica Policy Research
NASEM	National Academies of Science, Engineering, and Medicine
NASS	National Agricultural Statistics Service
NCES	National Center for Education Statistics
NCHS	National Center for Health Statistics
NCS-X	National Crime Statistics Exchange

NCVAS	National Center for Veterans Analysis and Statistics
NIBRS	National Incident Based Reporting System
NPSAS	National Postsecondary Student Aid Study
NSLDS	National Student Loan Data System
NSLTCP	National Study of Long-Term Care Providers
OECD	Organization for Economic Cooperation and Development
OMB	Office of Management and Budget
QCEW	Quarterly Census of Employment and Wages
RTI	Research Triangle Institute
SNAP	Supplemental Nutrition Assistance Program
SOI	Statistics of Income
Stats NZ	Statistics New Zealand
TSE	Total Survey Error
UK	United Kingdom
UMD	University of Maryland
UN	United Nations
USDA	United States Department of Agriculture
VA	U.S. Department of Veterans Affairs
VBA	Veterans Benefits Administration
VHA	Veterans Health Administration
WSS	Washington Statistical Society

Executive summary

U.S. federal statistical programs have integrated datasets together to create new or improved statistical products for many years. Recently, leadership of the federal statistical system asked the Federal Committee on Statistical Methodology (FCSM) to start work to develop quality standards for integrated data used in statistical programs. In support of such standards, the FCSM is identifying best practices for reporting on quality dimensions of integrated data products created and produced in federal statistical programs. This report summarizes the activities of the FCSM Working Group in its first year.

Building upon the well-established framework provided by Total Survey Error (TSE, see Groves et al., 2009), the FCSM has conducted a series of activities to explore quality reporting of integrated data. These activities included workshops to examine quality measurements for data used as an input to an integrated data product, quality assessment used in various integrated data processing methods, and quality measurement and reporting for output data products. Speakers at these workshops reported on a wide variety of topics, including features of structured and unstructured data, impacts of changes in data ownership and control with external data, modeling, linkage and fusion methods for integrating data, TSE and quality frameworks, and the need for transparent reporting targeted to different stakeholders and data users. The FCSM has identified several relevant research topics to improve methods to integrate data, appropriate measures to assess data quality for integrated data, and the communication of data-quality results to multiple stakeholder groups. The FCSM is holding additional workshops in specific areas and developing a detailed research agenda to identify methodological and theoretical work needed to inform documentation strategies for integrated data. Although the generic term 'survey data' is used throughout, differences between household and establishment surveys are recognized in the FCSM Working Group's activities and the research agenda described in this report.

From these initial activities, the FCSM is evaluating quality reporting frameworks that can be adapted for use by the federal statistical programs. The framework used by the European Statistical System (ESS; ESS, 2015) combined with principles from TSE and a recent report from the National Academies of Science, Engineering and Mathematics (NASEM, 2017) were starting points for the evaluation. The FCSM also reviewed other international standards and frameworks used for quality reporting, with attention to those applicable to integrated data. Many existing frameworks share many elements in common, including defining data quality as "fitness for use" in which "good" or "high" quality data meets its intended purpose in operations, decision-making, and planning. The FCSM has

begun to assess practices used for current integrated data products in the context of points raised during the workshops and existing data quality frameworks.

The FCSM and the Interagency Council on Statistical Policy (ICSP) are working to create a unified framework that benefits from the work already completed by ICSP (2018), ESS (2015), and NASEM (2017). Quality dimensions under consideration for this framework include *relevance, accuracy, reliability, timeliness, accessibility, clarity, comparability, coherence, transparency* and *interpretability, granularity, privacy, objectivity, confidentiality, integrity,* and *precision*. This joint effort, along with the FCSM's ongoing research activities and the results thereof, will contribute to development of a framework for transparent reporting of data quality for integrated data products, including appropriate statistical standards. Whether federal statistics come from traditional household and establishment surveys and carefully designed administrative records systems or from sources not initially designed for statistical purposes, data quality must be communicated transparently and understood to provide the best available statistical information to the public and to ensure the statistics are used wisely.

Introduction

The Federal Committee on Statistical Methodology (FCSM) established a Working Group on Transparent Quality Reporting in the Integration of Multiple Data Sources (hereafter FCSM Working Group) in 2017 to identify best practices for reporting on quality dimensions of integrated data products created and produced in federal statistical programs. This report summarizes the activities of the FCSM Working Group in its first year.

Integrated data merge or combine in some fashion two or more data sources, such as survey data from a statistical agency and administrative data from a government program. Integrated data are also known as multiple-source, blended, hybrid and mixed data. Statistical agencies have used integrated data for many years to produce a diverse array of products, including statistical estimates, public-use and restricted-use micro-level datasets, and analytical reports. Survey data can include both household and establishment surveys. Non-survey data used in these products can include administrative records, sensor data, imagery, crowd-sourced data, web scrapped data, and outputs of models. The scale and scope of agencies' uses of integrated data have grown in recent years. It is expected that this trend will continue and perhaps even accelerate.

FCSM established the Working Group in response to the request by the Interagency Council on Statistical Policy (ICSP) for detailed analysis on the statistical quality of integrated data in order to inform eventual statistical standards or guidelines. The ICSP, chaired by the Chief Statistician of the U.S. and comprised of the heads of the principal federal statistics agencies and representatives from other statistical programs, considers integrated data a key strategy to modernize the Federal Statistical System (FSS). Integrated data sources can be used to develop statistical reports and estimates that help inform policymakers and the American public. However, it is important for users of statistical products to be aware of the strengths and limitations of data in order to appropriately use the data and interpret results; poor quality statistics can misguide decisions with costly consequences. All data have potential errors and errors can be compounded when data from different sources are integrated to produce statistics. The initial focus of the FCSM Working Group has been the need for transparent quality reporting to inform users, however, quality dimensions are also used by statistical agencies to inform decisions and identify resource needs.

While the FCSM Working Group's activities began in 2017, their work is consistent with the Federal Data Strategy, part of the 2018 President's Management Agenda, for leveraging data as a strategic asset (strategy.data.gov, accessed on 9/30/18). The draft strategy principles include to:

- "… comprehensively document processes and products to inform data users."
- "… harness data from multiple sources, and acquire new data only when necessary.

The ICSP statement on *Principles for Modernizing Production of Federal Statistics* (ICSP, 2018) recognizes the shift from reliance on surveys and the challenges of new and blended (i.e. integrated) data sources, indicates that ICSP and FCSM will work together on a data quality framework for integrated data, and features transparent reporting on data of information quality as one of its three principles for using integrated and non-statistical data for statistical purposes. The FCSM Working Group's activities provide information to support both efforts.

Introducing reporting standards for integrated data would represent an application of 2002 Office of Management and Budget (OMB) Information Quality Act guidance (Federal Register, 2002). Following the direction of Congress in 2001, OMB issued government-wide guidelines in 2002 that "provide policy and procedural guidance to Federal agencies for ensuring and maximizing the quality, objectivity, utility, and integrity of information (including statistical information) disseminated by Federal agencies." The OMB guidance for ensuring and maximizing the quality, objectivity, utility, and integration of information is not limited to survey data and survey-based estimates. It applies to *all* information disseminated by Federal agencies. The work of the FCSM Working Group may someday contribute to detailed standards issued by OMB for statistical agencies to report on the quality of integrated data. Such OMB reporting standards would represent an application of 2002 OMB guidance.

The efforts of the FCSM Working Group during its first year were informed by the National Academies of Science, Engineering, and Medicine (NASEM) report titled *Federal Statistics, Multiple Data Sources, and Privacy Protection: Next Steps* (available at http://nap.edu/24893) (NASEM, 2017). This NASEM report discusses improving federal statistics for policy and social science research using multiple data sources and state-of-the-art estimation methods. Chapter 6 of the NASEM report, in particular, explores the issues of moving beyond existing quality frameworks, including Total Survey Error (TSE), to fully address the quality and confidentiality and access considerations inherent with integrating multiple data sources. The TSE framework is a paradigm for looking at errors (i.e. differences between the survey response observed and the true value the survey was measuring) stemming from the design, collection, and processing of survey data (Biemer et al 2017, Groves et al. 2009). The NASEM report calls for a broader framework of assessing the quality of administrative and private sector data, such as organic and found data that could be used alone or in combination to produce

federal statistics that go beyond TSE. Specifically, the NASEM report calls for a more comprehensive quality framework to include dimensions that capture user needs, including:

- *timeliness,*
- *relevance,*
- *accuracy,*
- *accessibility,*
- *coherence,*
- *integrity,*
- *privacy,*
- *transparency* and *interpretability;*
- *granularity.*

Since the Fall of 2017, the FCSM Working Group has been actively gathering information. Mathematica Policy Research (MPR) and the Joint Program in Survey Methodology (JPSM) of the University of Maryland (UMD) have been partners in these activities. The FCSM Working Group, under the auspices of the Washington Statistical Society (WSS), co-sponsored three public workshops to gather information on the quality of input data, data processing, and statistical outputs. In sessions at the 2018 FCSM Research and Policy Conference in March 2018, FCSM Working Group members reported on the three FCSM/WSS workshops and discussed related research priorities. In addition to the information from the three workshops and the FCSM Conference, the Statistics of Income (SOI) Division of the Internal Revenue Service sponsored a report by MPR that examined data quality frameworks and standards used outside the U.S. by national statistical offices and international organizations.

This report, summarizing the FCSM Working Group activities, includes points raised by presenters and external collaborators in these activities and identifies some future directions. Building from the initial workshops and activities, the FCSM Working Group is assessing quality reporting for integrated data products in statistical agencies and is identifying the research needed to inform the development of a framework for documenting data quality. Appendices in this report include the list of presentations from the FCSM/WSS workshops, selected examples of integrated data activities currently underway at statistical agencies, examples of current data quality assessments used at statistical agencies, a list of elements that can be organized by quality dimensions, along with indicators, for use in

an instrument for assessing quality for integrated data, and a list of possible research areas under consideration by the FCSM Working Group.

In collaboration with FSS staff and colleagues outside the FSS, the FCSM Working Group has already produced several products that are described in this report. Future products are expected. Products that are authored by members of the FCSM Working Group are being disseminated through the FCSM publications; products authored by others will be made available to the public through posting on the FCSM website, currently hosted by the National Center for Education Statistics (NCES).

FCSM Working Group Activities

As noted above, to better understand the range of characteristics related to integrated data that could be considered quality dimensions for the purpose of documentation and other data reporting, the FCSM and WSS co-hosted three workshops that explored current practices with respect to reporting on data quality. Quality was divided into three parts—input data, data processing, and statistical outputs — as the quality of integrated data is the sum of the quality of each of these components. A list of the presentations is in Appendix I, with links to the presentations found at http://washstat.org/presentations/. The Economic Research Service (ERS) of the USDA sponsored a JPSM report, *Findings from the Integrated Data Workshops Hosted by the Federal Committee on Statistical Methodology and Washington Statistical Society* (Brown et al, 2018). The final version was completed in August 2018 and can be found at http://nces.ed.gov/FCSM/pdf/Workshop_Summary.pdf.

Workshop 1, Input Data Quality

The goal of the first workshop was to identify features of data used as inputs to integrated data relevant for transparent reporting. In addition, the workshop opened by describing the goals for the all three workshops and the overall efforts of the FCSM Working Group on Transparent Reporting of Integrated Data.

A presentation was made on the findings from Chapter 6 of the NASEM report, *Quality Frameworks for Statistics Using Multiple Data Sources* (NASEM, 2017). In addition to presenting recommendations from the report, several points were emphasized, including: the importance of focus on coverage error with increasing use of non-survey data; the need to understand the data generation process for newer data sources to assess the construct-measurement gap; measurement of linkage errors; and, the importance of effective communication with the general public to ensure trust in resulting data and statistical estimates.

The remaining presentations and discussions of the first workshop were organized around identifying data quality standards and issues for external data considered as input for integrated data products, including administrative data, less structured data and unstructured data. (Examples of structured data include administrative data and private-sector data on prices. Unstructured data include satellite imagery, traffic patterns, sensor data, web scraped data, and medical images.) Some important points raised for the FCSM Working Group, included:

- Administrative data sources share many parallels with traditional survey data with respect to the assessment of data quality. There are fewer parallels for less structured and unstructured data sources.
- Clear and appropriate communication from data producers about the original reason for the collection of external data that will be used for statistical purposes is essential.
- Incorporating external data sources reduces the amount of control that the statistical agency has over data delivery, data quality and transparency. Agencies must trust that data providers will meet production expectations but agencies will also need to develop contingency plans in the event that data from a certain source cease to be available, cannot be delivered, or are markedly changed. Control of external data is particularly important for time trends, key products of statistical agencies.

Workshop 2, Data Processing

The focus of the second workshop was data processing for integrated data products, which include record linkage, statistical matching/data fusion, harmonization across data sources, and disclosure avoidance. As in the first workshop, an overarching theme of the second workshop was the importance of transparent and clear communication with technical and non-technical data users, from users of micro-data files to users of estimates conveyed in reports or tabulations. Some points raised for the FCSM Working Group at the second workshop include:

- High quality "truth decks" (e.g., hand-matched subsamples) are important for determining the quality of a linked data product.
- Transparency about modeling assumptions and other data blending techniques, and the biases these might create in resulting integrated estimates and data products, is critical.
- Bayesian multivariate models and other statistical modeling approaches may provide promising theoretical frameworks for combining datasets in ways other than record linkage (e.g. statistical matching.) However, these models do not yet incorporate understandings of quality profiles for the different input data sources and do not have standard measures for outputs.
- Statistical agencies should plan for data harmonization when designing their surveys and other primary collections of statistical data for future integration of multiple data sources.

- It is possible to be transparent both about the methods used for disclosure avoidance and about the error properties of the resulting estimates after application of disclosure avoidance measures.
- Some indicators of quality for one level of granularity do not necessarily apply to other levels.
- Varying quality profiles of input data are often the impetus for creating integrated data products.

Workshop 3, Statistical Outputs

The third workshop focused on the quality of statistical outputs. This last workshop also provided a synthesis from all three FCSM/WSS workshops. Building on the importance of communication, workshop organizers discussed different transparency reporting levels that could be used to determine the amount of documentation provided alongside a final data product. Data products may have users of varying sophistication and the level of detail provided for one type of user may not be appropriate or necessary for another user. Three general categories of transparency were identified:

- High Transparency for academics, agency specialists, subject-matter experts
- Moderate Transparency for policy makers, professional journalists, students
- Low Transparency for the general public

Several frameworks for reporting on the quality of integrated data were presented during the workshop. A way to expand the TSE model, developed originally for the evaluation of survey data, to incorporate datasets and hybrid estimates based on integrated data was presented as was a review of quality assessment models developed elsewhere in the world including contrasting these models with the quality profiles that at one point were thought by some in the U.S. statistical system to be the way forward.

The workshops concluded with a call to rethink existing research paradigms in the world of integrated data. There is a need to assess quality at the estimate level but this presents challenges given that each individual estimate may require its own tailored quality assessment. It was also suggested that interdisciplinary research collaborations should be promoted, noting that no single discipline is likely to possess all of the expertise required to develop and work with integrated data products.

Additional points raised during the third workshop, including some ideas from the synthesis of the three workshops, include:

- Transparency is needed at each step of data product development, from the original motivation for collecting the input data, to the steps that were taken to harmonize multiple data sets, to the matching procedures and model assumptions chosen, all the way through to the evaluation techniques the producers used. Data processing decisions for every step of the collection and analysis process should be documented so that stakeholders can judge whether a given product is appropriate for their use.
- Levels of transparency need to provide tiered access for data users, allowing users to find the information they needed in the format that works the best for them.
- Fitness for use is apt to be very project specific. Quality dimensions will not have equal value to all users for all products.
- A unified research plan is needed along with more communication between the FSS and stakeholders as to the major priorities for facilitating the use of integrated data products.
- Interdisciplinary teams can be of particular value when working with integrated data products and measuring their quality.
- Quality reporting for estimates will differ from that for datasets.
- Existing quality frameworks for survey and registry data reflect the statistical agency's control over the collection and dissemination of its data. Integrated data, particularly the use of non-structured and 'found' data sources, shift the control of many quality dimensions from the front end (e.g. data collection, initial processing) to the back end (processing and harmonization)

2018 FCSM Research and Policy Conference

At the 2018 FCSM Research and Policy Conference held March 7-9, two sessions were dedicated to the three FCSM/WSS workshops. The first session provided overviews of the workshop presentations, with a summary of the main discussion points. The aim of the second session was to identify research priorities for enhancing the quality of integrated data, measuring the quality of integrated data, and communicating important information about the quality of integrated data to potential data users. In the second session, organized as a panel discussion, speakers and the audience offered perspectives on reporting standards and research priorities. More information about the 2018 FCSM Research and Policy Conference and links to presentations are at

http://www.copafs.org/seminars/fcsm2018.aspx. Summaries of the two FCSM Working Group sessions are described in Brown et al (2018). These sessions generated additional ideas for the FCSM Working Group to consider, both from Working Group members and members of the audience, including those listed below.

- For administrative data, relevant quality metrics could differ depending on the use of the data, such as (a) the substitution of respondent data by administrative data for a subset of variables in a sample survey or (b) the combination of administrative data files to produce estimates for a population that neither dataset covers fully.
- Due to the processing required to create an integrated dataset, the dependence on models and sensitivity to modeling assumptions tends to be greater for integrated data analyses than for analyses based on classical survey data.
    - There is an increasing need to develop survey methodologists who are skilled in the science and art of modeling.
    - Standards or best practices need to be developed for reporting on the results of sensitivity analyses. The uncertainty in an estimate and its sensitivity to model specification ought to be related, and both should be communicated to data users.
    - Especially with modeling, more knowledge about the data is often needed than the survey methodologists or statisticians necessarily have.
- A research agenda that addresses the issues most relevant to the FSS as a whole will help the entire system to move forward.
- Incorrect estimates published as a result of rushing into use of integrated data could lead to difficult questions from the public.
- Fresh approaches to handling the tradeoffs between disclosure and users' desire for granularity are needed.

International Guidelines and Standards

The Statistics of Income (SOI) Division of the Internal Revenue Service commissioned MPR to review quality standards from national statistical agencies outside the U.S. as well as international organizations like the United Nations (UN), with an eye on quality reporting for statistical estimates that combine survey data with other types of data. The MPR report *Transparency in the Reporting of Quality for Integrated Data: A Review of International Guidelines and Standards* (Czajka and Stange, 2018) was completed in April 2018 (available at https://www.mathematica-mpr.com/our-publications-and-findings/publications/transparency-in-the-reporting-of-quality-for-integrated-data-a-review-of-international-standards).

MPR reviewed quality standards from Eurostat and the European Statistical System (ESS), leaders in the development of reporting standards for statistical data generally, as well as standards and documents from select European countries, Canada, Australia, the Organization for Economic Cooperation and Development (OECD), and the International Monetary Fund (IMF). As in the workshops, the literature across these countries and organizations is nearly uniform in defining data quality as "fitness for use" in which "good" or "high" quality data meets its intended purpose in operations, decision-making, and planning.

The chapters in the MPR report included detailed summaries of international standards used in the European Union and in selected national statistics organizations; literature on extending the TSE framework to integrated data, with a focus on the two-phase framework by Zhang (2012) and its use by Statistics New Zealand (Stats NZ); a review of quality assessments applied by selected international organizations when using administrative data for official statistics; and a review of the use of "Big Data" for official statistics in selected organizations outside the U.S. Based on their evaluations, the MPR report concluded with several observations for the FCSM Working Group, including these briefly summarized below:

- The concept of quality as expressed in a wide array of quality frameworks for statistical data is characterized by several features:
    - Quality is commonly defined as fitness for use.
    - Quality is multi-dimensional; five dimensions appear almost universally in quality frameworks around the world: (1) *relevance*, (2) *accuracy* and *reliability*, (3) *timeliness* and *punctuality*, (4) *coherence* and *comparability*, and (5) *accessibility* and *clarity*.

  - There are trade-offs among the dimensions.
  - While the multi-dimensional formulation of quality suggests comparable importance among the dimensions, discussions of quality in the international literature give disproportionate attention to *accuracy*.
  - A number of other dimensions appear in some national statistical organizations' quality frameworks; examples include interpretability*, credibility, methodological soundness, serviceability, assurances of integrity,* and *confidentiality*. *Granularity*, a dimension of quality singled out in NASEM (2017), is cited only rarely in international quality frameworks and supporting documents.

- Only one national statistical organization—Stats NZ—has developed a quality framework explicitly designed to address integrated data, building on Zhang's (2012) adaptation of the TSE model (Stats NZ, 2016).
- The distinction between the original purpose of an administrative data source and its use as one of multiple sources in an integrated dataset for statistical purposes is discussed repeatedly, albeit in different ways.
- Modeling is addressed extensively in the ESS Handbook's recommendations on quality reporting.
- The impact of methods of statistical disclosure control on the quality of statistical estimates is addressed in the *U.K. Guidelines for Measuring Statistical Output Quality* (Office of National Statistics, 2013). This topic received little attention elsewhere but can be expected to grow in importance.
- The MPR report raised the questions of whether quality could be measured usefully at the dataset level—as in Zhang's (2012) two-phase framework—or restricted to the individual estimate—as in the TSE model of Groves et al. (2009).
- As in the U.S., quality assurance frameworks from other countries are designed primarily for use with survey data, with more limited attention to administrative data—and generally in the form of registers. Efforts to develop quality assurance frameworks and data quality standards for Big Data are recent and in the early stages of development.

Moving forward

Framework for transparent data quality reporting

To make the most of previous work on data quality, the FCSM Working Group is exploring a set of quality attributes to be included in quality framework drawn from the concepts and dimensions already familiar to statistical agencies, combining the TSE model with dimensions used by the ESS: *relevance, accuracy, reliability, timeliness, accessibility, clarity, comparability* and *coherence*. As illustrated in the FCSM/WSS Workshops, constructs in the TSE have counterparts in non-statistical and administrative data. The TSE framework is a way of inventorying and bringing together the sources of error from individual building blocks of an integrated data set, though additional research is needed to understand how the individual errors cancel or compound one another in the final estimate.

The FCSM Working Group found the ESS framework, which includes relatively substantial documentation, to be a useful reference for developing a data quality framework for U.S. statistical agencies. The ESS framework was discussed at length by the CNSTAT panel (NASEM, 2017); some of the concepts and terminology associated with the ESS data quality framework are already familiar to U.S. statistical agencies; and ESS provides many specific definitions of terms in its supporting documentation. However, reporting on some components of the ESS framework could be overly burdensome for many agencies. Further assessment of the applicability of the ESS framework for transparent quality reporting for integrated data in the U.S. will require balanced consideration of multiple factors, including stakeholder information needs, empirical results on dominant quality features for integrated data, and reporting burden.

While some of the quality dimensions included in the ICSP statement "*Principles for Modernizing Production of Federal Statistics*" (ICSP, 2018) are dimensions in the ESS framework, the ICSP's list also includes the following: *objectivity, confidentiality, integrity* and *precision*. Generally, a data quality framework would include how agencies manage the quality of information as well as elements of transparent reporting. The primary focus of (ICSP, 2018) is to set forth an overarching plan/strategy/set of principles to modernize the FSS, whereas most of the ESS guidance and metrics are about process.

The FCSM Working Group is working with ICSP to combine the quality dimensions in ICSP (2018) and the dimensions being considered from the ESS into a unified framework. To do this, a

common understanding of the dimensions under consideration is needed. While the current working definitions are similar, they are not identical. The following definitions are currently being used by ESS (ESS Quality Glossary (ESS, 2012)). These definitions have been the starting points for the FCSM Working Group's discussions of quality reporting and are consistent with the dimensions cited in ICSP (2018).

- *Relevance* is an attribute of statistics measuring the degree to which statistics meet current and potential needs of the users.
- *Accuracy* is an attribute of statistics measuring the closeness of estimates to the unknown true values.
- *Reliability* is an attribute of statistics that measure as faithfully, accurately and consistently as possible the reality that they are designed to represent and implying that scientific criteria are used for the selection of sources, methods and procedures.
- *Timeliness* is an attribute of statistics measuring the period between the availability of the information and the event or phenomenon it describes.
- *Punctuality* is an attribute of statistics measuring the delay between the date of the release of the data and the target date (the date by which the data should have been delivered or released).
- *Accessibility* is an attribute of statistics describing the set of conditions and modalities by which users can obtain data.
- *Clarity* is an attribute of statistics describing the extent to which easily comprehensible metadata are available, where these metadata are necessary to give a full understanding of statistical data.
- *Comparability* is an attribute of statistics measuring the extent to which differences between statistics can be attributed to differences between the true values of the statistical characteristics.
- *Coherence* is an attribute of statistics measuring the adequacy of the data to be reliably combined in different ways and for various uses.

In addition to dimensions in ICSP (2018) and ESS (2015), a statistical quality framework could include elements such as *granularity,* which was recommended by NASEM (2017, Chapter 6). The

growing use of integrated data is motivated, in part, as a response to user need for greater geographic, temporal, and demographic subgroup detail. While indicators of *granularity*, when appropriate, could be incorporated within other quality dimensions, such as *accuracy, precision*, and *relevance*, keeping *granularity* as a separate dimension increases its prominence. On the other hand, *granularity* as a quality dimension may not apply to all estimates; as noted by MPR, *granularity* appears infrequently among the international standards they reviewed. FCSM will continue to evaluate *granularity* as a dimension of quality.

Aspects of quality dimensions for integrated data may overlap or have alternate interpretations from other disciplines. For example, the dimension *accuracy* includes both bias (coverage error, nonresponse error, measurement error and processing error) and variability (sampling error), while indicators of *precision* are also related to variability. Furthermore, the dimension *precision* can have constructs in other disciplines that differ from its statistical use here (e.g. numerical precision is the number of digits in a number; precision can also indicate, resolution (or granularity), the smallest interval measurable by the scientific instrument (i.e. our data collections)).

Possible elements that could be used in a data quality instrument, including dimensions and indicators, are given in Appendix IV.

Current practices for transparent quality reporting

In the summer of 2018, members of the FCSM Working Group began an ongoing project designed to identify best practices for transparent quality reporting for integrated data. Several ICSP agencies are volunteers in the project, which evaluates current reports issued by the agencies for data products created by the integration of multiple data sources. Specifically, the project is evaluating quality dimensions in these reports that describe the integrated data products and surveying users of the agency's statistical product(s) to solicit their appraisals of the agency's reporting practices. The domains that are of interest are the national income and product accounts (which provide Gross Domestic Product and other statistics), integration of administrative and survey data, private-sector data (which are becoming integrated into an array of federal statistical products), and integration of data from multiple surveys.

Pursuing longer term research

The project on identifying best practices for transparent quality reporting for integrated data is one step in methodological and empirical research needed to fully understand documentation strategies for the quality of integrated data. FCSM Working Group is holding additional workshops in 2018 on current research and practice for specific topics identified as critical to quality reporting for integrated data; the FCSM Metadata workshop was held in September 2018 and a workshop on sensitivity analysis is planned for later in 2018. In addition, the theme for the FCSM Geospatial Interest Group's October workshop is "Data Quality Standards When Integrating Geospatial Data Sets". Appendix V identifies some specific high-priority topics under consideration. For most or all of these topics, the FCSM Working Group anticipates that the research will involve a combination of work in mathematical statistics (e.g., on the statistical properties of the proposed methods); behavioral sciences (e.g., on the underlying social processes that lead to specific issues with population coverage, incomplete data or measurement error); computer science (e.g., on the tools used to implement the proposed methods); data science (e.g. uncertainty quantification) and substantive areas like economics, public health and agriculture. Most or all of the topics included in Appendix V could potentially lead to a mixture of research projects that are best carried out by (a) federal civil service personnel, (b) contracting/cooperative agreements, (c) academic peer-reviewed grants or (d) crowdsourcing (e.g., in development of some open-source software).

The FCSM Working Group anticipates that all of the topics listed in Appendix V are potentially important for long-term improvement of agency work with integrated data sources, and related transparent communication with stakeholders, though the most appropriate topics will likely vary across agencies. Indeed, much of the most productive research areas may flow from case-specific work with specific statistical programs or products. In addition, some elements of all of these topics are likely to fit well with academic research funded through the NSF or other peer-reviewed granting organizations.

Although the focus of this working group is on transparent reporting of data quality for integrated data, during the workshops it became evident that research efforts need to continue on the best methods to integrate data, the appropriate measures of data quality, and the most effective means of communicating data quality. Beginning in the fall of 2018, the list of topics (Appendix V) will be evaluated and more fully developed.

Conclusion

Activities of the FCSM Working Group in 2017-2018 provide a solid foundation for fulfilling the FCSM's charge to develop standards for integrated data reporting. Whether federal statistics come from traditional household and establishment surveys and carefully designed administrative record systems or from sources such as transactions, sensors, imagery, and web content not designed for statistical purposes, data quality must be communicated transparently and understood to ensure that the nation is provided the best available statistical information and that the statistics can be used wisely.

References

Biemer, P., de Leeuw, E., Eckman, S., Edwards, B., Kreuter, F., Lyberg, L., Tucker C., and West, B. 2017. *Total Survey Error in Practice*. Hoboken, New Jersey: John Wiley & Sons.

Brown, A., Caporaso, A., Abraham, K., and Kreuter, F. 2018. *Findings from the Integrated Data Workshops hosted by the Federal Committee on Statistical Methodology and Washington Statistical Society.* Found at https://www.census.gov/fedcasic/fc2018/ppt/4CBrown.pdf.

Czajka, J. and Stange, M. 2018. *Transparency in the Reporting of Quality for Integrated Data: A Review of International Guidelines and Standards*. Washington, DC: Mathematica. Found at https://www.mathematica-mpr.com/our-publications-and-findings/publications/transparency-in-the-reporting-of-quality-for-integrated-data-a-review-of-international-standards.

European Statistical System. 2012. *ESS Quality Glossary*. Found at http://ec.europa.eu/eurostat/web/quality/quality-reporting.

European Statistical System. 2015. *ESS Handbook for quality reports, 2014 edition*. Luxembourg: Eurostat, 2015. Found at http://ec.europa.eu/eurostat/web/quality/quality-reporting.

Federal Register Vol. 67, No. 36. 2002. Office of Management and Budget. *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies*. Found at https://www.gpo.gov/fdsys/pkg/FR-2002-02-22/pdf/R2-59.pdf.

Groves, R.M., Fowler Jr., F. J., Couper, M., Lepkowski, J. S., Singer, E., and Tourrangeau, R. 2009. *Survey Methodology*. *Revised edition.* New York: Wiley.

Interagency Council on Statistical Policy. 2018. *Principles for Modernizing Production of Federal Statistics.* Found at http://nces.ed.gov/FCSM/pdf/Principles.pdf

Iwig, W., Berning, M., Marck, P., and Prell, M. 2013. *Data Quality Assessment Tool for Administrative Data*, Federal Committee on Statistical Methodology Working Paper 46. February 2013. Found at https://nces.ed.gov/FCSM/policy_wp.asp.

McEntarfer, E. 2013. "Understanding the U.S. Labor Market: The Longitudinal Employer-Household Dynamics (LEHD) Program" *U.S. Census Bureau, Center for Economic Studies and Research Data Centers Research Report: 2012*, U.S. Government Printing Office, Washington, DC, 2013. Found at https://www.census.gov/ces/pdf/2012_Research_Report.pdf.

National Academies of Sciences, Engineering, and Medicine. 2017. *Federal Statistics, Multiple Data Sources, and Privacy Protection: Next Steps*. Washington, DC: The National Academies Press. Found at https://doi.org/10.17226/24893.

National Center for Health Statistics, Office of Analysis and Epidemiology. 2018. *The Linkage of the National Center for Health Statistics Survey Data to the National Death Index – 2015 Linked Mortality File (LMF): Methodology Overview and Analytic Considerations*. Hyattsville, Maryland. Found at https://www.cdc.gov/nchs/data-linkage/mortality-methods.htm.

National Center for Veterans Analysis and Statistics. 2018. *Profile of Veterans: 2016*. Found at https://www.va.gov/vetdata/About_Us.asp.

Office for National Statistics. 2013. *Guidelines for Measuring Statistical Output Quality.* Version 4.1. United Kingdom. Found at https://www.statisticsauthority.gov.uk/wp-content/uploads/2017/01/Guidelines-for-Measuring-Statistical-Outputs-Quality.pdf.

Statistics New Zealand. 2016. *Guide to Reporting on Administrative Data Quality.* Wellington, NZ: Statistics New Zealand. Found at http://archive.stats.govt.nz/methods/data-integration/guide-to-reporting-on-admin-data-quality.aspx.

Zhang, L.C. 2012. "Topics of statistical theory for register-based statistics and data integration." *Statistica Neerlandica*, vol. 66, no. 1, pp. 41-63.

Appendix I. Federal Committee on Statistical Methodology/Washington Statistical Society (FCSM/WSS) Workshops on the Quality of Integrated Data

First Workshop: Input Data (December 1, 2017)

Eltinge, John. 2017. "*Introduction to FCSM/WSS Workshops on the Quality of Integrated Data*"

Chapman, Chris. 2017. "*Overview of the Input Workshop*"

Groves, Robert. 2017. "*Advancing the Paradigm of Combining Data Sources – Considerations from the National Academy of Sciences*"

Cohen, Steven. 2017. "*The Integration of Administrative and Survey Data in Support of Medical Expenditure Analyses: Utility and Challenges*"

Berning, Michael and David Sheppard. 2017. "*Quality of Administrative Records as Source Data"*

Murphy, Bonnie and Crystal Konny. 2017. "*Quality Considerations for Administrative Data Used for the Producer Price Index (PPI) and Consumer Price Index (CPI) development*"

Muth, Mary. 2017. "*Assessment of Commercial Store and Household Scanner Data: Methods, Content, and Cautions*"

Elkin, Peter. 2017. "*The Improvement in Sensitivity and Often Specificity when Adding Unstructured to Structured Data*"

Johnson, David. 2017. "*Satellite Remote Sensing Imagery Quality and Timeliness: Considerations for Use in Regional Estimation of Crop Production*"

Mahapatra, Subrat. 2017. "*Sensing Data Quality in Sensor-Based Data*"

Rigobon, Roberto. 2017. "*Web-scraped Data, Consideration of Quality Issues for Federal Statistics*"

Second Workshop: Data Processing (January 25, 2018)

Eltinge, John. 2018. "*Introduction*"

Chapman, Chris. 2018. "*Recap of First Workshop: Lessons Learned*"

Shafer, Joseph. 2018. "*Overview of Second Workshop*"

Steorts, Rebecca. 2018. "*Entity Resolution: Measuring and Reporting Quality*"

Winkler, William. 2018. "*Discussion of 'Entity Resolution: Measuring and Reporting Quality'*"

Reist, Ben. 2018. "*Leveraging Survey Methods to Improve Administrative Record Estimates*"

Jang, Don. 2018. "*Data Harmonization in Survey Data Integration*"

Holan, Scott. 2018. "*Recent Advances in Spatial and Spatio-Temporal Change of Support for Official Statistics*"

Reiter, Jerry. 2018. "*Blending Data through Statistical Matching, Modeling, and Imputation*"

Mulrow, Ed. 2018. "*Discussion of 'Blending Data through Statistical Matching, Modeling, and Imputation*'"

Sweeney, Latanya. 2018. "*The elusive sweet spots of privacy and utility*"

Abowd, John. 2018. "*Discussion of 'The elusive sweet spots of privacy and utility'*"

Third Workshop: Statistical Outputs (February 26, 2018)

Eltinge, John. 2018. "Introduction"

Chapman, Chris. 2018. "Recap of First Workshop: Lessons Learned"

Schafer, Joseph. 2018. "Recap of Second Workshop: Lessons Learned"

Young, Linda. 2018. "Overview of Third Workshop"

Langton, Lynn. 2018. "*Identifying and Addressing a Break (Blip) in Series*"

Eltinge, John. 2018. "*Discussion of 'Identifying and Addressing a Break (Blip) in Series*'"

Raghunathan, Trivellore. 2018. "*Combining Information from Multiple Data Sources: Challenges and Opportunities*"

Bell, William. 2018. "*Discussion of 'Combining Information from Multiple Data Sources: Challenges and Opportunities*'"

Biemer, Paul. 2018. "*Assessing and Improving the Accuracy of Estimators from Blended Data*"

Czajka, John. 2018. "*Transparency in the Reporting of Quality for Integrated Data: International Standards*"

Kreuter, Frauke. 2018. "*Summary*"

Appendix II. Integrated Data Projects in the Federal Statistical System

Statistical agencies are currently producing integrated data products, including those that include data not originally collected for statistical purposes (i.e. non-statistical data). Some of the non-statistical sources used in the following examples, and in other applications by agencies across the FSS involve both structured data (administrative records) and unstructured data (e.g. satellite imagery). Selected examples of integrated data sources used by ICSP agencies include the following.

- In 2012, The Bureau of Justice Statistics (BJS) spearheaded an initiative titled *National Crime Statistics Exchange (NCS-X)* in order to spur the development of the National Incident Based Reporting System (NIBRS) into a nationally representative source of detailed crime statistics. Initially, NCS-X recruited a select sample of agencies to report NIBRS data to the FBI. In 2013, BJS and the FBI signed a joint statement of support for NCS-X and agreed to work together to expand NIBRS coverage through funding, training, and technical assistance to law enforcement agencies across the United States. Information found at https://www.bjs.gov/content/ncsx.cfm.
- The Bureau of Economic Analysis (BEA) maintains a long-standing integrated data system for the System of National Accounts, which combines information from many sources to characterize quantitatively the structure of the economy and to generate a variety of economic statistics. More recently, the Bureau of Transportation Statistics (BTS) took a similar hybrid approach to estimating the movement of goods throughout the United States in its Freight Analysis Framework, Information found at https://www.bts.gov/faf.
- The Bureau of Labor Statistics (BLS) uses "alternative" (non-survey) data in the Producer Price Index and the Consumer Price Index to create sampling frames; to supplement, validate, and impute survey data; and to reduce cost and respondent burden (Presented at FCSM/WSS workshop on Input Data, Murphy and Konny, 2017).
- The U.S. Census Bureau (Census), among many examples, combines Unemployment Insurance earnings data with Quarterly Census of Employment and Wages (QCEW) data and other Federal administrative records to produce Longitudinal Employer-Household Dynamics (LEHD) statistics on employment and workforce changes as well as the connections between workers' home and employment locations. (McEntarfer, 2013)

- The Economic Research Service (ERS) uses proprietary data on retail food prices to combine with survey data on household food acquisitions and purchases; to calculate the cost of the Thrifty Food Plan, the basis for benefits in the Supplemental Nutrition Assistance Program (SNAP, formerly the Food Stamp Program); and to construct a Quarterly Food-at-Home Price Database (Presented at FCSM/WSS workshop on Input Data, Muth, 2017)
- The National Agricultural Statistics Service (NASS) uses satellite remote-sensing imagery, combined with other data sources, to add information to reports on crop production, yield estimates, and area. The spatial images improve both the quality and the timeliness of the information in the reports. (Presented at FCSM/WSS workshop on Input Data, Johnson, 2017)
- The National Center for Education Statistics (NCES) regularly collects data for National Postsecondary Student Aid Study (NPSAS); information found at https://nces.ed.gov/surveys/npsas/. NPSAS integrates several administrative data sources in combination with student survey data. Administrative data are drawn from public-sector and private-sector providers and include data from the Integrated Postsecondary Education Data System (IPEDS), the National Student Loan Data System (NSLDS), the Central Processing System (CPS), data from the Veterans Benefits Administration (VBA), the National Student Clearinghouse, and SAT and ACT results. With the exception of the SAT and ACT data, these administrative data sources are integrated into NPSAS to create better and more complete information on how students pay for postsecondary education. For a recent example of how these data are used, please see the recently released "2015–16 National Postsecondary Student Aid Study (NPSAS:16): Student Financial Aid Estimates for 2015–16" which describes the percentages of students receiving various types of financial aid and average amounts received, by type of institution attended, attendance pattern, dependency status, and income level.
- The National Center for Health Statistics (NCHS) links its population health survey data with vital records from the National Death Index. The integrated data are available for health and mortality researchers on a restricted-use basis; a separate version is available for public use. (NCHS, 2018)
- The biennial National Study of Long-Term Care Providers (NSLTCP) at NCHS produces statistical information to monitor state and national trends in the supply, provision, use and selected characteristics of paid, regulated post-acute and long-term care services. NSLTCP currently includes five sectors – adult day services centers, assisted living and similar residential

care communities, home health agencies, hospices, and nursing homes. NSLTCP uses CMS administrative data (certification, claims, clinical assessment) for three sectors for which nationally representative data are available (home health, hospice, nursing home), as well as nationally representative primary multi-mode survey data collection for two sectors (adult day, residential care) for which nationally representative data do not exist. This NSLTCP design allows NCHS to harmonize content across multiple administrative and survey data sources to enable comparisons among the five post-acute and long-term care sectors. NCHS uses this cost-effective, sustainable strategy to efficiently understand a spectrum of long-term care and post-acute care sectors, leveraging both existing administrative data and primary survey data collection. Information found at https://www.cdc.gov/nchs/nsltcp/index.htm.

- The National Center for Veterans Analysis and Statistics (NCVAS) prepared *Profile of Veterans: 2016*, the most recent of the agency's annual report, using the official Veteran population projection from the U.S. Department of Veterans Affairs (VA), the U.S. Veterans Eligibility Trends and Statistics, and the Census Bureau's American Community Survey (NCVAS, 2018).
- The Veterans Health Administration (VHA) links data from Veterans' VA electronic health record to data from the Centers for Medicare and Medicaid Services, including enrollment and utilization through Medicare Advantage and Medicaid, information on patients with end-stage renal disease, care quality, access, and cost. These data are available to approved VA researchers or affiliates. Information available from the VA Information Resource Center at https://www.virec.research.va.gov/VACMS/About.asp#About.

Appendix III. Data Quality Assessments at Statistical Agencies

There are numerous specific examples of statistical agencies considering several dimensions of data quality as part of their ongoing statistical programs. Three illustrative examples are described here from the Census Bureau, BLS, and ERS.

- The Data Acquisition and Curation area (DAC) of the Census Bureau manages over 150 interagency agreements to share administrative data and outside survey data sources (Presented at FCSM/WSS workshop on Input Data Quality, Berning and Sheppard, 2017). DAC evaluates possible data sources with a Data Quality Assessment Tool for Administrative Data (Iwig et al., 2013). The tool has six qualitative and quantitative dimensions that largely overlap with the dimensions of the ESS framework—*relevance, accuracy, coherence, accessibility, interpretability,* and *institutional environment*. The tool prompts many data quality assessment questions. For example, will the data source continue to be available and consistent over time? That issue crops up frequently as agencies consider integrating non-survey sources into production. Another question is about the laws governing the use of the data, which is an issue about accessibility.

- The BLS uses multiple data sources to construct prices indexes, with the goal to improve *accuracy* and *timeliness*—two key dimensions of data quality. As described in the first workshop (Presented at FCSM/WSS workshop on Input Data Quality, Murphy and Konny, 2017), when determining whether an alternative data source is fit for a particular use within the agency, the BLS uses a tool it calls an "Alternative Data Matrix". Several familiar dimensions are touched on by various rows in the matrix, including several used or recommended in other frameworks: *granularity, coverage, timeliness, sampling procedures*. Measures more specific to external data include *viability, reliability of the source, quality of the descriptive data* and *whether the methods are understood*.

- Ongoing research conducted by ERS is assessing the quality of store and household scanner data obtained from propriety sources; scanners are the laser readers that detect the barcodes on products, providing a record of which particular item was bought and its price. Several issues associated with the retail food price data concern the dimensions of *coherence* and *comparability*. For example, there were differences in the data provided by stores on whether data were available for individual stores (or aggregate for all locations) and whether prices were

affected by coupons or loyalty cards. (Presented at FCSM/WSS workshop on Input Data Quality, Muth, 2017).

- NCHS assesses the quality of the NCHS surveys linked to the National Death Index by comparing the estimates from the linked files to mortality estimates from the National Vital Statistics System. Results from earlier years of linked data can be found here: Mortality Experience of the 1986-2000 National Health Interview Survey Linked Mortality Files Participants, https://www.cdc.gov/nchs/data/series/sr_02/sr02_147.pdf.

Appendix IV. Possible Elements of a Data Quality Instrument

The following list identifies elements that may be useful in an instrument for transparent quality reporting. Many of the components in the outline below, particularly for sections 1-7, are drawn from the European Statistical System's *ESS Handbook for Quality Reports, 2014 Edition (2015)*.

1. Introduction

1.1. A description of the process and its outputs

1.2. A brief history of the statistical process and outputs in questions.

1.3. The broad statistical domain to which the outputs belong; related statistical outputs.

1.4. The boundary of the quality report at hand and references to related quality reports.

1.5. An overview of all output produced by the statistical process.

1.6. References to other documentation, especially on methodology.

2. Relevance, Assessment of User Needs and Perceptions

2.1. A content-oriented description of all statistical outputs.

2.2. Definitions of statistical target concepts (population, definition of units and aggregation formula) including discrepancies from ESS/international concepts. (May also be discussed under Coherence and Comparability.)

2.3. Information on completeness compared with relevant regulations/guidelines.

2.4. Available quality indicators.

2.5. Means of obtaining information on users and uses.

2.6. Description and classification of users.

2.7. Uses for which users want the outputs.

2.8. Unmet user needs, including reasons for not meeting them.

2.9. Users and uses given special consideration.

2.10. Means of obtaining user views.

2.11. Main results regarding user satisfaction.

2.12. Date of most recent user satisfaction survey.

3. Accuracy and reliability

3.1. Overall accuracy

3.1.1. Identification of the main sources of error for the main variables.

3.1.2. If micro-data are accessible for research purposes, it may be necessary to make additional comments to assist such uses.

3.1.3. A summary assessment of all sources of error with special focus on the key estimates.

3.1.4. An assessment of the *potential for bias* (sign and order of magnitude) for each key indicator in quantitative or quality terms.

3.2. Sampling Errors

3.2.1. Always applicable:

3.2.2. Where sampling is used there should be a section on sampling errors.

3.2.3. As far as possible sampling error should be presented for estimates of change in addition to estimate of level. If necessary, reasonable assumptions can be used.

3.2.4. If probability sampling is used:

3.2.4.1. There should be a presentation of sampling errors calculated according to formulas that should also be made available. If the estimators include adjustments for non-sampling errors, for example non-response, this should be explained and included also in the accuracy assessment.

3.2.4.2. The most appropriate presentational device should be chosen, normally CVs, range of CVs, or confidence intervals.

3.2.4.3. If outliers have received special treatment in estimation, this must be clearly described.

3.3. If non-probability sampling is used:

3.3.1. For sampling with cut-off an assessment of the accuracy due to the cut-off procedure should be included in addition to the presentation of sampling error for the sampled portion of the population.

3.3.2. For other forms of non-probability a sampling method can be invoked for the estimation of sampling error. A motivation for the chosen model and discussion of sampling bias should be included.

3.4. Coverage Errors

3.4.1. Quantitative information on overcoverage and multiple listings.

3.4.2. An assessment, preferably quantitative, on the extent of undercoverage and the bias risks associated with it.

3.4.3. Actions taken for reduction of undercoverage and associated bias risks.

3.4.4. Information on the frame: reference period, updating actions, and references to other documents on frame quality.

3.5. Measurement Errors

3.5.1. Identification and general assessment of the main risks in terms of measurement error.

3.5.2. If available, assessment based on comparisons with external data, re-interviews, experiments or data editing.

3.5.3. The efforts made in questionnaire design and testing, information on interviewer training and other work on error reduction.

3.5.4. Questionnaires used should be annexed (if very long by hyperlink)

3.6. Non-response errors

3.6.1. Non-response rates according to the most relevant definitions for the whole survey and for important sub-domains.

3.6.2. Item non-response rates for key variables.

3.6.3. A breakdown of non-respondents according to cause for non-response.

3.6.4. A qualitative statement on the bias risks associated with non-response.

3.6.5. Measures to reduce non-response.

3.6.6. Technical treatment of non-response at the estimation stage.

3.7. Processing Errors for Micro-data

3.7.1. Identification of the main issues regarding processing errors for the statistical process and its outputs.

3.7.2. Where relevant and available, an analysis of processing errors affecting individual observations should be presented; else a qualitative assessment should be included.

3.8. Accuracy for a Census

3.8.1. An evaluation/assessment of undercoverage and overcoverage.

3.8.1.1. A description of methods used to correct for undercoverage and overcoverage.

3.8.1.2. A description of methods and an assessment of the accuracy if a cut-off threshold is used.

3.8.2. An evaluation/assessment of measurement errors.

3.8.3. An evaluation/assessment of non-response errors.

3.8.4. An evaluation/assessment of processing errors.

3.9. ***Accuracy for a Statistical Process using Administrative Source(s)***

3.9.1. An evaluation/assessment of overcoverage, undercoverage and item non-response (missing data).

3.9.2. An evaluation/assessment of measurement errors.

3.9.3. For integration of several registers, an evaluation/assessment of the errors in units' identifiers and, in case of records linkage, errors in linkage. For event-reporting systems, an estimate/assessment of the rate of unreported events.

3.10. Accuracy for a Statistical Process Involving Multiple Data Sources

3.10.1. An overall description of the organisation of the process, the various segments and a summary of the quality aspects.

3.10.2. For each segment, the items as specified in the appropriate sections in these guidelines. These items should be grouped by error type.

3.10.3. When revisions of the estimates are released some information should be provided.

3.11. Accuracy for Price or Other Economic Index Process

3.11.1. Information on all sampling dimensions (for weights, products, outlets/companies, etc.)

3.11.2. Any attempt at estimating or assessing the sampling error in all or some of these dimensions.

3.11.3. Quality adjustment methods (including replacement and re-sampling rules) for at least major product groups.

3.11.4. Assessment of other types of error, where they could have a significant influence.

3.12. Accuracy for a Statistical Compilation

3.12.1. Information and indicators relating to accuracy for example as defined in the IMF's Data Quality Assessment Framework (DQAF) or other relevant, well accepted standard.

3.12.2. Analysis of revisions between successively published estimates.

3.12.3. For National Accounts:

3.12.4. Analysis of the causes of the statistical discrepancy

3.12.5. Assessment of non-observed economy.

3.13. Seasonal Adjustment

3.13.1. A short description of the method used.

3.13.2. A report on quality aspects in line with the ESS guidelines on seasonal adjustment.

3.14. Model Assumptions and Associated Errors

3.14.1. Models related to a specific source of error should be presented in the section concerned. This is recommended also in the case of a cut-off threshold and model-based estimate.

3.14.2. Domain specific models, for example, as needed to define the target of estimation itself, should be thoroughly described and their validity for the data at hand assessed.

3.15. Imputation

3.15.1. Information on the extent to which imputation is used.

3.15.2. A short description of the methods used and their effects on the estimates. (Typically this information will be reported in the section(s) dealing with the errors that imputation is helping to correct rather than in a separate section.)

3.16. Revisions

3.16.1. The revision policy

3.16.2. The number of revisions (planned and unplanned).

3.16.3. The average size of revisions (one or more measures).

3.16.4. The main reasons for revisions, and the extent to which the revisions improved accuracy.

4. Timeliness and Punctuality

4.1. For annual or more frequent releases: the average production time for each release of data.

4.2. For annual or more frequent releases: the percentage of releases delivered on time, based on scheduled release data.

4.3. The reasons for non-punctual releases explained.

5. Coherence and Comparability

5.1. General:

5.1.1. Brief descriptions of all conceptual and methodological metadata elements that could affect coherence/comparability.

5.1.2. An assessment (preferably quantitative) of the possible effect of each reported difference on the output values.

5.1.3. Differences between the statistical process and the corresponding European regulation/standard and/or international standard (if any).

5.2. Comparability—geographical:

5.2.1. A quantitative assessment of comparability across regions based on the (weighted) number of differences in metadata elements.

5.2.2. At ESS level, a coherence/comparability matrix summarising by region the possible sources of lack of comparability relative to a specified standard.

5.2.3. Mirror Statistics: Assessment of discrepancies (if any).

5.3. Comparability—over time:

5.3.1. Reference periods at which series breaks (if any) occurred, the reasons for them and treatments of them.

5.4. Coherence—National Accounts:

5.4.1. Where relevant, the results of comparisons with National Account framework and feedback from National Accounts with respect to coherence and accuracy problems.

5.5. Internal Coherence:

5.5.1. Any lack of coherence in the output of the statistical process itself.

6. Accessibility and Clarity, Dissemination Format

6.1. A description of the conditions of access to data: media, support, pricing policies, possible restrictions, etc.

6.2. A summary description of the information (metadata) accompanying the statistics (documentation, explanation, quality limitations, etc.).

6.3. The description should refer to both less sophisticated and more advanced users and how their needs have been taken into account.

6.4. A summary of user feedback on accessibility, clarity and dissemination format.

7. Statistical processing

7.1. Indicate if the data set is based on a survey, on administrative data sources, on a mix of multiple data sources or on data from other statistics activities. If samples surveys are used, some sample characteristics should also be given (e.g., population size, gross and net sample size, type of sampling design, reporting domain, etc.). If administrative registers are used, the description of registers should be given (source, primary purpose, etc.).

7.2. Include the frequency of data collection (e.g., monthly, quarterly, annually, continuous).

7.3. Describe the [data collection] method used, in case of surveys, to gather data from respondents (e.g. sampling methods, postal survey, CAPI, on-line survey, etc.)

7.4. Describe the procedures for checking and validating the source and output data and how the results of these validations are monitored and used.

7.5. Describe the data compilation process (e.g. imputation, weighting, adjustment for non-response, calibration, model used etc.).

7.6. Describe the time series to be adjusted and the statistical procedures and the used for adjusted the series.

8. Web Scraping

8.1. While not covered in the European Statistical System's *ESS handbook for quality reports, 2014 edition*, web scraping software is increasingly used as a means of extracting data from public facing websites that is not otherwise available as a structured or semi-structured data file. Even in cases where statistical agencies do not explicitly request data acquired through web scraping, alternative data provided to statistical agencies in structured format by data brokers may contain data procured using this extraction technique.

8.2. Web scraping as a low cost/high value means of collecting hard to find data has brought it to the attention of US statistical agencies as they seek to meet the growing expectations of the modern data consumers. However, the agencies leveraging data products acquired using this technique must be mindful of attendant legal considerations and limitations of the resulting data.

8.3. Indicate if the data set (or a subset of the data set) was acquired through web scraping.

8.4. A summary description of the information (metadata) accompanying the data set (documentation, explanation, quality limitations, volatility, etc.).

8.5. List the source web sites from which the data was extracted.

8.6. Describe the methods/products used to extract the source data.

8.7. Address legal considerations endemic to web scraping (e.g., use of copyright protected data, adherence to fair use standards, avoiding burden to the service of the sites being scraped, adherence to Terms of Use of the sites being scraped).

8.8. Include the frequency of data collection (e.g., monthly, quarterly, annually, continuous).

8.9. Describe the procedures for checking and validating the source and output data and how the results of these validations are monitored and used.

8.10. Describe the data compilation process (e.g. imputation, weighting, calibration, model used etc.).

9. Additional Criteria

9.1. Additional criteria that may warrant exploration for a draft instrument include:

9.2. Descriptions of modeling work, including outcome and predictor variables; model form and related parameterization; estimation and inference methods; goodness-of-fit results and other diagnostics; and related measures of inferential uncertainty

9.3. Quality measures appropriate for data sources like output from sensors, images, and object-oriented data.

9.4. Criteria appropriate for non-quantitative measures of quality.

9.5. Multiple forms of metadata

Appendix V. Research Topics and Activities

This Appendix lists research topics for informing transparent quality reporting for integrated data being considered by the FCSM Working Group and some possible activities that could inform these topics. A research agenda will be developed further by the FCSM Working Group in 2018 and 2019.

Research topics

1. Tools for improved transparent reporting on data quality, and improvement thereof, with emphasis on:
    a. Extension of survey-nonresponse theory and methods to develop diagnostic and adjustment methods to account for limitations of non-survey data sources, including:
        i. (sub)population coverage rates and associated biases
        ii. incomplete data patterns at the group, unit and item levels
    b. Extension of survey-measurement-error theory and methods to develop diagnostic and adjustment methods to account for data errors in non-survey data sources, including:
        i. unit problems
        ii. variable specification problems
        iii. measurement-quality issues specific to a given data type, including relatively simple categorical or continuous variables from structured data, as well as more complex data from text, sensors or images
    c. Extension of previously developed theory and methods for record linkage and data fusion to improve diagnostic and adjustment methods for:
        i. Improved measurement and modeling of the quality of record linkage, going beyond simple record linkage rates as such.
        ii. Improved procedures for data fusion
        iii. Develop a better understanding of the effects of having specific linkage variables, e.g. SSN, on linkage quality relative to less unique PII and the appropriate metrics for communicating to data users why a particular estimate from a linkage effort is high or low quality as they consider future projects and data collections

d. Extension of previously developed methods and theory for assessing quality of integrated data.

   i. Develop metrics for the quality dimensions of the survey data that are critical for high quality integrated data. For example, how much non-response bias is tolerable when calibrating non-survey data for national estimates?

   ii. Consider whether there are metrics related to latent modeling and latent variables that could inform quality reporting of integrated data.

e. Extension of previously developed methods and theory for measurement and improvement of quantitative features of the quality of published estimates (e.g., table entries or estimates of model parameters). In some cases, the predominant goal of the research would be to develop improved methods for quantitative assessment of the overall level of uncertainty attributable to a given published estimate. In other cases, the goals of the research would also include assessment of the uncertainty attributable to specific sources, potentially leading to specific quality-improvement interventions. Potential research areas include:

   i. Data-integration and quality assessment approaches from published literature on small-domain estimation and multiple-frame/multiple-mode surveys

   ii. Diagnostic and adjustment methods for "break in series" phenomena (e.g. bridge studies)

   iii. Sensitivity analyses appropriate for cases in which there is limited empirical information available on the magnitude of some error sources

   iv. "Propagation of error" techniques that account for the combined effects of multiple quality issues

   v. Empirical results regarding the level of temporal or cross-sectional granularity at which the applicable quality measures can be produced in a stable and informative way.

   vi. Examination of potential differences in relevant measurement and quality reporting for survey-integrated data by type of survey (e.g. establishment or household), when applicable.

f. Quantitative measures of output quality of microdata releases. This potentially includes datasets constructed through one or more of record linkage, data fusion and full-scale synthetic data creation

   i. Identifying the best communication strategies to transparently report (at the multiple levels of transparency) on the different strategies that can be used for linking records.

   ii. Identification and use of "truth decks" against which linked microdata files can be validated; conducting and reporting upon these validation studies.

   iii. Understanding linkage consent and the impact on multiple dimensions of quality of linked microdata files, overall and for subpopulations.

   iv. Understanding the impact of different quality protections on the linked microdata files and the implications for fitness of use for a range of likely analyses.

   v. Standardization of metadata for output microdata files.

   vi. Development of statistical methods for determining appropriate dissemination of linked micro-data files, which can have relatively high levels of disclosure risk.

g. Practical graphical display of selected quality measures covered in (a)-(e); this potentially includes both geographic and non-geographic data publications.

h. Output data quality - qualitative components like relevance and timeliness.

2. Tools to evaluate and improve key components of risk and cost

3. Methods for the empirical assessment of the connection between customary measures of data quality (e.g., mean squared errors, or confidence interval widths and coverage rates), cost and risk; and the perceived value delivered to selected groups of data users. One of many sub-cases would involve assessment of the risk that problems in one or more dimensions of quality could lead decision makers into a wrong decision, and whether the wrong decision causes inconvenience, inefficiency, serious harm, or worse.

Research activities

1. Development of open-source or open-access software to address one or more of the topics listed above.

2. Development of standardized toolsets for development of integrated-data products

3. Expanded training in selected areas of methodology and technology related to the topics listed above.
4. Develop a detailed inventory of quality reporting for structured, unstructured and integrated data developed elsewhere in government (outside the FSS) and in the private sector.
5. Explore effective means of communicating the wide array of concepts associated with producing estimates from integrated data. As discussed in the last workshop, data integration puts more of the effort on the back end than on the front end. With front end effort, communication of the assumptions becomes more standard. From the back end, the assumptions, methods, processes etc. are much broader. Guidance from university colleagues in statistical education or related areas with experience with risk communication may help guide the needed research.