

SEMI E151-0309

GUIDE FOR UNDERSTANDING DATA QUALITY

This standard was technically approved by the global Information & Control Committee. This edition was approved for publication by the global Audits and Reviews Subcommittee on January 2, 2009. It was available at www.semi.org in February 2009 and on CD-ROM in March 2009.

1 Purpose

1.1 The purpose of this guide is to provide common terminology and definitions that can be used when addressing the quality of semiconductor equipment generated data. It establishes a set of common definitions that help specify, communicate, address, and discuss the quality of data produced by the production equipment.

2 Scope

2.1 This guide contains the following component:

2.1.1 Background and description of an approach to understanding data quality and associated terminology.

NOTICE: This standard does not purport to address safety issues, if any, associated with its use. It is the responsibility of the users of this standard to establish appropriate safety and health practices and determine the applicability of regulatory or other limitations prior to use.

3 Limitations

3.1 The term “data,” as utilized in this guide, is limited to data that is or can be collected from production equipment in the semiconductor industry via a SEMI standard communication interface.

3.2 This guide does not define the protocol of messaging used to transport or communicate data values. That messaging infrastructure is covered in other SEMI standards. The data quality definitions, attributes, and associated elements contained in this guide are intended to apply specifically to the value content of equipment messages that report data.

3.3 This guide does not specify how the quality of equipment generated and reported data may be improved or corrected. The quality of the messaging and communication infrastructure is outside the scope of this guide.

4 Referenced Standards and Documents

4.1 Definitions or descriptions of the following terms used in this specification may be found in the SEMI International Standards: Compilation of Terms (available on the SEMI Web site, <http://www.semi.org>).

4.2 SEMI Standards

SEMI E5 — SEMI Equipment Communications Standard 2 Message Content (SECS-II)

SEMI E54 — Sensor/Actuator Network Standard

SEMI E120 — Specification for the Common Equipment Model (CEM)

SEMI E125 — Specification for Equipment Self Description (EqSD)

SEMI E132 — Specification for Equipment Client Authentication and Authorization

SEMI E134 — Specification for Data Collection Management (DCM)

4.3 IEEE Standard¹

IEEE Std 100-1988 — Standard Dictionary of Electrical and Electronics Terms (4th Edition, ANSI/IEEE). Referenced definitions reprinted, with permission.

¹ Institute of Electrical and Electronics Engineers, IEEE Operations Center, 445 Hoes Lane, P.O. Box 1331, Piscataway, New Jersey 08855-1331, USA. Telephone: 732.981.0060; Fax: 732.981.1721; <http://www.ieee.org>

4.4 ISMI Documents

Factory and Equipment Clock Synchronization and Time-stamping Guidelines: Version 2.0; <http://ismi.semtech.org/docubase/abstracts/4781beng.htm>²

ISMI Data Quality Guidelines Version 1; <http://ismi.semtech.org/docubase/abstracts/4843ceng.htm>³

4.5 NIST Documents⁴

http://boulder.nist.gov/div818/stdterms/NIST_Defs.pdf

<http://tf.nist.gov/timefreq/general/glossary.htm>

4.6 Other References

McGraw-Hill Dictionary of Scientific and Technical Terms 5th Edition.⁵ Referenced definitions reprinted, with permission.

NOTICE: Unless otherwise indicated, all documents cited shall be the latest published versions.

5 Terminology

5.1 Terminology is defined as abbreviations, acronyms and definitions.

5.2 *Abbreviations and Acronyms* — The following abbreviations and acronyms are used in this guide.

5.2.1 *APC* — Advanced Process Control

5.2.2 *EDA* — Equipment Data Acquisition

5.2.3 *GEM* — Generic Equipment Model

5.2.4 *SECS-II* — SEMI Equipment Communications Standard 2 Message Content

5.2.5 *SML* — SECS Message Language

5.2.6 *XML* — Extensible Markup Language

5.3 Definitions

5.3.1 *absolute accuracy* — accuracy as measured from a reference that must be specified. [IEEE]

5.3.2 *accuracy* — closeness of the agreement between the result of a measurement and a true value of the measurand.

5.3.3 *application* — software used to record, analyze, or visualize equipment data which in turn will be used to control the equipment processes. Examples of these applications are: Advance Process Control, maintenance, cell controllers, database, and storage software systems.

5.3.4 *application latency* — the interval of time between the time data is received and the time data results are applied by the application.

5.3.5 *bias* — a fixed deviation from the true value that remains constant over replicated measurements within the statistical precision of the measurement. [IEEE]

5.3.6 *cycle* — (1) a periodic movement in a time series. [IEEE] (2) a period of time in which a set of events is completed.

5.3.7 *drift* — a deviation in the output from a set value that is affected by physical or environmental conditions.

2 International SEMATECH Manufacturing Initiative, 2706 Montopolis Drive, Austin, TX 78741, USA. Telephone: 512.356.3500; <http://www.ismi.semtech.org>

3 International SEMATECH Manufacturing Initiative, 2706 Montopolis Drive, Austin, TX 78741, USA. Telephone: 512.356.3500; <http://www.ismi.semtech.org>

4 National Institute of Standards and Technology, 100 Bureau Drive, Stop 3460, Gaithersburg, MD 20899-3460, USA. Telephone: 301.975.6478; <http://www.nist.gov>

5 McGraw-Hill Education, Two Penn Plaza, New York, NY 10121-2298, USA. Telephone: 212.904.2574, Fax: 212.904.6285; <http://www.mheducation.com>

5.3.8 *end-to-end latency* — with respect to the reporting of data from equipment to a requesting application, the total sum of tool data latency, network data latency, and the time it takes the signal at a receiving node to reach the application from which the latency is being measured.

5.3.9 *hysteresis* — phenomenon seen in the elastic and electromagnetic behavior of materials, in which a lag occurs between the application or removal of a force or field and its effect.

5.3.10 *max* — the maximum, i.e., the greatest value something assumes. [IEEE]

5.3.11 *min* — the minimum, i.e., the smallest value something assumes. [IEEE]

5.3.12 *network latency* — the amount of time taken by a signal or message to travel from the equipment to its destination through the factory network.

5.3.13 *offset* — a component of error that is constant and independent of the input, often used to denote bias. [IEEE]

5.3.14 *range* — (1) the capacity span of an instrument. (2) the difference between the highest and lowest value that a quantitative function may assume. [IEEE]

5.3.15 *repeatability* — closeness of the agreement between the results of successive measurements of the same measurand carried out under the same conditions of measurement. [NIST]

5.3.16 *reproducibility* — measurement with the same result under a set of conditions of measurement, that includes different locations, operators, measuring systems, and replicate measurements on the same or similar objects.

5.3.17 *resolution* — smallest difference between values of a measurand that can be meaningfully distinguished. [NIST]

5.3.18 *rules* — a set of instructions or behavior that results from an operation, request, or command.

5.3.19 *sampling rate* — defines the number of samples per second (or per other time unit) taken from a continuous signal to make a discrete signal.

5.3.20 *sensitivity* — (1) the ratio of the response or change induced in the output to a stimulus or change in the input [McGraw-Hill]. (2) the change in the response of a measuring device divided by the corresponding change in the stimulus. [NIST]

5.3.21 *stability* — the ability to maintain a value, remain unchanged, or stay within a set of limits or parameters.

5.3.22 *standard communication interface* — any SEMI standard communication port in the equipment used for the purpose of controlling, collection and reporting data. Examples of these interfaces are SECS-II, EDA, and Sensor Bus.

5.3.23 *tool latency* — the interval of time required for a signal to travel inside the equipment from its source and be available to an external user for retrieval. It is influenced by sampling time, reaction time, retrieval time, time constants, indirect measurement factors, processing, encoding, decoding, and packaging.

5.3.24 *transient response* — (1) a change in the steady-state condition of a parameter. (2) that part of a change in a variable, such as current, voltage, or speed which may be initiated by a change in steady-state conditions or an outside influence that decays or disappears following its appearance. [IEEE]

6 Data Quality Guide Graphical Constructs

6.1 This section describes a particular convention to help the reader of this guide understand the data quality approach used in this document and to aid in the interpretation of the identified aspects, attributes, and elements that impact data quality. This convention is utilized here strictly as an understanding aid; it does not impose any specification on data quality.

6.2 As shown in Figure 1, the oval graphical construct represents a data quality aspect. Data quality aspects are directly associated with data quality attributes which are represented as rectangular boxes. Data quality attributes which impact data quality are defined in the main body of this guide and are represented as rectangular boxes. Data quality elements are factors that influence data quality attributes. Data quality elements are represented by rectangles and are associated with a data quality attribute. Data quality associations are represented as a line connecting related