



Worldwide leaders in public and management accounting

Criteria for describing a set of data and evaluating its integrity

Contents

Notice to Readers

Criteria for Describing a Set of Data and Evaluating Its Integrity

- 3 Introduction
 - 5 Preparing a Description
 - 6 Presentation of a Set of Data
 - 6 Criteria
 - 11 Effective Date
-

Glossary

Notice to Readers

The 2020 *Criteria for Describing a Set of Data and Evaluating Its Integrity* were established by the Assurance Services Executive Committee (ASEC) of the AICPA for use when defining a set of data, documenting that definition in a description to help users understand a set of data and its individual members, and for evaluating the integrity of a set of data relative to its description.

ASEC, in establishing and developing these criteria, followed due process procedures, including exposure of the criteria for public comment. BL section 360R, *Implementing Resolutions Under Section 3.6 Committees*,¹ designates ASEC as a senior technical committee with the authority to make public statements without clearance from the AICPA Council or the Board of Directors. Paragraph .A44 of AT-C section 105, *Concepts Common to All Attestation Engagements*,² indicates that criteria promulgated by a body designated by the Council of the AICPA under the AICPA Code of Professional Conduct are, by definition, considered to be suitable. Accordingly, these criteria are suitable criteria for preparing and evaluating the presentation of the description of a set of data.

In accordance with AICPA Statements on Standards for Attestation Engagements, criteria to be applied in the preparation and evaluation of the subject matter should be suitable.³ Attributes of *suitable criteria*⁴ are as follows:

- a. **Relevance.** Criteria are relevant to the subject matter.
- b. **Objectivity.** Criteria are free from bias.
- c. **Measurability.** Criteria permit reasonably consistent measurements, qualitative or quantitative, of subject matter.
- d. **Completeness.** Criteria are complete when subject matter prepared in accordance with them does not omit relevant factors that could reasonably be expected to affect decisions of the intended users made on the basis of that subject matter.

In addition to being suitable, criteria used in an attestation engagement should be available⁵ to intended users. The publication of this criteria makes it available to users.

It is important to understand the distinction between standards and criteria. Professional standards such as the attestation standards establish the requirements and application guidance for CPAs engaged to perform and report during attestation engagements. In contrast, criteria are the benchmarks used to measure or evaluate the subject matter addressed by an attestation engagement. The engaging party may specify any criteria for use in the preparation and evaluation of the subject matter, provided that they are suitable, in accordance with the attestation standards. Accordingly, there is no requirement that a CPA use this set of criteria when measuring or evaluating a set of data.

¹ All BL sections can be found in AICPA Professional Standards.

² All AT-C sections can be found in AICPA Professional Standards.

³ Paragraph .25b of AT-C section 105, Concepts Common to All Attestation Engagements

⁴ Paragraph .A42 of AT-C section 105

⁵ Paragraph .25b of AT-C section 105

Criteria for Describing a Set of Data and Evaluating Its Integrity

Introduction

1. *Data* is typically defined as facts and statistics collected about the characteristics or attributes of events and instances for reference or analysis or for use as a basis for further calculation, reasoning, discussion, and informing decision-making. A *set of data* is a defined collection of data regarding events that share common characteristics or relationships. For example, when an entity enters into a transaction to factor receivables without recourse, it may be required by the factoring company to provide a set of data regarding the receivables factored and the entities from which the amounts are due.
2. Collecting and transforming raw data into a set that is useful for decision-making usually involves
 - defining the data to be collected.
 - collecting relevant data based on the definition.
 - recording the data.
 - aggregating, stratifying, processing, or grouping the data for use.

For example, data on the daily high and low temperatures for a community is determined by defining the individual thermometers from which the data is to be collected and the times and units of measurement (for example, Fahrenheit) at which the temperature will be measured. Air temperature is continuously measured by a thermometer (data), but it may only be measured, read, and recorded once per minute, creating a set of data of the recorded temperature readings of that particular thermometer for that moment or a range of time-based observation points. The data can then be combined with data from other thermometers, which vary in terms of location in the community or type of thermometer, to determine the daily high and low temperatures. Gas and electric companies can use this data to project consumption.

3. Data and sets of data have varying degrees of structure, which needs to be captured as part of the definition of the set of data. They may be highly structured (for example, phone numbers), partly structured (for example, email and object-oriented databases), or unstructured (for example, a series of characters in a sentence or pixels used to create a television image). However, even highly structured data may have unstructured elements (for example, the description field in a journal entry), and data that appears to be unstructured may have structured elements (for example, the data stored in a computer file appears to be random characters unless the user knows how the data was formatted and stored).
4. In order for a user of a set of data to have context regarding the data's integrity, a definition of that data should have the following attributes:
 - a. It is relevant to the events or instances that cause members to be included in the set of data.
 - b. It defines the relationship between the members of the set of data.
 - c. It permits determination about whether data about an event or instance is appropriately included in (that is, a data point is a member of the set of data) or excluded from the set of data.
 - d. It states the characteristics of the members to be measured and recorded (elements) and the validity, completeness, and accuracy of those measurements.
 - e. It permits the determination of whether the data is valid, complete, accurate, and current.

5. A set of data has integrity when it is consistent with the definition. The contents of a set of data may also be defined and described after the data has been created or obtained by performing procedures to evaluate the members of the set. For example, a forensic accountant may obtain a database of transactions and perform procedures to evaluate the nature of the transactions in the database and the information contained in each field.
6. A set of data that is collected and recorded, prepared, or otherwise obtained has a purpose, and the importance of the integrity of a set of data will vary depending on its purpose. For example, airline pilots have a greater need for accurate data regarding wind speed at an airport than does a person deciding whether to wear a jacket. Consequently, an airline pilot has a greater need for wind speed data to have integrity than a person making clothing decisions. Furthermore, the data the pilot relies on requires a greater degree of precision, completeness, and accuracy.
7. To reduce the risk that a set of data is misused or misinterpreted by users other than the collectors of the data, the description of the definition of the set of data needs to be available to users to allow them to understand the nature of data within the set, its purpose, and what determines membership in the set. This information also enables users to understand the limitations of the set of data.
8. This document presents criteria for use when preparing a description of the definition of a set of data, or evaluating whether a definition presents a set of data in accordance with the criteria, and provides implementation guidance that represents important characteristics of the criteria.
9. The criteria can be used to assist management, boards of directors, internal auditors, and other stakeholders in determining the relevance of the data to the users' purpose and in making decisions based on that set of data. Management may also use the criteria when defining the set of data and documenting that definition in a description to help users understand a set of data and its individual members.
10. The criteria may also be used by a CPA in an attestation engagement on the integrity of the set of data or on the description alone. An attestation engagement is performed in accordance with the AICPA Statements on Standards for Attestation Engagements. In such an examination or review engagement, the CPA uses the criteria when evaluating the description and, when applicable, whether the set of data is consistent with its definition. The criteria may also be used by a CPA in an agreed-upon procedures engagement.
11. The attestation engagement is performed in accordance with AT-C section 105, *Concepts Common to All Attestation Engagements*,¹ and other applicable sections of the attestation standards, depending on the type of attestation service being performed (for example, AT-C section 205, *Examination Engagements*). For example, a social networking website for professionals within an industry may wish to provide an anonymized set of data regarding network participants to a financial institution in order to obtain financing. In such a situation, an examination on the integrity of the set of data may be useful in obtaining more favorable terms.
12. The criteria in this document may also be used when the CPA is engaged to provide other nonattest or advisory services to a client. For instance, a CPA may use the criteria when engaged to assist management in defining a set of data, documenting management's description of that definition, and collecting, processing, and presenting the set of data for use in internal reporting. Consulting services are performed in accordance with the guidance in Statement on Standards for Consulting Services No. 1, *Consulting Services: Definitions and Standards* (CS section 100).²

¹ All AT-C sections can be found in AICPA *Professional Standards*.

² CS sections can be found in AICPA *Professional Standards*.

13. In addition, these criteria may be useful to responsible parties or organizations that are developing suitable criteria for a particular type of a set of data that has a specific definition and purpose. In such instances, these parties may find these criteria useful in establishing the specific definition for the type of a set of data.
14. These criteria are not intended for use in the context of engagements to report on financial statements in accordance with generally accepted auditing standards or Statements on Standards for Accounting and Review Services.

Preparing a Description

15. A description is intended to provide users of the set of data with context that will enable them to understand the data and make informed decisions based on that set of data.
16. Included in paragraph 21 are the criteria for use when defining a set of data and documenting the description of that definition. For the description to fairly present the set of data, the context and disclosures addressed by the criteria should be included in the description. In addition, the description is in accordance with the criteria when each of the relevant elements required by the criteria is included (that is, it is sufficient to avoid the description from being misleading due to material omissions or misrepresentations).
17. For some elements or sets of data, there may be a specific definition (for example, a definition specified by an industry group) that is relevant to understanding the data. Factors that may help the user understand the set of data include the following:
 - The population
 - The nature of each element
 - The source of the data
 - Units of measurement, and the accuracy, correctness, or precision of measurements
 - Uncertainty or confidence in the elements
 - The date of measurement or period of occurrence

In such situations, the description of the set of data may need to include the identification of that definition. In addition, when the characteristics of an element of the set of data differ from a widely used or expected definition, it may be useful to users to specifically state that fact to reduce the risk of misunderstanding.

18. Implementation guidance accompanies each criterion in paragraph 21. This guidance presents factors to consider when making judgments about the nature and extent of disclosures required by each criterion, but it does not address all possible situations.
19. The description may be presented using various formats, such as narratives, tables, or graphics, or a combination thereof. The degree of detail to be included in the description generally is a matter of judgment unless there are contractual, legal, or regulatory requirements that specify the detail to be included in the description.

Presentation of a Set of Data

20. A description presents a set of data in accordance with the description criteria if
- the description is in accordance with the description criteria.
 - members of the set of data are appropriately included in the population of events or instances represented by the set of data.
 - events or instances that should be included in the population are included in the set of data.
 - the elements of members of a set of data are valid, complete, accurate, and current based on the description.

Criteria

21. The following table presents the criteria for defining and documenting a description of the definition of a set of data and evaluating that set of data. The implementation guidance in the right column presents factors to consider when applying the criteria but does not address all possible situations.

Criteria	Implementation Guidance
<p>DI1: The description of the set of data includes the purpose of the set of data.</p>	<p>A set of data is usually collected and recorded or prepared for a specific purpose. In some instances that purpose is formally articulated and is based on a specific intended use by a specific intended user. However, in other situations, a set of data may be collected for multiple or dissimilar purposes, for example, the recording of messages issued to the CPU of a computer, which may be useful if an error occurs in the execution of an application. The act of deciding to collect and record or otherwise prepare a set of data results in the need to decide what data to include in a set.</p> <p>Determining the data to include in a set of data requires many decisions related to its purpose. The persons making the determination need to identify the purpose and may need to identify intended users and how those persons are intended to use the set of data. Failure to identify the purpose can result in the set of data being used for a purpose for which the set of data is not relevant or not valid, complete, accurate, or current.</p> <p>Similarly, users of a set of data need to determine whether their planned use of the set of data aligns with its purpose.</p> <p>Examples of the purpose of a set of data include the following:</p> <ol style="list-style-type: none"> U.S. Census data. One purpose of the set of data from the U.S. Census conducted every 10 years is to support the U.S. Congress in determining the apportionment of the House of Representatives. Automobile dealer inventories. A purpose of inventory data from an automobile dealership is to help the automobile manufacturer make decisions about production volumes and pricing and marketing strategies.

Criteria	Implementation Guidance
<p>D12: The description of the set of data is complete and accurate and includes the following:</p> <ol style="list-style-type: none"> The population of events or instances included in the set of data The nature of each element (field) of the set of data (that is, the event or instance to which the data element relates) The sources of the data within the set The units of measurement of data elements The accuracy, correctness, or precision of measurement The uncertainty or confidence interval inherent in each data element and in the population of those elements The time periods over which the set of data was measured or the period of time during which the events the data relates to occurred The factors in addition to date or period of time that determined the inclusion or exclusion of items in the data elements and population 	<p>For the set of data to be useful, the description of the set of data provided must be valid, complete, accurate, and current. Misrepresentations or failure to include information likely to be relevant to the decisions of intended users may result in a lack of integrity of the set of data as well as of the description of the set of data, which in turn, may cause users of the set of data to make erroneous or insufficiently informed decisions.</p> <p>Whether creating or using a set of data, it is necessary to understand what is and is not included in the set, as well as what elements of the data are included (for example, a field within a record). This information is provided in a definition of the data contained in the set. An inappropriate definition may result in errors, either due to a single error or to the aggregation of errors in the set of data, which may include the following:</p> <ul style="list-style-type: none"> Incomplete identification of events or instances Erroneously included events or instances Inaccuracy in the measurement or recording of data elements Data that is not fit for its purpose <p>The definition may accompany the set of data or be available separately. For simple sets of data, the definition may be implicit in the presentation of data. For example, the column headings of an on-screen television guide list channels, times, and programs. However, for most sets of data, a formal description, which may include the following, is necessary to use the set of data appropriately:</p> <ol style="list-style-type: none"> Identification of the Population The population consists of the events or instances included in the set of data. Users generally need to understand the factors that determined the inclusion or exclusion of an event or instance from the set, which are typically communicated in the identification of the population. For example, the date of sale as defined by generally accepted accounting principles (GAAP) determines whether a transaction is included in the sales journal for a particular period of time. For the 2010 U.S. census data, the data population includes the households that responded to the census form or that were interviewed by census takers, over a specified period of time. <p>There may be situations in which the set of data is known to be incomplete or has missing members of the population (for example, records may be missing for a particular date due to the failure of computer storage or an error in recovering data after a system disruption) or incomplete elements for members of the set of data. When the set of data is incomplete, the identification of the population should include information on the missing or incomplete members of the population to permit users to understand the nature of those members of the set of data and permit those users to evaluate whether the set of data is sufficient for their purposes.</p>

Criteria	Implementation Guidance
	<p>For example, a missing temperature reading from a thermometer measuring hourly air temperature may not be material for a set of data used to determine the average daily low temperature for the year. However, the loss of wind speed measurements during a hurricane may significantly affect the usefulness of a set of data used for establishing building codes in a hurricane-prone area.</p> <p>b. Nature of the Elements</p> <p>Each member of a set of data comprises elements that are the characteristics of the data that relate to a particular event or instance that has been recorded. In a traditional database, a member of a set of data is usually represented by a record, and the elements are recorded in the fields for that record. Each recorded characteristic is an element. Types of characteristics are measured consistently between events or instances, and the shared nature of the characteristics is described and associated with the elements in which they are recorded. Whether recording characteristics or using the recorded data, the definition of each element needs to be understood in order to reduce the risk of misstatement of the characteristic.</p> <p>For example, “attendance at a ball game” might mean tickets sold, persons passing through the turnstiles, or all persons at the venue who were not playing the game or otherwise employed in game-related activities or serving game patrons. The attendance reported will vary depending on the nature of the different definitions of <i>attendance</i>.</p> <p>For example, “size of a warehouse” might represent the number of square feet in the warehouse floor that houses products for resale, the number of a certain type of inventory item that may be stored in the space, or the number of divisible storage areas within the warehouse. As a result, the size of the warehouse will usually be reported in the manner that aligns with the purpose of the set of data.</p> <p>c. Sources of the Data</p> <p>Understanding the source of the data within the set is necessary in order to reduce the risk of misinterpreting the set of data during its use. For example, oral and tympanic measures of body temperature yield different results; therefore, a researcher using such a set of data needs to understand its source. In addition, understanding the source of the data may affect a user’s evaluation of the appropriateness of the set of data for the user’s purpose. For example, sets of data from independent sources, such as commercial data providers, may be less biased or more reliable than sets of data produced by the entity in some circumstances.</p>

Criteria	Implementation Guidance
	<p>Sources of data should be identified at a level of specificity that would permit the user of the set of data to obtain the same or a similar set of data given the appropriate circumstances or a user to understand the characteristics of the source of the data.</p> <p>In addition to understanding the source of the data, in some instances, it may be useful to provide information on how the set of data was collected or transformed and other information that can help the user evaluate the sufficiency of the completeness of the set of data.</p> <p>d. Units of Measurement</p> <p>Measurement of most elements requires the use of a unit of measurement. Because many elements have alternative units of measurement (for example, length measured in meters or feet, volume measured in liters or gallons, or area measured in square feet or square yards), identification of the unit of measurement for each element is necessary in order to avoid misunderstanding and potentially erroneous results. For example, in 1999, the Mars Climate Orbiter failed when one piece of software provided data in a different unit of measurement than was expected by the software that used the data.</p> <p>Unless obvious from the nature of the element, the unit of measurement should be identified.</p> <p>e. Accuracy, Correctness, or Precision of Measurement</p> <p>The measurement of many characteristics has a limit to its accuracy, correctness, or precision. For example, a thermometer that measures with a precision of two degrees may be sufficient for cooking but not for controlling a chemical reaction. In order to collect data accurately, the collector of the data needs to understand the level of precision set forth in the definition of the set of data, and in order to use the set of data appropriately, the user may need to understand the level of precision used in the collection of the data.</p> <p>For example, monetary amounts that are measured at a precision of the nearest \$1,000 are likely to be sufficient when estimating the balance of an account for analytical purposes; however, data used to prepare an account reconciliation is likely to be precise to the nearest cent.</p> <p>f. Uncertainty in the Data and Population</p> <p>Many types of data have characteristics that are uncertain at the time of measurement and will only be determined at a later time. Until the uncertainty is resolved, an estimate may be recorded based on the underlying characteristics. For example, a meteorologist may forecast rain at a certain location on a certain date. However, providing the measure of uncertainty (for example, the chance of rain is 80 percent based on the historical data for the conditions measured) allows a user to make a more informed decision.</p>

Criteria	Implementation Guidance
	<p>There is no prescribed method for communicating uncertainty. Examples of such methods include the following:</p> <ul style="list-style-type: none"> i. Providing the standard deviation of the element ii. Providing information on historical variations iii. Reporting the margin of error associated with polling data iv. Describing a range of possible values for the elements <p>In addition to describing the uncertainty, it may be useful to describe the persons, or the qualifications of the persons, determining the recorded value of an element that exhibits uncertainty. For example, it may be useful for the user to know that a financial estimate was prepared by the actuarial department of an insurance company.</p> <p>g. Date of Measurement or Period of Occurrence</p> <p>Identification of the measurement date of an element or population, or the period of time over which events occurred, is critical to both the measurement and use of data.</p> <p>h. Other Factors</p> <p>The characteristics of data vary depending on the nature of events or instances that form the subject population of the set of data. In addition, the characteristics of data may vary due to variation or changes in measurement. Consequently, the definition of the set of data may require inclusion of information regarding other factors in addition to those specified here. Examples of such characteristics include ownership, classification for security and privacy purposes, access privileges, version, retention and disposal requirements, lineage and audit trail information, and assurance-related information. They also may include changes in the consistency of measurement taken over a period of time. For example, the thermometer used to record hourly temperatures may be replaced by a newer model, resulting in a more precise reading and the elimination of a bias in the older thermometer.</p> <p>For some types of elements or sets of data, there may be specific definitions that are relevant in the determination of the population, the nature of each element, the source of the data, units of measurement, accuracy, correctness or precision of measurement, uncertainty or confidence in the elements, or other factors, and which are included in other suitable criteria. In such situations, the description of the set of data may need to include the identification of such criteria. For example, a set of data containing an extract of the general ledger of an entity would usually state that the determination of the existence and occurrence, completeness, accuracy, and valuation, and rights and obligations relating to the data provided, are based on GAAP. In addition, when the characteristics of the data differ from widely used criteria or an expected set of criteria, it may be useful to users to specifically state that fact.</p>

Criteria	Implementation Guidance
<p>D13: The description of the set of data identifies any information that has not been included within the set of data or description but is necessary to understand each data element and the population.</p>	<p>Whether creating or using a set of data, a significant amount of metadata is needed. The data description is a key source of metadata, but it is usually not practical to include all the information needed to use the set of data in the data description. For example, the information required to use a company’s financial statements includes knowledge of GAAP. When additional information beyond the data description is needed to use the set of data properly and the need for such information cannot be presumed to exist on the part of the user, the additional information should be identified. For example, in a set of data about crude oil inventories, the description may need to refer to the API (American Petroleum Institute) gravity definitions of each grade held. Similarly, if the description of the definition of census data did not include the definition of a household, the description would make reference to where that definition is available.</p>

Effective Date

22. The criteria are effective upon issuance.

Glossary

criteria. The benchmarks used to measure or evaluate the subject matter.

current. The subject matter is current if the present reality or condition of the subject matter is being represented. This is evaluated relative to the time period or cutoff date of the information relative to its purpose and the timing of its use.

data. Facts and statistics collected about the characteristics or attributes of events and instances for reference or analysis or for use as a basis for further calculation, reasoning, discussion, and informing data-based decision-making.

element (field). A particular shared characteristic of members of a set of data that has been measured and recorded in the set of data, for example, a field in a record.

member of a set of data, or member. Data regarding a particular event or instance that is included in a set of data, for example, a record of a transaction that has been recorded in a database.

metadata. A set of data that describes and gives further detail about other data. This includes the appropriate context required to understand the information.

record. A basic data structure comprising a collection of fields that share a relationship.

set of data. A defined collection of data regarding events or instances that share common characteristics or relationships.

valid. The data and the elements of the data represent what they purport to represent.



Worldwide leaders in public and management accounting

Copyright © 2020 by Association of International Certified Professional Accountants New York, NY 10036-8775 All rights reserved. CGMA and Chartered Global Management Accountant are trademarks of the Association of International Certified Professional Accountants and are registered in the United States and other countries. The Globe Design is a trademark owned by the Association of International Certified Professional Accountants. For information about the procedure for requesting permission to make copies of any part of this work, please email copyright-permissions@aicpa-cima.com with your request. Otherwise, requests should be written and mailed to Permissions Department, 220 Leigh Farm Road, Durham, NC 27707-8110 USA.
1 2 3 4 5 6 7 8 9 0 BRAAS 1 9 8 1912-03214