

From Quality to Information Quality in Official Statistics

Ron S. Kenett¹ and Galit Shmueli²

The term *quality of statistical data*, developed and used in official statistics and international organizations such as the International Monetary Fund (IMF) and the Organisation for Economic Co-operation and Development (OECD), refers to the usefulness of summary statistics generated by producers of official statistics. Similarly, in the context of survey quality, official agencies such as Eurostat, National Center for Science and Engineering Statistics (NCSES), and Statistics Canada have created dimensions for evaluating the quality of a survey and its ability to report ‘accurate survey data’.

The concept of Information Quality, or InfoQ provides a general framework applicable to data analysis in a broader sense than summary statistics: InfoQ is defined as “the potential of a data set to achieve a specific (scientific or practical) goal by using a given empirical analysis method.” It relies on identifying and examining the relationships between four components: the analysis goal, the data, the data analysis, and the utility. The InfoQ framework relies on deconstructing the InfoQ concept into eight dimensions used for InfoQ assessment.

In this article, we compare and contrast the InfoQ framework and dimensions with those typically used by statistical agencies. We discuss how the InfoQ approach can support the use of official statistics not only by governments for policy decision making, but also by other stakeholders, such as industry, by integrating official and organizational data.

Key words: InfoQ; survey data; decision making; industry; government.

1. Introduction

Official statistics are produced by a variety of organizations including central bureaus of statistics, regulatory health care agencies, educational systems, and national banks. A common trend is the integration of official statistics and organizational data to derive insights at local and global levels.

One example is provided by the Intesa Sanpaolo Bank in Italy, which maintains an integrated database to support analytic research requests by management and various decision makers (Foresti et al. 2012). The bank uses regression models applied to internal data integrated with data from a range of official statistics providers, such as:

- Financial statements (CEBI),
- EPO patents (Thomson Scientific),
- Foreign direct investment (Reprint),
- ISO certificates (Accredia),

¹ KPA Ltd., Box 2525, Raanana 43100, Israel and University of Turin, Turin, Italy. Email: ron@kpa-group.com

² National Tsing Hua University, Institute of Service Science, 2, Hsinchu, 30013 Taiwan. Email: galit.shmueli@iss.nthu.edu.tw

Acknowledgment: The second author is partially funded by grant 104-2410-H-007-001-MY2 by Ministry of Science and Technology in Taiwan.

- Trademarks (UIBM, OIHM, USPTO, WIPO),
- Credit ratings (CEBI, Intesa Sanpaolo), and
- Corporate group charts (Intesa Sanpaolo).

In another example, data from surveys of companies in the north of Italy are combined with official data from the Italian stock exchange to calibrate the survey data (Dalla Valle 2014; Dalla Valle and Kenett 2015). The survey data is from Assolombarda, an Italian association of about 5,000 manufacturing and service companies located in the north of Italy. Assolombarda periodically collects data through questionnaires sent to the associated firms in order to gather information about the economic climate, firms' activity and production, and the number and types of employees. The FTSE-MIB is the official Italian stock market index for the Italian national stock exchange and consists of the 40 most-traded stock classes on the exchange. Dalla Valle and Kenett (2015) apply Bayesian networks to a dataset that integrates the Assolombarda surveys with information from the balance sheets of the 40 largest Italian firms on the Italian stock market. Proper data integration is key to this study.

In this article, we focus on information derived from an analysis of official statistics data with or without integration with other data sets. The objective is to provide decision makers with high-quality information. We use the Information Quality concept and framework by Kenett and Shmueli (2014; 2016) to evaluate the quality of such information to decision makers or other stakeholders.

Information Quality, or InfoQ, provides a general framework applicable to data analysis in a broader sense than summary statistics. InfoQ relies on identifying and examining the relationships between four components: the analysis goal, the data, the data analysis, and the utility. The first and last components, analysis goal and utility, require a translation between the particular domain and the world of statistics, data mining or another data-analytic field.

Official statistics "need to be used to be useful" (Forbes and Brown 2012, 91), and utility is one of the overarching concepts in official statistics. An issue that can lead to misconceptions and therefore a challenge to the translation of domain to statistics and back is that many of the terms used in official statistics have specific meanings in this context which are based on, but not identical to, their meaning in everyday usage. Forbes and Brown (2012, 91) state: "All staff producing statistics must understand that the conceptual frameworks underlying their work translate the real world into models that interpret reality and make it measurable for statistical purposes. . . . The first step in conceptual framework development is to define the issue or question(s) that statistical information is needed to inform. That is, to define the objectives for the framework, and then work through those to create its structure and definitions. An important element of conceptual thinking is understanding the relationship between the issues and questions to be informed and the definitions themselves."

In an interview-based study of 58 educators and policy makers, Hambleton (2002) found that the majority misinterpreted the official statistics proficiency reports that compare results across grades and across years. This finding was particularly distressing as policy makers rely on such reports for the appropriation of funds and other key decisions. In terms of information quality, the quality of the information provided by the reports was low. The translation from statistics to domain-specific insights was faulty.

The US Environmental Protection Agency (EPA), together with the Department of Defense and Department of Energy, launched the Quality Assurance Project Plan

(see EPA 2005, 35), which presents “steps . . . to ensure that environmental data collected are of the correct type and quality required for a specific decision or use.” They used the term Data Quality Objectives to describe “statements that express the project objectives (or decisions) that the data will be expected to inform or support.” These statements relate to descriptive goals, such as “Determine with greater than 95% confidence that contaminated surface soil will not pose a human exposure hazard.” These statements are used to guide the data-collection process. They are also used to assess the resulting data quality.

Central bureaus of statistics are now combining surveys with administrative data in dynamically updated studies that have replaced the traditional census approach, so that the proper integration of data sources is becoming a critical requirement. We suggest that evaluating InfoQ can significantly contribute to the range of examples described above. The article proceeds as follows: Section 2 reviews the InfoQ dimensions proposed by Kenett and Shmueli (2014), putting them in the context of official statistics research studies. Section 3 presents quality standards applicable to official statistics and their relationship with InfoQ dimensions, and Section 4 describes standards used in customer surveys and their relationship to InfoQ. We conclude with a discussion and suggest directions for further work in Section 5.

2. Information Quality (InfoQ) and Official Statistics

InfoQ is defined as “the potential of a data set to achieve a specific (scientific or practical) goal by using a given empirical analysis method” (Kenett and Shmueli 2014, 3). InfoQ is determined by the data (X), the data analysis (f) and the analysis goal (g), as well as by the relationships between them. Utility is measured using specific metric(s) (U). By examining each of these components and their relationships, we can learn about the contribution of a given study as a source of knowledge and insight. The components of InfoQ have been mapped onto eight dimensions that represent a deconstruction of the concept. Here, we present the eight InfoQ dimensions and provide some guiding questions that can be used in planning, designing, and evaluating official statistics reports.

2.1. Data Resolution

Data resolution refers to the measurement scale and aggregation level of the data. The data’s measurement scale should be carefully evaluated in terms of its suitability to the goal, the analysis methods used, and the required resolution of the utility U . Questions one could ask to figure out the strength of this dimension include:

- Is the data scale used aligned with the stated goal of the study?
- How reliable and precise are the data sources and data-collection instruments used in the study?
- Is the data analysis suitable for the data aggregation level?

A low rating on data resolution can be indicative of low trust in the usefulness of the study’s findings.

An example of data resolution is provided by Google’s ability to predict the prevalence of flu based on the type and extent of Internet search queries (www.ft.com/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html#axzz2y6ASfagk). These predictions

match the official figures published by the Centers for Disease Control and Prevention (CDC) quite well. The point is that Google's tracking has only a day's delay, compared to the week or more it takes for the CDC to assemble a picture based on reports from doctors' surgeries. Google is faster because it is tracking the outbreak by finding a correlation between what people search for online and whether they have flu symptoms. An application aiming to achieving an online prediction of flu prevalence might combine the weekly updated CDC official statistics with the dynamically updated Google estimates.

2.2. Data Structure

Data structure relates to the type(s) of data and data characteristics such as corrupted and missing values due to the study design or data-collection mechanism. Data types include structured numerical data in different forms (e.g., cross-sectional, time series, network data) as well as unstructured, non-numerical data (e.g., text, text with hyperlinks, audio, video, and semantic data). The InfoQ level of a certain data type depends on the goal at hand. Questions to ask to figure out the strength of this dimension include:

- Is the type of data used aligned with the stated goal of the study?
- Are data-integrity details (corrupted/missing values) described and handled appropriately?
- Are the analysis methods suitable for the data structure?

A low rating on data structure can be indicative of poor data coverage in terms of the project goals. For example, using a cross-sectional analysis method to analyze a time series warrants special attention when the goal is parameter inference, but is of less concern if the goal is forecasting future values. Another example is removing records with missing data when missingness might not be random. A paper analyzing online transactions with the objective of evaluating actual behavior versus declared behavior also needs data on declared behavior through focused queries or questionnaires. Without this, the structure of the data will not provide adequate information quality.

2.3. Data Integration

With the variety of data sources and data types available today, studies sometimes integrate data from multiple sources and/or types to create new knowledge regarding the goal at hand. Such integration can increase InfoQ, but in other cases it can reduce InfoQ, for example by creating privacy breaches (for a video of methodologies discussed during the 2011 international privacy data day, see https://www.youtube.com/watch?v=QES3-X0U1Q_Q). Questions to ask to figure out the strength of this dimension include:

- Are the data integrated from multiple sources? If so, what is the credibility of each source?
- How is the integration performed? Are there linkage issues that lead to dropping crucial information?
- Does the data integration add value in terms of the stated goal?
- Does the data integration cause any privacy or confidentiality concerns?

A low rating on data integration can be indicative of missed potential in data analysis.

A prime example of data integration is the fusion feature in Google (<https://support.google.com/fusiontables/answer/2571232>). In the risk in open source software (RISCOSS) FP7 project, a methodology was developed that aggregates quantitative data captured from OSS communities with qualitative expert opinion through an assessment of risk scenarios to derive risk indicators using Bayesian networks (see www.riscoss.eu). Another example, focused on analyzing semantic data for risk assessment based on the MUSING FP6 project, is presented in Kenett and Raanan (2010). Another example of data integration is the combination of structured and unstructured semantic data (Figini et al. 2010). See also Penny and Reale (2004) and Vicard and Scanu (2012).

2.4. *Temporal Relevance*

The process of deriving knowledge from data can be placed on a timeline that includes the periods of data collection, data analysis, and usage of results as well as the temporal gaps between these three stages. The different durations and gaps can each affect InfoQ. The data-collection duration can increase or decrease InfoQ, depending on the study goal, for example studying longitudinal effects versus a cross-sectional goal. Similarly, if the collection period includes uncontrollable transitions, this can be useful or disruptive, depending on the study goal. Questions to ask to figure out the strength of this dimension include:

- Considering the data collection, data analysis and deployment stages, are any of them time-sensitive?
- Does the time gap between data collection and analysis cause any concern?
- Is the time gap between the data collection and analysis and the intended use of the model (e.g., in terms of policy recommendations) of any concern?

A low rating on temporal relevance can be indicative of an analysis with low relevance to decision makers due to data collected in a different contextual condition. This can happen in economic studies with policy implications that are based on old data.

2.5. *Chronology of Data and Goal*

The choice of variables to collect, the temporal relationship between them, and their meaning in the context of the goal at hand affects InfoQ. Questions to ask to figure out the strength of this dimension include:

- If the stated goal is predictive, are all the predictor variables expected to be available at the time of prediction?
- If the stated goal is causal, do the causal variables precede the effects?
- In a causal study, are there issues of endogeneity (reverse-causation)?

A low rating on chronology of data and goal can be indicative of low relevance of a specific data analysis due to misaligned timing. A customer-satisfaction survey that was designed to be used as input to the annual budget planning cycle becomes irrelevant if its results are communicated after the annual budget is finalized (Kenett and Salini 2012).

2.6. Generalizability

The utility of $f(X|g)$ is dependent on the ability to generalize f to the appropriate population. Two types of generalizability are statistical generalizability and scientific generalizability. Statistical generalizability refers to inferring from a sample to a target population. Scientific generalizability refers to applying a model based on a particular target population to other populations. This can mean either generalizing an estimated population pattern/model f to other populations, or applying f estimated from one population to predict individual observations in other populations. Determining the level of generalizability requires careful characterization of g . Generalizability is related to the concepts of reproducibility, repeatability, and replicability (Drummond 2009; Banks 2011; Kenett and Zacks 2014; McNutt 2014; Kenett and Shmueli 2015). Reproducibility represents insights that are replicable (but not necessarily identical), while repeatability is about achieving the same results in a repeated experiment. Replicability is used most often in genome wide association studies where a follow up experiment is conducted to identify a subset of genes as active, following a large study investigating thousands of genes. Repeatability relates to data quality and analysis quality, while reproducibility relates to InfoQ. Questions to ask to figure out the strength of this dimension include:

- Is the stated goal statistical or scientific generalizability?
- For statistical generalizability in the case of inference, does the study under review answer the question “What population does the sample represent?”
- For generalizability in the case of a stated predictive goal (predicting the values of new observations; forecasting future values), are the results generalizable to the data to be predicted?

In the context of item response studies, Georg Rasch used the term *specific objectivity* to describe that case essential to measurement in which “comparisons between individuals become independent of which particular instruments – tests or items or other stimuli – have been used. Symmetrically, it is thought to be possible to compare stimuli belonging to the same class – measuring the same thing – independent of which particular individuals, within a class considered, were instrumental for comparison“ (Rasch 1977, 58). The term *general objectivity* is reserved for the case in which absolute measures (i.e., amounts) are independent of which instrument (within a class considered) is employed, and no other object is required. By “absolute” we mean the measure “is not dependent on, or without reference to, anything else; not relative” (ibid.). Similar constructs apply to economic and sociological studies.

2.7. Operationalization

Two types of operationalization are considered: construct operationalization and action operationalization of the analysis results. Constructs are abstractions that describe a phenomenon of theoretical interest. Measurable data are an operationalization of underlying constructs. The relationship between the underlying construct and its operationalization can vary, and its level relative to the goal is another important aspect of InfoQ. The role of construct operationalization is dependent on the goal, and especially on whether the goal is explanatory, predictive, or descriptive. In explanatory models, based on underlying causal theories, multiple operationalizations might be acceptable for

representing the construct of interest. As long as the data are assumed to measure the construct, the variable is considered adequate. In contrast, in a predictive task, where the goal is to create sufficiently accurate predictions of a certain measurable variable, the choice of operationalized variable is critical. Action operationalizing results refers to three questions posed by [W. Edwards Deming \(1982\)](#):

- What do you want to accomplish?
- By what method will you accomplish it?
- How will you know when you have accomplished it?

Questions to ask to figure out the strength of construct operationalization include:

- Are the measured variables themselves of interest to the study goal, or is their underlying construct of interest?
- What are the justifications for the choice of variables?

Questions to ask to figure out the strength of operationalizing results include:

- Who can be affected (positively or negatively) by the research findings?
- What can he or she do about it?
- Who else?

A low rating on operationalization indicates that the study might have academic value but in fact has no practical impact.

2.8. Communication

Effective communication of the analysis and its utility directly impacts InfoQ. There are plenty of examples where the miscommunication of valid results has led to problematic outcomes. For a study of how to make National Assessment of Educational Progress (NAEP) and state test score reporting scales and reports more understandable, see [Hambleton \(2002\)](#). Questions that a reviewer should ask to figure out the strength of this dimension include:

- Is the exposition of the goal, data, and analysis clear?
- Is the exposition level appropriate for the readership of this report?

A low rating on communication can indicate that poor communication might cover the true value of the analysis and, thereby, dump the value of the information provided by the analysis.

3. Quality Standards for Official Statistics

A concept of *Quality of Statistical Data* was developed and used in European official statistics and international organizations such as the International Monetary Fund (IMF) and the Organisation for Economic Co-operation and Development (OECD). This concept refers to the usefulness of summary statistics produced by national statistics agencies and other producers of official statistics. In this context, quality is evaluated in terms of the statistics' usefulness for a particular goal. The OECD uses seven dimensions for quality assessment: *relevance, accuracy, timeliness and punctuality, accessibility, interpretability,*

coherence, and *credibility* (see chap. 5 in Giovanini 2008). Eurostat's quality dimensions are *relevance of statistical concept*, *accuracy of estimates*, *timeliness and punctuality in disseminating results*, *accessibility and clarity of the information*, *comparability*, *coherence*, and *completeness*. (Eurostat 2003, 2009). See also Statistics Canada (2002), UK Department of Health (2004) and Office for National Statistics (2007).

In the United States, the National Center for Science and Engineering Statistics (NCSES), formerly the Division of Science Resources Statistics, was established within the National Science Foundation with general responsibility for statistical data. Part of its mandate is to provide information that is useful to practitioners, researchers, policy-makers, and the public. NCSES prepares about 30 reports a year based on surveys.

The purpose of survey standards is to set a framework for assuring data and reporting quality. Guidance documents are meant to help

- (1) increase the reliability and validity of data,
- (2) promote a common understanding of the desired methodology and processes,
- (3) avoid duplication and promote the efficient transfer of ideas, and
- (4) remove ambiguities and inconsistencies.

The goal is to provide the clearest possible presentations of data and their analysis. Guidelines typically focus on technical issues involved in the work rather than issues of contract management or publication formats (Biemer et al. 2003, 2012).

Specifically, NCSES aims to adhere to the ideals set forth in "Principles and Practices for a Federal Statistical Agency" (Citro and Straf 2006). As NCSES is a US federal statistical agency, NCSES surveys must follow guidelines and policies as set forth in the Paperwork Reduction Act and other legislation related to surveys. For example, NCSES surveys must follow the implementation guidance, survey clearance policies, response-rate requirements, and related orders prepared by the Office of Management and Budget (OMB). The following standards are based on US government standards for statistical surveys (see www.nsf.gov/statistics/). We list them in Table 1 with an annotation mapping them onto InfoQ dimensions, where relevant.

I. Development of Concepts, Methods, and Design

Survey Planning

Standard 1.1: Agencies initiating a new survey or major revision of an existing survey must develop a written plan that sets forth a justification, including: goals and objectives; potential users; the decisions the survey is designed to inform; key survey estimates; the precision required of the estimates (e.g., the size of differences that need to be detected); the tabulations and analytic results that will inform decisions and other uses; related and previous surveys; steps taken to prevent unnecessary duplication with other sources of information; when and how frequently users need the data; and the level of detail needed in tabulations, confidential microdata, and public-use data files.

This standard requires the explicit declaration of goals and methods for communicating results. It also raises the issue of data resolution in terms of dissemination and generalization (estimate precision).

Table 1. Relationship between NCSES standards and InfoQ dimensions.

NCSES standard		Data resolution	Data structure	Data integration	Temporal relevance	Data-Goal chronology	Generalizability	Operationalization	Communication
I. Development of concepts, methods and design	Survey planning	■					■		■
	Survey design	■	■	■			■		
	Survey response rate						■	■	
	Survey pretesting	■	■						
II. Collection of data	Developing sampling frames					■		■	
	Required notifications to potential survey respondents							■	
	Data-collection methodology								
III. Processing and editing of data	Data editing								
	Nonresponse analysis and response-rate calculation						■		
	Coding								
	Data protection								
	Evaluation								■
IV. Production of estimates and projections	Developing estimates and projections						■		
V. Data analysis	Analysis and report planning								
	Inference and comparisons						■		
VI. Review procedures	Review of information products								
VII. Dissemination of information products	Releasing information					■			■
	Data protection and disclosure avoidance for dissemination								
	Survey documentation								■
	Documentation and release of public-use microdata								

Survey Design

Standard 1.2: Agencies must develop a survey design, including defining the target population, designing the sampling plan, specifying the data-collection instrument and methods, developing a realistic timetable and cost estimate, and selecting samples using generally accepted statistical methods (e.g., probabilistic methods that can provide estimates of sampling error). Any use of nonprobability sampling methods (e.g., cut-off or

model-based samples) must be justified statistically and be able to measure estimation error. The size and design of the sample must reflect the level of detail needed in tabulations and other data products, and the precision required of key estimates. Documentation of each of these activities and resulting decisions must be maintained in the project files for use in documentation (see Standards 7.3 and 7.4).

This standard advises on data resolution, data structure and data integration. The questionnaire design addresses the issue of construct operationalization, and estimation error relates to generalizability.

Survey Response Rates

Standard 1.3: Agencies must design the survey to achieve the highest practical rates of response, commensurate with the importance of survey uses, respondent burden, and data-collection costs, to ensure that survey results are representative of the target population so that they can be used with confidence to inform decisions. Nonresponse bias analyses must be conducted when unit or item response rates or other factors suggest the potential for bias to occur.

The main focus here lies on statistical generalization. In a sense, this standard also deals with action operationalization. The survey must be designed and conducted in a way that encourages respondents to take action and respond.

Pretesting Survey Systems

Standard 1.4: Agencies must ensure that all components of a survey function are as intended when implemented in the full-scale survey and that measurement error is controlled by conducting a pretest of the survey components or by having successfully fielded the survey components on a previous occasion.

Pretesting relates to data resolution and the question of whether the collection instrument is sufficiently reliable and precise.

II. Collection of Data

Developing Sampling Frames

Standard 2.1: Agencies must ensure that the frames for the planned sample survey or census are appropriate for the study design and are evaluated against the target population for quality.

Sampling-frame development is crucial for statistical generalization. Here we also ensure the chronology of data and goal in terms of the survey deployment.

Required Notifications to Potential Survey Respondents

Standard 2.2: Agencies must ensure that each collection of information instrument clearly states the reasons the information is planned to be collected; the way such information is planned to be used to further the proper performance of the functions of the agency; whether responses to the collection of information are voluntary or mandatory (citing authority); the nature and extent of confidentiality to be provided, if any, citing authority;

an estimate of the average respondent burden together with a request that the public direct to the agency any comments concerning the accuracy of this burden estimate and any suggestions for reducing this burden; the control number; and a statement that an agency may not conduct and a person is not required to respond to an information collection request unless it displays a currently valid control number.

This is another aspect of action operationalization.

Data-Collection Methodology

Standard 2.3: Agencies must design and administer their data-collection instruments and methods in a manner that achieves the best balance between maximizing data quality and controlling measurement error while minimizing respondent burden and cost.

III. Processing and Editing of Data

The standards in this section focus upon the data component, and in particular upon assuring data quality and confidentiality.

Data Editing

Standard 3.1: Agencies must edit data appropriately, based on available information, to mitigate or correct detectable errors.

Nonresponse Analysis and Response-Rate Calculation

Standard 3.2: Agencies must appropriately measure, adjust for, report, and analyze unit and item nonresponse to assess their effects on data quality and to inform users. Response rates must be computed using standard formulas to measure the proportion of the eligible sample that is represented by the responding units in each study, as an indicator of potential nonresponse bias.

Coding

Standard 3.3: Agencies must add codes to collected data to identify aspects of data quality from the collection (e.g., missing data) in order to allow users to appropriately analyze the data. Codes added to convert information collected as text into a form that permits immediate analysis must use standardized codes, when available, to enhance comparability.

Data Protection

Standard 3.4: Agencies must implement safeguards throughout the production process to ensure that survey data are handled to avoid disclosure.

Evaluation

Standard 3.5: Agencies must evaluate the quality of the data and make the evaluation public (through technical notes and documentation included in reports of results or

through a separate report) to allow users to interpret results of analyses, and to help designers of recurring surveys focus improvement efforts.

This relates to communication.

IV. Production of Estimates and Projections

Developing Estimates and Projections

Standard 4.1: Agencies must use accepted theory and methods when deriving both direct survey-based estimates as well as model-based estimates and projections that use survey data. Error estimates must be calculated and disseminated to support assessment of the appropriateness of the uses of the estimates or projections. Agencies must plan and implement evaluations to assess the quality of the estimates and projections.

This standard is aimed at statistical generalizability and focuses on the quality of the data analysis (deriving estimates can be considered part of the data-analysis component).

V. Data Analysis

Analysis and Report Planning

Standard 5.1: Agencies must develop a plan for the analysis of survey data prior to the start of a specific analysis to ensure that statistical tests are used appropriately and that adequate resources are available to complete the analysis.

This standard once again focuses on analysis quality.

Inference and Comparisons

Standard 5.2: Agencies must base statements of comparisons and other statistical conclusions derived from survey data on acceptable statistical practice.

VI. Review Procedures

Review of Information Products

Standard 6.1: Agencies are responsible for the quality of information that they disseminate and must institute appropriate content/subject-matter, statistical, and methodological review procedures to comply with OMB and agency Information Quality Guidelines.

VII. Dissemination of Information Products

Releasing Information

Standard 7.1: Agencies must release information intended for the general public according to a dissemination plan that provides for equivalent, timely access to all users and provides information to the public about the agencies' dissemination policies and procedures, including those related to any planned or unanticipated data revisions.

This standard touches on the chronology of data and goal as well as on communication.

Data Protection and Disclosure Avoidance for Dissemination

Standard 7.2: When releasing information products, agencies must ensure strict compliance with any confidentiality pledge to the respondents and all applicable federal legislation and regulations.

Survey Documentation

Standard 7.3: Agencies must produce survey documentation that includes those materials necessary to understand how to properly analyze data from each survey, as well as the information necessary to replicate and evaluate each survey's results (see also Standard 1.2). Survey documentation must be readily accessible to users, unless it is necessary to restrict access to protect confidentiality.

Proper documentation is essential for proper communication.

Documentation and Release of Public-Use Microdata

Standard 7.4: Agencies that release microdata to the public must include documentation clearly describing how the information is constructed and provide the metadata necessary for users to access and manipulate the data (see also Standard 1.2). Public-use microdata documentation and metadata must be readily accessible to users.

This standard aims at adequate communication of the data (not the results).

These standards provide a comprehensive framework for the various activities involved in planning and implementing official statistics surveys.

The next section focuses on customer-satisfaction surveys, such as the surveys on Service of General Interest (SGI) conducted within the European Union (EU).

4. Standards for Customer Surveys

Customer satisfaction, according to the ISO10004:2010 standards of the International Organization for Standardization (ISO), is the “customer’s perception of the degree to which the customer’s requirements have been fulfilled”. It is “determined by the gap between the customer’s expectations and the customer’s perception of the product [or service] as delivered by the organization” ([ISO/TS 10004 2010](#)).

The ISO describes the importance of standards on its website:

“ISO is a non-governmental organization that forms a bridge between the public and private sectors. Standards ensure desirable characteristics of products and services such as quality, environmental friendliness, safety, reliability, efficiency and interchangeability – and at an economical cost.”

ISO’s work program ranges from standards for traditional activities such as agriculture and construction, through mechanical engineering, manufacturing and distribution, to transport, medical devices, information and communication technologies, standards for good management practice and for services. Its primary aim is to share concepts, definitions and tools to guarantee that products and services meet expectations. When standards are absent, products may turn out to be of poor quality, might be incompatible with available equipment, or could be unreliable or even dangerous.

The goals and objectives of customer-satisfaction surveys are described clearly in ISO 10004:

“The information obtained from monitoring and measuring customer satisfaction can help identify opportunities for improvement of the organization’s strategies, products, processes and characteristics that are valued by customers, and serve the organization’s objectives. Such improvements can strengthen customer confidence and result in commercial and other benefits”.

In the following, we give a brief description of the ISO 10004 standard.

ISO 10004 Guidelines for Monitoring and Measuring Customer Satisfaction

The rationale of the ISO 10004 standard – as reported in Clause 1 – is to provide “guidance in defining and implementing processes to monitor and measure customer satisfaction”. It is intended for use “by organizations regardless of type, size or product provided” but it is related only “to customers external to the organization”.

The ISO approach outlines three phases in the processes of measuring and monitoring customer satisfaction:

- (1) Planning (Clause 6);
- (2) Operation (Clause 7);
- (3) Maintenance and Improvement (Clause 8).

Planning

The planning phase “refers to the definition of the purposes and objectives of measuring customer satisfaction and the determination of the frequency of data gathering (regularly, on an occasional basis, dictated by business needs or specific events)”. For example, an organization might be interested in investigating reasons for customer complaints after the release of a new product or causes of a loss of market share. Alternatively, it might want to compare its position relative to other organizations regularly. Moreover, “[i]nformation regarding customer satisfaction might be obtained indirectly from the organization’s internal processes (e.g., customer complaints handling) or from external sources (e.g., reported in the media) or directly from customers”.

In determining the frequency of data collection, this clause relates to chronology of data and goal as well as to temporal relevance. “[D]efinition of . . . customer satisfaction” concerns construct operationalization. The collection of data from different sources indirectly touches on data structure and resolution. However, the use of “or” for choice of data source indicates no intention to integrate data.

Operation

The operation phase represents the core of the standard and introduces the operational steps an organization should follow in order to meet the requirements of ISO 10004. These steps are:

- (a) identify the customers (current or potential) and their expectations,
- (b) gather customer-satisfaction data directly from customers through a survey and/or indirectly examining existing sources of information, after having identified

Table 2. Relationship between ISO 10004 guidelines and InfoQ dimensions.

ISO 10004 Phase	Data Resolution	Data Structure	Data Integration	Temporal Relevance	Data-Goal Chronology	Generalizability	Operationalization	Communication
Planning								
Operation								
Maintenance and improvement								

the main characteristics related to customer satisfaction (product, delivery, or organizational characteristic),

- (c) analyze customer-satisfaction data after having chosen the appropriate method of analysis,
- (d) communicate customer-satisfaction information,
- (e) monitor customer satisfaction at defined intervals to control that “the customer-satisfaction information is consistent with, or validated by, other relevant business performance indicators” (Clause 7.6.5).

Statistical issues mentioned in ISO 10004 relate to the number of customers to be surveyed (sample size), the method of sampling (Clause 7.3.3.3 and Annex C.3.1, C3.2), and the choice of the scale of measurement (Clause 7.3.3.4 and Annex C.4).

Identifying the population of interest and sample design relate to generalization. Communication is central to step (d). Step (e) refers to data integration, and the choice of measurement scale relates to data resolution.

Maintenance and Improvement

The maintenance and improvement phase includes periodic review, evaluation, and continual improvement of processes for monitoring and measuring customer satisfaction.

This phase aims at maintaining generalizability and temporal relevance, as well as the appropriateness of construct operationalization (“reviewing the indirect indicators of customer satisfaction”). Data integration is used to validate the information against other sources, and communication and actionable operationalization are also mentioned. [Table 2](#) summarizes the relationship between the three phases of the ISO 10004 and the eight InfoQ dimensions.

5. Conclusions and Discussion

The present article begins by referring to two examples where official statistics data are combined with organizational data in order to derive information of higher quality through analysis. In the Intesa Sanpaolo Bank example, the competitiveness of an enterprise was

assessed using factors such as innovation and R&D, intangibles (such as human capital, brands, quality, and environmental awareness), and foreign direct investment. Some of the challenges encountered when establishing a coherent integrated database included incomplete matching using “tax ID number” as the key, since patent, certification and trademark archives contain only the business name and address of the enterprise. As a result, an algorithm was developed for matching a business name and address to other databases containing both the same information and the tax ID number. With this approach, different business names and addresses may appear for the same enterprise (for instance, abbreviated names, acronyms with or without full stops, presence of the abbreviated legal form, etc.). The tax ID number of an enterprise may also change over the years. Handling these issues properly is key to the quality of information generated by regression analysis. These aspects are related to data resolution, data structure, data integration, temporal relevance, and chronology of data and goal, four of the InfoQ dimensions. Therefore, considering each of these InfoQ dimensions, with their associated questions, can help guide the analyst to detect and formalize the challenges to the final information quality.

In the study of [Dalla Valle \(2014\)](#), data from a survey about the economic climate, firms’ activity and production, and the number and types of employees of 167 firms located in the provinces of Milan and Lodi in the North of Italy were combined with official statistics data from the Italian national stock exchange. An analysis combining Vines and Bayesian networks permits the proper calibration of the data, thus strengthening the quality of the information derived from the survey. The InfoQ dimensions involved in this work also include data resolution, data structure, data integration, temporal relevance and chronology of data and goal. These two examples demonstrate how concern for the quality of the information derived from an analysis of a given data set requires that attention be paid to several dimensions beyond the quality of the analysis method used. The eight InfoQ dimensions provide a general template for identifying and evaluating such challenges.

With the increased availability of data sources and ubiquity of analytic technologies, the challenge of transforming data into information and knowledge is growing in importance ([Kenett 2008](#)). Official statistics play a critical role in this context and applied research, using official statistics, needs to ensure the generation of high-information quality. In this article, we discuss the various elements that determine the quality of such information and describe several proposed approaches for achieving it. Specifically, we compare the InfoQ concept of [Kenett and Shmueli \(2014\)](#) with NCSES and ISO standards. InfoQ is a general approach that has been applied to a wide range of applied research applications such as education, healthcare, risk management and customer-satisfaction surveys (for more examples see [Kenett and Shmueli 2016](#)). Here we discuss examples of how official statistics data and data from internal sources are integrated to generate higher information quality. These various guidelines and initiatives focus on identifying what was learned from the data analysis and was actually done within the analysis framework. These two aspects are related to the issue of reproducibility and replicability of research. The terminology in this context is not unified ([Kenett and Shmueli 2015](#)). For example, quoting [Drummond \(2009\)](#):

“Reproducibility requires changes; replicability avoids them. A critical point of reproducing an experimental result is that irrelevant things are intentionally not replicated. One might say, one should replicate the result not the experiment.”

In contrast, Banks (2011) notes:

“As a former editor of the Journal of the American Statistical Association, my own sense is that very few applied papers are perfectly reproducible. Most do not come with code or data, and even if they did, I expect a careful check would find discrepancies from the published paper. The reasons are innocent: code written by graduate students is continually tweaked and has sketchy documentation. The exact data cleaning procedures are not perfectly remembered when the final version of the paper is written, or may be muddled by miscommunication among multiple authors. And even if a conscientious researcher provided a full description of every cleaning step, every model fitting choice, and all aspects of variable selection, the resulting paper would be so long and tedious that no doubt the foolish editor would demand that it be shortened.”

The expanded view of information quality, embedded in the InfoQ dimensions, is an attempt to clearly map both reproducibility and replicability components in a research study. From any report derived from data analysis we would like to understand both what was learned and how it was achieved. Our work is an initial step in such an endeavor and further tools, and methodology needs to be developed in order to support it.

6. References

- Banks, D. 2011. “Reproducible Research: A Range of Response.” *Statistics, Politics, and Policy* 2: 2151–7509. Doi: <http://dx.doi.org/10.2202/2151-7509.1023>.
- Biemer, P. and L. Lyberg. 2003. *Introduction to Survey Quality*. Hoboken: John Wiley and Sons.
- Biemer, P.P., D. Trewin, H. Bergdahl, L. Japac, and Å. Pettersson. 2012. A Tool for Managing Product Quality. Eur. Conf. Quality in Official Statistics, Athens.
- Citro, C.F. and M.L. Straf. 2006. Principles and Practices for a Federal Statistical Agency: Fifth Edition, The National Washington, DC: Academies Press.
- Dalla Valle, L. 2014. “Official Statistics Data Integration Using Copulas.” *Quality Technology and Quantitative Management* 11: 111–131. Doi: <http://dx.doi.org/10.1080/16843703.2014.11673329>.
- Dalla Valle, L. and R.S. Kenett. 2015. “Official Statistics Data Integration to Enhanced Information Quality.” *Quality and Reliability Engineering International* 31: 1281–1300. Doi: <http://dx.doi.org/10.1002/qre.1859>.
- Deming, W.E. 1982. *Out of the Crisis*. Cambridge: MIT Press.
- Drummond, C. 2009. “Replicability is not Reproducibility: Nor is it Good Science.” In *Proceedings of the Evaluation Methods for Machine Learning Workshop at the 26th ICML*, June 14–18, Montreal, Canada. Available at: https://www.researchgate.net/publication/228709155_Replicability_is_Not_Reproducibility_Nor_is_It_Good_Science (accessed September 5, 2016).

- EPA. 2005. Uniform Federal Policy for Quality Assurance Project Plans: Evaluating, Assessing, and Documenting Environmental Data Collection and Use Programs. Available at: www.epa.gov/fedfac/pdf/ufp_qapp_v1_0305.pdf (accessed 20 October, 2015).
- Eurostat. 2003. *Standard Quality Report*. Luxembourg: Eurostat.
- Eurostat. 2009. *Handbook for Quality Reports*. Luxembourg: Eurostat.
- Figini, S., R.S. Kenett, and S. Salini. 2010. "Optimal Scaling for Risk Assessment: Merging of Operational and Financial Data." *Quality and Reliability Engineering International* 26: 887–897. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/qre.1158/full> (accessed September 2016).
- Forbes, S. and D. Brown. 2012. "Conceptual Thinking in National Statistics Offices." *Statistical Journal of the IAOS* 28: 89–98. Doi: <http://dx.doi.org/10.3233/SJI-2012-0751>.
- Foresti, G., F. Guelpa, and S. Trenti. 2012. "Enterprise in a Globalised Context and Public and Private Statistical Setups." *Proceedings of the 46th Scientific Meeting of the Italian Statistical Society*. (SIS), June 20–22, Rome. Available at: <http://meetings.sis-statistica.org/index.php/sm/sm2012/paper/viewFile/2393/267> (accessed 22 September, 2016)
- Giovanini, E. 2008. *Understanding Economic Statistics*. Paris: OECD Publishing. Available at: <https://www.oecd.org/std/41746710.pdf> (accessed 22 October, 2016).
- Hambleton, R.K. 2002. "How Can We Make NAEP and State Test Score Reporting Scales and Reports More Understandable?" In *Assessment in educational reform*, edited by R.W. Lissitz and W.D. Schafer. 192–205. Boston, MA: Allyn & Bacon.
- ISO/TS 10004. 2010 (E). *Quality Management — Customer Satisfaction — Guidelines for Monitoring and Measuring*, First Edition, the International Organization for Standardization, Geneva, Switzerland.
- Kenett, R.S. 2008. "From Data to Information to Knowledge." *Six Sigma Forum Magazine*: 32–33.
- Kenett, R.S. and Y. Raanan. 2010. *Operational Risk Management: A Practical Approach to Intelligent Data Analysis*. Chichester, UK: John Wiley and Sons.
- Kenett, R.S. and S. Salini. 2012. *Modern Analysis of Customer Satisfaction Surveys: With Applications Using R*. Chichester, UK: John Wiley and Sons.
- Kenett, R.S. and G. Shmueli. 2014. "On Information Quality (with discussion)." *Journal of the Royal Statistical Society, Series A* 177: 3–38.
- Kenett, R.S. and G. Shmueli. 2015. "Clarifying the Terminology that Describes Scientific Reproducibility (correspondence)." *Nature Methods* 12: 699. Doi: <http://dx.doi.org/10.1038/nmeth.3489>.
- Kenett, R.S. and G. Shmueli. 2016. *Information Quality: The Potential of Data and Analytics to Generate Knowledge*. Chichester, UK: John Wiley and Sons.
- Kenett, R.S. and S. Zacks. 2014. *Modern Industrial Statistics: With Applications in R, MINITAB and JMP*. Chichester, UK: John Wiley and Sons.
- McNutt, M. 2014. "Reproducibility." *Science* 343: 229. Doi: <http://dx.doi.org/10.1126/science.1250475>.
- Office for National Statistics 2007. *Guidelines for measuring statistical quality*. London: Office for National Statistics.

- Penny, R.N. and M. Reale. 2004. "Using Graphical Modelling in Official Statistics." *Quaderni di Statistica* 6: 31–48. Available at: <http://www.labstat.it/home/wp-content/uploads/2015/03/reale.pdf> (accessed September 2016).
- Rasch, G. 1977. "On Specific Objectivity: An Attempt at Formalizing the Request for Generality and Validity of Scientific Statements." *The Danish Yearbook of Philosophy* 14: 58–93.
- Statistics Canada 2002. *Statistics Canada's quality assurance framework*. Ottawa: Statistics Canada.
- UK Department of Health 2004. *A strategy for NHS information quality assurance—consultation draft*. London: Department of Health.
- Vicard, P. and M. Scanu. 2012. "Applications of Bayesian Networks in Official Statistics." In *Advanced Statistical Methods for the Analysis of Large Data-Sets*, edited by A. di Ciaccio, M. Coli, and J.M. Angulo Ibanez. 113–123. Berlin: Springer Verlag.

Received September 2014

Revised September 2015

Accepted August 2016