Data Quality Guidelines: Version 1

International SEMATECH Manufacturing Initiative
Technology Transfer #07024843C-ENG

Data Quality Guidelines: Version 1
Technology Transfer #07024843C-ENG
International SEMATECH Manufacturing Initiative
July 24, 2008

Abstract: These guidelines from the ISMI Data Quality project (MFGM031M) define the data, data quality, and data capability reporting required of semiconductor equipment and describes initial thoughts on an evaluation method to ensure that these capabilities and characteristics are provided correctly. It is intended to guide equipment suppliers, third-party application builders, and users of data towards measuring, producing, and verifying data quality in a comprehensive and consistent manner. This revision allows the document to be accessible to suppliers and other persons who can benefit from it. It is aligned with SEMI Data Quality Guide document 4613. See also Technology Transfer 08074943A-ENG, *Data Quality Evaluation Methods*.

Keywords: Data Quality, Automated Data Collection

Authors: Gino Crispieri

Approvals: Gino Crispieri, Author
Harvey Wohlwend, Project Manager
Olaf Rothe, Program Manager
Joe Draina, Director
Laurie Modrey, Technology Transfer Team Leader

Table of Contents

List of Figures

Acknowledgments

The author wishes to acknowledge the ISMI member company Equipment Data Quality Working Group (Group786) participation in developing this document.

The author of this guidelines document would like to also acknowledge the co-chairs of the SEMI Data Quality Task Force, James Moyne, John Pace, and Harvey Wohlwend for their work in this area that led to the publication of this guideline document.

1 EXECUTIVE SUMMARY

1.1 Background

Semiconductor processing has become more stringent due to market requirements for higher performance devices that require critical dimensions possible only with higher levels of data quality and process control. IC makers have been forced to drive equipment and third-party applications for even more detailed control of quality and efficiency for the processes and tasks they run. This has increased the importance of data quality for process analysis and control. Data quality is defined as data generated by the equipment that conforms to the requirements needed for reliable, repeatable, accurate, and successful processing.

As the industry moves towards e-Manufacturing, the data that semiconductor equipment generates become even more critical to improving equipment productivity and the quality of that data becomes of paramount importance since critical decisions depend on it. As more data become available from process equipment for software applications such as run-to-run control, fault detection and classification (FDC), statistical process control (SPC), e-Diagnostics, and others, data quality *must* be assured. It follows, therefore, that incomplete, inaccurate, or untimely data will lead to poor decisions impacting the factory's productivity. Data must be provided with sufficient accuracy, resolution, and sampling frequency to allow high fidelity extraction of relevant data features for process/equipment characterization, process control, fault detection, and failure diagnosis.

Event data and context information must be

- Complete
- Consistent
- Correct
- Reflect the actual time of and conditions pertaining to the occurrence of the event
- Transferred in a timely manner (needed to achieve fault interdiction)

1.2 Scope

This document defines data quality concepts, definitions, and guidelines to help equipment suppliers, third-party application builders, and users of data. It also introduces an evaluation method (see Technology Transfer #08074943A-ENG, *Data Quality Evaluation Methods*, for the full methodology) to ensure that these capabilities and characteristics are provided correctly. These guidelines move towards measuring, producing, and verifying data quality in a consistent manner. Data within the scope of this document include all data used for configuration, history collection, control, and diagnostics, i.e., control variables/signals, sensor values, events, and monitoring (equipment health) data available from the equipment through SECS/GEM, equipment data acquisition (EDA), or other defined interfaces (standardized method) for e-Manufacturing.

Data quality guidelines specified in this document are associated with the following:

- Data aspects, attributes, and elements
  - Availability
  - Interpretability

  - Usability
  - Interface
  - Mechanism
  - Protocol
  - Format
  - Units
  - Order
  - Accuracy
  - Precision
  - Frequency
  - Latency

Other terms include Min, Max, Bias, Drift, Repeatability, Resolution, Reproducibility, Transient Response, Sampling Rate, Tool Latency, Application latency, and Network Latency.

- Commands/Responses
  - Each request elicits a reply
  - Event reports are sent in the right order
  - Data reported are correct
  - Events are accurate
  - Messages are well formed
  - Alarms include context data

2 REFERENCE DOCUMENTS

Many past efforts to define and address data quality have produced requirements for control systems and data quality evaluation methods and standards. Some of these documents and standards are listed below.

2.1 SEMI Standards

The following standards are applicable to this document:

SEMI E5	*SEMI Equipment Communications Standard (SECS-II)*
SEMI E10	*Definition and Measurement of Equipment Reliability, Availability, and Maintainability*
SEMI E30	*Generic Equipment Model (GEM) for Communications and Control of SEMI Equipment*
SEMI E40	*Standard for Processing Management (PM)*
SEMI E54	*Sensor/Actuator Network Standard*
SEMI E79	*Definition and Measurement of Equipment Productivity*
SEMI E87	*Provisional Specification for Carrier Management System (CMS)*

SEMI E90	*Substrate Tracking System (STS)*
SEMI E94	*Control Job Management (CJM)*
SEMI E116	*Equipment Performance Tracking (EPT)*
SEMI E120	*Common Equipment Model (CEM)*
SEMI E125	*Specification for Equipment Self-Description (EqSD)*
SEMI E126	*Specification for Equipment Quality Information Parameters (EQIP)*
SEMI E133	*Provisional Specification for Automated Process Control System Interface (PCS)*
SEMI E134	*Specification for Data Collection Management (DCM)*
SEMI E148	*Specification for Time Synchronization and Definition of the Clock Object*
SEMI 4613	*Guide for Understanding Data Quality, Draft Document*

2.2 Other References

[1] *Control Systems Requirements Specification (CSRS) v2.0*, SEMATECH Technology Transfer 96123222B-ENG, Dec 1996.

[2] *Automatic Data Collection for Operational and Equipment Performance: TP2 Project Final Report*, SEMATECH Technology Transfer 01104187A-ENG, Oct 2001.

[3] *Factory and Equipment Clock Synchronization and Time-stamping Guidelines: Version 1.0*, SEMATECH Technology Transfer 06094781A-ENG, Sep 2006.

[4] *NIST Time and Frequency Glossary*.

[5] EEC Guidelines, Phase 2.5, Jul 2002.

[6] EEC High-level Requirements for APC.

[7] *Data Quality Guidelines and Test Methods v0.6*, SEMATECH Technology Transfer 02064280A-ENG, Jun 2002.

[8] *Automatic Data Collection (ADC) Requirements for Tracking Tool Performance*, SEMATECH Technology Transfer 03034388A-ENG, Mar 2003.

[9] SEMI Data Quality ballot, SEMI 3652A, Aug. 2004.

[10] *Semiconductor Factory and Equipment Clock Synchronization for e-Manufacturing*, SEMATECH Technology Transfer 04094557A-ENG, Sep 2004.

[11] “Time Synchronization,” *APC Symposium*, Sep 2005.

[12] *Using NTP: Introduction and Recommended Practices*, SEMATECH Technology Transfer 06024736A-ENG, Feb 2006.

[13] Leo L. Pipino, Yang W. Lee, and Richard Y. Wang, “Data Quality Assessment,” *Communications of the ACM*, Vol 45, No 4, Apr 2002.

[14] Elizabeth M. Pierce, “Assessing Data Quality with Control Matrices,” *Communications of the ACM*, Vol 47, No 2, Feb 2004.

[15] ISMI Data Quality Evaluation Method, SEMATECH Technology Transfer 08074943A-ENG, July 2008.

3 TERMINOLOGY

3.1 Terms Used

Absolute accuracy – Accuracy as measured from a reference that must be specified.

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.

Actuator – An analog or digital output device used to effect changes in the physical environment. Examples of actuators include mass flow controllers (MFCs) and open/closed valves.

Advanced Process Control (APC) – The manufacturing discipline for applying control strategies and/or employing analysis and computation mechanisms to recommend optimized machine settings and detect faults and determine their cause (SEMI E133).

Carrier – The cassette and/or box, standard mechanical interface (SMIF) pod, or front opening unified pod (FOUP) used to transport wafers through the factory.

Clock – A device used to provide real-time date and time information.

Data Quality – Data generated by the equipment that conforms to the requirements needed for reliable, repeatable, accurate, and successful processing. Data reported by the equipment need to be consistent, correct, and timely transferred.

Distribution – The outcome or result from the collection of values from a process, system, or population that describes its behavior or function. Distributions are used to build up notions of integrability and specifically to describe a generalized function

Equipment – Equipment (manufacturing equipment) performs one or more of the following manufacturing functions in the factory: material process, material transport, material measurement, or material storage. Equipment is made up of various parts: modules, subsystems, and sensors/actuators. Except for linked equipments, equipment has at least one carrier port. Equipment communicates with the factory control system.

Equipment Suppliers – All suppliers of equipment and devices used for manufacturing semiconductor devices.

Equipment Element – A component of the equipment that behaves as a unit, performs work, and may or may not contain lower level components.

Event Timestamp – Time and date the equipment or an application assigns to an event when it occurs or is reported.

Fault Detection – The technique of monitoring and analyzing variations in process data to predict anomalies (SEMI E133).

I/O Device – Any type of sensor or actuator or aggregation of sensor and/or actuator.

Interface – Any standard communication port in the equipment used for the purpose of controlling, collection, and reporting of data. Examples are SECS-II, EDA, and Sensor Bus.

Linked Equipment – Two or more equipment physically and logically connected and functioning as a single installation of equipment. In this case, the individual component equipment is modeled as a high level module of the linked equipment.

Lot Name – A unique designation that is used by the factory for identifying and tracking product-material through the factory.

Manufacturing Execution System (MES) – The factory system responsible for managing the manufacturing process, including logistics and process flow.

Material – 1) Any material used in or required by the manufacturing process. Material is classified as consumable, durable, or product. 2) An abstraction of the various types of things used during manufacturing, such as wafers, reticles, carriers, or durables that require some management.

Material Location – A reference to a place within the equipment of an equipment component that can hold material, such as the top surface of an indexer or substrate chuck or the end effecter of a substrate handler.

Measurement Equipment – Equipment whose intended function is to measure or inspect the product and to report results. Measurement of the product is the factory's means of gaining feedback on the manufacturing process.

Module (equipment module) – A major component of the equipment that contains at least one material location and executes only one task at a time.

Process equipment – Equipment whose intended function is to process material, i.e., to add value to the material.

Range — The capacity span of an instrument. The difference between the highest and lowest value that a quantitative function may assume

Recipe Name (process recipe name) – The name given to a particular set of user-defined instructions to process the wafer(s) within a module.

Rules — A set of instructions or behavior resulting from an operation, request, or command.

Run-to-Run Control – The technique of modifying recipe parameters or selecting control parameters between runs to improve processing performance. A "run" can be a batch, lot, or an individual wafer (SEMI E133).

Sensitivity – The ratio of the response or change induced in the output to a stimulus or change in the input. The change in the response of a measuring device divided by the corresponding change in the stimulus.

Sensor – A component that responds to changes in the physical environment and provides an analog or digital input value.

Sensor/Actuator Device – A device consisting of one or more sensors and/or actuators on the physical equipment. See SEMI E54 for a precise definition of "sensor or actuator" and for a description of the internal structure of a sensor/actuator network Common Device Model Definition.

Stability – Maintainenance of a value, an unchanged state, or performacne within a set of limits or parameters.

Setup – 1. (verb) Performing one or more steps that puts the equipment into a known state in which it is ready to perform a specific process; 2. (noun) The state of the equipment once it has been set up.

Subassembly – A component of equipment that provides some limited functionality.

Substrate – Same as *Material*.

Subsystem – An intelligent aggregate that behaves as a unit. A subsystem is made up of sensors and/or actuators and may contain mechanical assemblies. Multiple modules may share subsystems.

Transient Response – A change in the steady-state condition of voltage, current, or both. 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.

3.2 Acronyms and Abbreviations

ALID	Alarm Event Identifier
AMHS	Automated Material Handling System
APC	Advanced Process Control
CEID	Collection Event Identifier
EDA	Equipment Data Acquisition
GEM	Generic Equipment Model (E30)
MES	Manufacturing Execution Systems
OEE	Overall Equipment Efficiency (SEMI E79)
RAMS	Reliability Availability Maintainability Standard (SEMI E10)
SEMI	Semiconductor Equipment and Materials International
SECS-II	Semiconductor Equipment Communications Standard (SEMI E5)
SVID	System Variable Identifier
YMS	Yield Management System

4 MOTIVATION FOR IMPROVING DATA QUALITY

Today, data are collected using various interfaces provided by a hierarchy of devices and subsystems that make up equipment. The type and amount of information known by each component in this architecture becomes increasingly specialized the closer the component is to the devices that make up the equipment. At the equipment controller level, much information is known about the execution context within which the rest of the equipment is operating. This includes, for example, information about active jobs, currently executing recipes, and system configuration parameters. At the subsystem and I/O device levels, less information about the execution context of the equipment is known, but much more is known about the low level operation of each of the constituent devices. This would include data such as the direct current (DC) bias of a power supply or temperature controller value to maintain the desired temperature.

Because of the complexity of semiconductor equipment communication networks, subsystem controllers, I/O devices, and equipment control software, the transmission of information from lower levels to higher ones will likely be affected by task priorities and hence cause the data reported to be inconsistent or of poor quality. Figure 1 shows an example of levels of process control applied to manufacturing time scales.

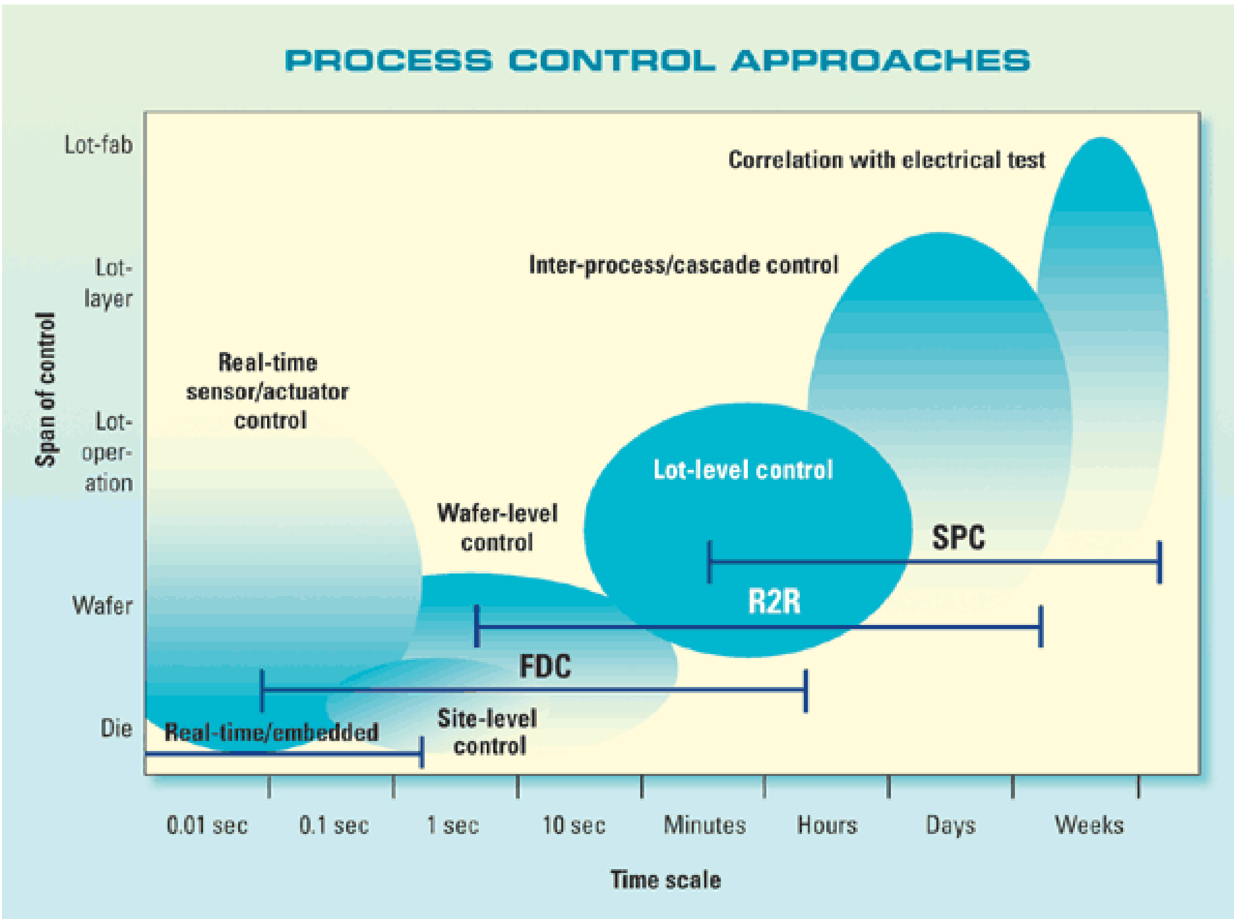

Figure 1 Levels of Process Control in Manufacturing

Several problems highlighted by many users, in particular International SEMATECH Manufacturing Initiative (ISMI) members, that cause poor data quality are discussed below.

4.1 Typical Data Quality Equipment Problems

Following are some of the top data and message quality problems reported by IC makers and identified in previous studies and projects [7]:

- Non-standard equipment interface behavior (deviation from the standards' definition)
- Documentation does not match software (undocumented features)
- Data collection has no defined limits
- Accessibility of event or data is not properly documented
- Data uses wrong format (float vs. text)

- Equipment processing affects data reporting and vice versa
- Short data collection periods with high report rates can take equipment down
- Missing data in requested data collection reports
- Faulty data (incorrect context data in reports)
- Message and data latency issues (data are not current)
- Time synchronization issues (timestamps)
- Inaccurate error reporting

4.1.1 Unavailable Messages and Variables

Often, one or more messages and/or variables required for standards compliance are not available as defined by the equipment manual or the standard. Typically, access to critical sensor data is not available through standard interfaces or other equipment interfaces at all. Making this type of data available usually requires negotiation between the IC maker and the equipment supplier and additional cost to the IC maker. Recipe parameters for equipment setup and process control are not available through the standard factory interface making it difficult to control processes remotely. Equipment documentation does not keep up with the software changes or accurately describe available events, alarms, or data. Non-conformance to SEMI standards like GEM or EDA limits equipment interface functionality in the factory.

4.1.2 Inconsistent Data Parameter Availability

Exposed equipment data parameters do not use correct formats, or their definitions are not consistent with accepted industry conventions for similar parameters on similar equipment. Differences in naming conventions and reported data types from equipment to equipment make it difficult to use the data for comparison and analysis. The data available inside the equipment sometimes are more accurate than the data reported. Sometimes, the resolution of the data is below what is needed for data analysis or control causing applications to make incorrect changes or wrong decisions. Data collection reports sometimes include or exclude data under undetermined conditions, i.e., data reporting is sometimes inconsistent for no apparent reason.

4.1.3 Duplicate Events

Identical event messages are issued with the same timestamp when only one message should have been issued. This causes the equipment controller and the factory applications to store duplicate information or malfunction.

4.1.4 Events Conflict with Equipment Process Flow

Equipment sends messages that violate the known logical sequence of events for the equipment process or the expected messaging flow to the factory control application. For example, if each wafer is supposed to undergo a 1) metal etch, 2) strip, and 3) cooling recipe, messages are sent out of order (1, 3, 2 or 2, 1, 3, etc.). In other cases, a completed event will be received before a start event. This problem is not a missing message but an out of order event report.

4.1.5 Missing Messages

Equipment sends messages on an intermittent basis. For example, a normally occurring “wafer complete” message is not received as expected after a “wafer start.” Messages are missed because of the interface reporting resolution or mechanism used to collect the event data is inconsistent, inefficient, unreliable, or poorly implemented.

4.1.6 Incorrect Message Received

Equipment sends the incorrect reply to a given request or the data included in the message are not correct. This problem causes the factory software to become out of synchronization and eventually stops the material process flow. Including the incorrect information in a message leads to incorrect decision making and may result in material being scrapped or damaged.

4.1.7 Missing Variables

Information is absent that is normally included in a given event report. Non-adherence to the standards used for communication and control causes data to be reported incorrectly, does not allow the user to specify certain parameters for control, or does not allow the parameters to be reported accurately. Data reporting synchronization inside the equipment is not done precisely leading to missing data in reports, creating uncertainty of the process data recorded. In general, this can affect traceability and knowing which wafer was processed or completed at the processing step.

4.1.8 Message Latency

Event messages are sent by the equipment long after a process problem has occurred, causing material to be accepted when it should be scrapped or not processed. Sometimes the latency will cause the context data to be reported with the wrong parameters or variables leading to incorrect process records or decisions. In some equipment, the event reports will stop being sent to the host at some process step and will not resume until a new trace report is re-initiated, resulting in missing data and incomplete traceability reports. Data in event or trace reports are not consistently reported causing jitter or missing data.

4.1.9 Data with Incorrect Timestamps

Timestamps from the equipment messages deviate from the factory “standard clock.” To obtain an accurate picture of the factory, the clocks for the equipment, the “middleware,” and the factory must all be synchronized. Timestamps reported from the equipment are either not accurate enough or they are not synchronized with the factory.

4.1.9.1 Conflicts in Timestamps

Data conflicts in timestamps occur in certain situations generating problems correlating data on other applications in the factory:

- Timestamps for duplicate messages – timestamps of redundant messages for the same event have conflicting times.
- Timestamps for different messages that represent the same process step conflict, e.g., the timestamp for the start of the first processing step conflicts with the timestamp for the corresponding module start message.
- Messages issued without a timestamp.
- Completion time (either from the event’s timestamp or from the message’s received timestamp) before the start time.

4.1.10 Communication Loss between Equipment and Control Entity

Equipment stops communicating with the factory without any clear reason. The factory cannot start correctly because the equipment's interface does not follow proper communication protocols when the equipment is introduced to the factory floor, resulting in long hours of troubleshooting and repair. Trace data collection plans can sometimes lock up the communication interface or the equipment itself, requiring a reboot of the equipment's control computer or, even worse, a restart of the whole equipment, resulting in both loss of production capacity and introduction of perturbations to equipment stability. This problem sometimes affects the user interface and other computers used internally in the supplier's equipment.

4.1.11 Incorrect Data Logs

User interface-displayed data and log records do not match factory communication logs. These records have messages missing, or the log data ares not accurate or complete.

4.1.12 Equipment Alarm Message Errors

The following basic alarm errors must be addressed:

- Severe alarm messages with severities 1, 2, 4, and 5 do *not* stop the equipment process as they should or non-severe alarm (warning) messages with severities 0, 3, 6, 7, and 8 stop the equipment process when they should not (SEMI E5-1105, section 7.9).
- Severe alarm messages are not issued in pairs (e.g., both set and clear messages are not issued).
- Alarm clear does not signify that the alarm condition no longer exists.
- Alarms are issued without a reason.

4.1.13 Equipment Provides Non-Standard Interfaces

Some equipment in the factory does not provide a standard interface to communicate with factory systems. Custom interfaces are not acceptable as they result in some areas in the factory not having supervisory software control and not being integrated with the factory or they require custom supervisory software development, debug, and maintenance.

5 DATA QUALITY CHARACTERISTICS, ATTRIBUTES, AND DEFINITIONS

SEMI Draft Document 4613, Guide for Understanding Data Quality, defines a standard terminology to understand and discuss data quality, its characteristics, and its attributes. This material is repeated here for convenience. The steps to conduct a data quality assessment are described. All three steps must be completed to evaluate the quality of data. Data quality starts with an interest in particular information to understand a process, analyze its results, or control it.

5.1 Data Availability, Interpretability and Usability Aspects

The primary focus in addressing data quality is the equipment data that are made available to the end user by transmission from the tool through standardized interfaces or the equipment-to-factory host or other factory-level data collection agent. An example is trace data as defined in SEMI E30 or the data available through the emerging EDA (Interface A) standard suite and methodology. While the quality of this transmitted data can be influenced by factors within the equipment as in the case of processing, the equipment data available to external users is the main concern here.

Three aspects of data quality have been identified—availability, interpretability, and usability as shown in Figure 2—that are used in this document to drive the data quality evaluation methodology.

Availability is the first step before beginning as assessment of data quality. Availability is achieved when the user has the means and privilege to get the data and has a degree of interest in a particular data set. *Interface* and *Mechanism* are two attributes of this data quality aspect as shown in Figure 3.

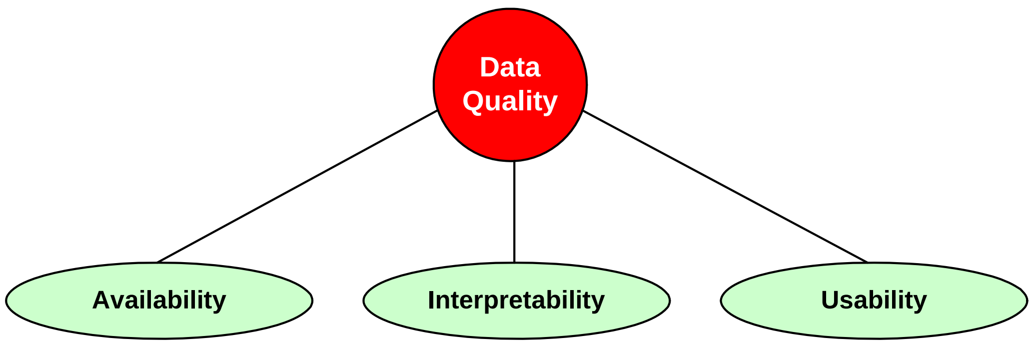

Figure 2 Data Quality Aspects

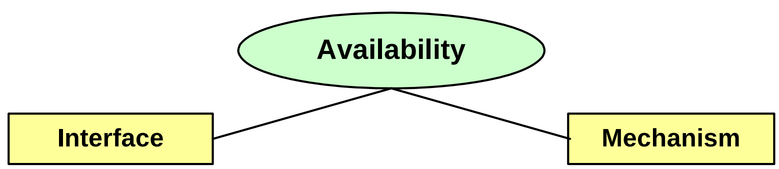

Figure 3 Data Accessibility Attributes

The following define the elements of the *Availability* of data quality:

Interface — Any standard communication port in the equipment used to control, collect, and report data. Examples are SECS-II and EDA.

Mechanism — In the context of this document, an event, equipment state, or special trigger used to collect or make data available to the user.

Interpretability —The means of understanding and deciding whether further analysis of the data will be meaningful and necessary. Interpretability is closely tied to the *protocol, format*, *order,* and *units* of the data of interest as shown in Figure 4.

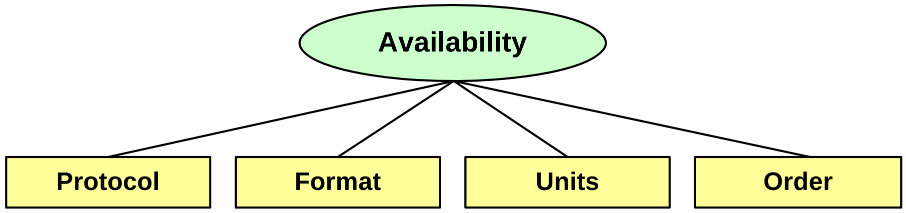

Figure 4 Data Interpretability Attributes

The following are definitions of the elements of *Interpretability*.

Protocol — The set of guidelines or rules for data representation, signaling, authentication, and error detection to send information over a communications interface. For example, syntax types based on SEMI standard protocols are XML and SML.

Format — The semantics of what something means or how it will be interpreted. There is a great difference in getting a parameter value in string, integer, or floating point.

Units — A standard measure of a quantity. Units help determine, express, or differentiate data. Units quantify the data. For example, PSI, Torr, and Bar are ISO standard units of measurement for pressure.

Order — Data that follow a specified pattern or sequence. This refers to whether the data request is an ordered array, an ordered sequence of values, or an ordered message request. Structure and/or content of data is important to its interpretation.

Usability is directly tied to the data specification. It is where actual data quality evaluation takes place and where many data attributes are verified. Figure 5 shows the attributes and elements related to the usability of data.

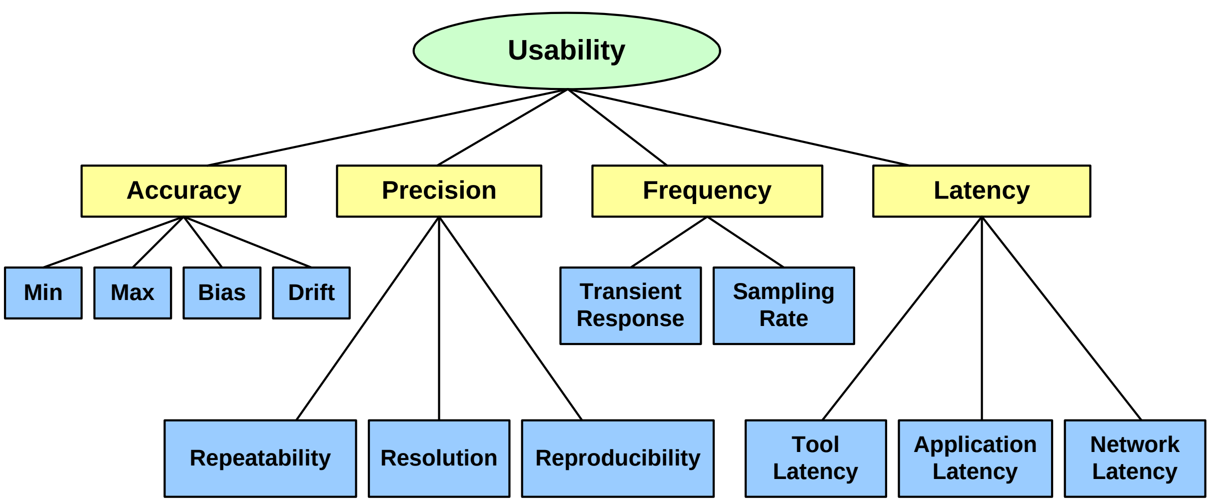

Figure 5 Data Usability Attributes and Elements

Accuracy, *Precision*, *Frequency*, and *Latency* are the usability attributes of data quality. Their definitions are as follows:

Accuracy — Closeness of the agreement between the result of a measurement and a true value of the measurand. Since the true value cannot be determined exactly, the measured or calculated value of highest available accuracy is typically taken to be the true value [3]. Figure 6 shows an example of high and low accuracy results represented as a collection of values within an intended target. Accuracy is directly affected by the *bias*, *min*, *max*, and *drift* inherent in the system or measurement. In many cases, the *max* or *min* permissible value is required to assess the accuracy, capacity, or capability of a system. Range is directly related to these two data quality elements. System *drift*, whether it is cyclic, random, or linear, can affect the accuracy of the desired process or measurement result. Unstable systems affect the ability to acquire accurate measurements or accurate data reporting.

Precision — The quality of the repeatability of measurement data, customarily expressed in terms of the standard deviation of the extended set of measurement results from a well defined (adequately specified) measurement process in a state of statistical control [4]. It is the closeness of agreement between indications obtained by replicate measurements on the same or similar objects under specified conditions. As shown in Figure 6, high precision represents a collection of data that is closely related or that has about the same value. Precision is measured by *repeatability*, *resolution*, and *reproducibility*. High data resolution does not imply precision for a system that is too sensitive, which in turn can produce excessive data noise. System hysteresis can affect the precision needed to achieve the same results over time.

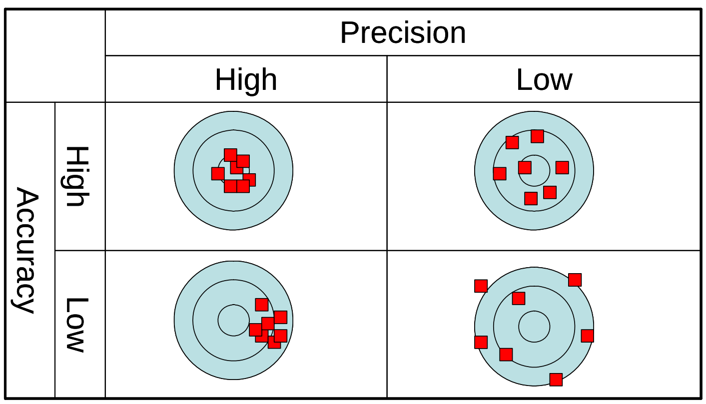

Figure 6 Accuracy vs. Precision

Frequency — Frequency is a measure of the number of occurrences of a repeating event per unit of time. It is also the time interval between samples in a periodic sampling control system. The *transient response* of a system can affect the rates at which data are collected. The s*ampling rate* is directly related to the data frequency such that a proper sampling plan can be used to collect the correct amount of data needed to determine the desired result or verify the output of a process within the required specification.

Latency — The time lag between sending a request and beginning to receive the data. Latency is an accumulation of multiple sources. In this document, three sources are identified: *tool latency*, *application latency*, and *network latency*. Tool latency is affected by all of the equipment's internal hardware and software responsible for retrieving the data and reporting it to the application that requested the information. Usually, *network latency* is negligible unless the message travels over several networks; then it becomes significant and acting upon the message or data result is critical. *Application latency* is not included in the scope of this document, but it is defined as the timestamp for the start or completion of the request from which *end-to-end latency* can be calculated.

Following are definitions of the elements associated with the *Usability*.

Application latency — The instance in time that a host or factory application sends or receives data.

Bias — A fixed deviation from the true value that remains constant over replicated measurements within the statistical precision of the measurement.

Drift — Slow change over time of a reported value.

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, plus the time it takes the signal at the receiving node to reach the application from which the latency is being measured.

Max — The maximum value of a function is the greatest value it assumes.

Min — The minimum value of a function is the lowest value it assumes.

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

Offset — A component error that is constant and independent of the input, often used to denote bias.

Order — Data that follow a specified pattern or sequence. Refers to the actual request for an ordered array, an ordered sequence of values, or an ordered message request. Structure and/or content of data is important to its interpretation.

Protocol — The set of guidelines or rules for data representation, signaling, authentication, and error detection required to send information over a communications interface. For example, syntax types based on SEMI standard protocols are XML and SML.

Repeatability — Closeness of the agreement between the results of successive measurements of the same measurand using the same measurement conditions.

Reproducibility— Measurement with the same result under a set of conditions of measurement, which includes different locations, operators, measuring systems, and replicate measurements on the same or similar objects.

Resolution — Smallest difference between values of a measurand that can be meaningfully distinguished.

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 includes sampling time, reaction time, retrieval time, time constants, indirect measurement factors, processing, encoding, decoding, and packaging.

Transient response — A change in the steady-state condition of voltage, current, or both. That part of a change in a variable, such as current, voltage, or speed, that may be initiated by a change in steady-state conditions or an outside influence that decays or disappears following its appearance.

6 EQUIPMENT DATA QUALITY GUIDELINES

The following guidelines identify what is required to improve data and to provide users with reliable data for interdiction, decision making, and factory-level and equipment-level process control. These guidelines indicate that data collected from the equipment has sufficient accuracy, resolution, and sampling frequency to allow high fidelity extraction of relevant data features for process/equipment characterization, process control, fault detection, and failure diagnosis. As part of these guidelines, event data and context information must be complete, consistent, and correct. They must also reflect the actual time of and conditions pertaining to the occurrence of the indicated event. Timely transfer of data is necessary for fault interdiction. Data streams provided by the equipment during reports must be complete, consistent, timely, accurate, and properly sampled. They must possess high enough resolution to serve their intended purposes. Data reported by the equipment that include event, alarms, and trace data must satisfy the following requirements.

6.1 Equipment Data Accessibility Documentation and Report

> The equipment must document and provide reports for all data and its associated parameters accessible from the equipment for configuration, data collection, and control. In some cases, process and data results must include information pertaining to their interpretability and usability attributes, ways to access that information, and any data collection limitations.

Who Implements: Equipment Suppliers

Who Uses: Equipment Suppliers, IC Makers, and Third-Party Applications

Remarks:

For accessibility, the messages to be sent to and received from the equipment depend on the interface used to communicate with the equipment. IC makers will evaluate potential data availability with respect to these guidelines. IC makers and suppliers will evaluate actual data availability by assessing against these guidelines.

In some cases, the supplier should document what data are not available and why. If data are used to control the equipment, then they should be made available external to the equipment. Data documentation must include the format, units, and mechanism to access the data contents. Data collection and analysis enable the understanding of data and the impact of recipe and control changes on the process. All data elements available from the equipment should be documented.

6.2 Equipment Data Collection and Reporting Reliability

> Data collection and reporting must not affect material processing and the reporting of data must not be affected by material processing.

Who Implements: Equipment Suppliers

Who Uses: Equipment Suppliers, IC Makers, and Third-Party Applications

Remarks:

The user setting up an application must not be required to determine whether sufficient equipment resources and capabilities are available for reliable data reporting.

Event transmission must meet pre-specified reliability requirements, such that event reliability includes the following:

- An event occurs that is reported in both a timely and accurate fashion
- Event reports are not transmitted for any events that did not occur
- Event or trace reporting must not affect processing

6.3 Data Transfer and Interpretability from Equipment

Data transfer and interpretability must not require a special software application. SEMI or another international standard protocol for data access and control must be used.

Who Implements: Equipment Suppliers, Third-Party Applications

Who Uses: Equipment Suppliers and IC Makers

Remarks:

All data must be transferable without customized software for specific equipment and/or applications.

Data structures must be interpretable by the equipment and receiver based on a standard data type or SEMI-defined data type.

Data communication must use either SEMI-based protocols or other standard protocol excluding any supplier proprietary communication protocol or format unless specified by the user.

6.4 Reported Data Accuracy and Resolution

The supplier must provide values for data accuracy and resolution for those parameters that are accessible and used for process and control.

Who Implements: Equipment Suppliers

Who Uses: Equipment Suppliers, Third-Party Applications, and IC Makers

Remarks:

Parametric data must have sufficient resolution and accuracy to distinguish signal noise from potentially subtle trends in the data. The typical order of magnitude of some process tolerances is in the single digits or less than 10%. Assuming 1% process sensitivity, a resolution one order of magnitude lower is required, or 0.1%, to have an acceptable signal-to-noise ratio.

- In some special cases, data must be provided with a resolution of not less that 0.1% of the full scale or 0.1% of the normal operating range, whichever is finer resolution.
  - For example, for a sensor that has a full scale operating range between 0 and 12 V, a resolution of at least 0.012 V is needed. If the normal operating range is between 9 and 11 V (a 2 V range), a resolution of 0.002 V is needed.
- If a parameter provides a resolution of 0.1%, it requires an absolute accuracy of 0.05% such that the least significant bit of the data has significant meaning.

6.5 Reported Data Sample Period and Throughput

Data sampling frequency must be just larger than the required signal bandwidth to enable perfect reconstruction of the signal from the sampled data. In other words, exact reconstruction of a continuous time baseband signal from its samples is possible if the signal has band limits and the sampling frequency is greater than twice the signal bandwidth (Nyquist Sample theorem).

Who Implements: Equipment Suppliers

Who Uses: Equipment Suppliers, Third-Party Applications, and IC Makers

Remarks:

Data must have a sufficient resolution across time to distinguish signal characteristics that occur at a high speed. In the near future, sample periods shorter than 100 ms are expected to be common for those parameters that are critical to achieving important production specifications. For such parameters, it is recommended that the minimum sample period be at 1% or less of the duration of a related recipe step.

- An example would be 100 ms for a 10-second recipe step.

Internal equipment sampling rate requirements are process-dependent. External equipment transmission rate requirements are application-dependent. This includes additional capabilities that would be required for e-Diagnostics or other use cases.

In the future, typical equipment will provide 50 variables per chamber at a maximum on-equipment sample rate of 10 Hz. For some specialized processes, such as rapid thermal and flash anneal, an additional smaller number of (up to, for example, a total of 30–40) critical variables may require an internal equipment sample rate of 200 Hz.

The actual data content and required number of simultaneous traces may vary depending on equipment type and/or process stations. For key process stations, such as chambers in an RF plasma etching system, it is highly recommended that each process station support its own independent trace of at least 50 parameters with maximum sample periods of 100 ms.

6.6 Data Report Latency and Buffering

Data must be made available in a timely manner to support process interdiction and control.

Who Implements: Equipment Suppliers

Who uses: Equipment Suppliers, Third-Party Applications, and IC Makers

Remarks:

In general, data must be transmitted as soon as possible to support real-time analyses and to timely interdiction where it is warranted.

The number of data samples contained in each independent off-equipment transmission must be constant and conform to on-equipment buffering parameters that are configured by the equipment user.

For discrete event data, the latency from the occurrence of a physical event to possible transmission from the data port of the equipment should be no longer than 100 ms. “Possible transmission” allows for the situation that either the network or the client application may be unavailable to receive the data. As a result of such outages, data may (but is not required to) be buffered at the equipment.

6.7 Data Sampling Period Uniformity and Acquisition Interval

Data must be reported with uniform sample rates and without a loss of data.

Who Implements: Equipment Suppliers

Who Uses: Equipment Suppliers, Third-Party Applications, and IC Makers

Remarks:

In the absence of such uniform sampling rate, additional pre-processing of data is required to apply many conventional time series analyses; generally this is an interpolation of the desired spacing of data points based on existing data points and is a burden to the IC maker. Ensuring that all requested data is included in the data report, avoiding missing data or duplicate values is paramount to enabling process control.

Figure 7 shows the nominal sample times for the first four samples of a trace. Jitter is conventionally measured from cycle-to-cycle and has more to do with precision.

- In the example, one instance of jitter may be calculated as $|(t_1 - t_0) - (t_2 - t_1)|$, where t_0, t_1, t_2, … t_n are successive actual sample times. Drift is the slope of a linear trend in the actual sample inter-arrival times. The drift can be positive when the reporting delayincreases or negative when the reporting is faster than the requested interval period.

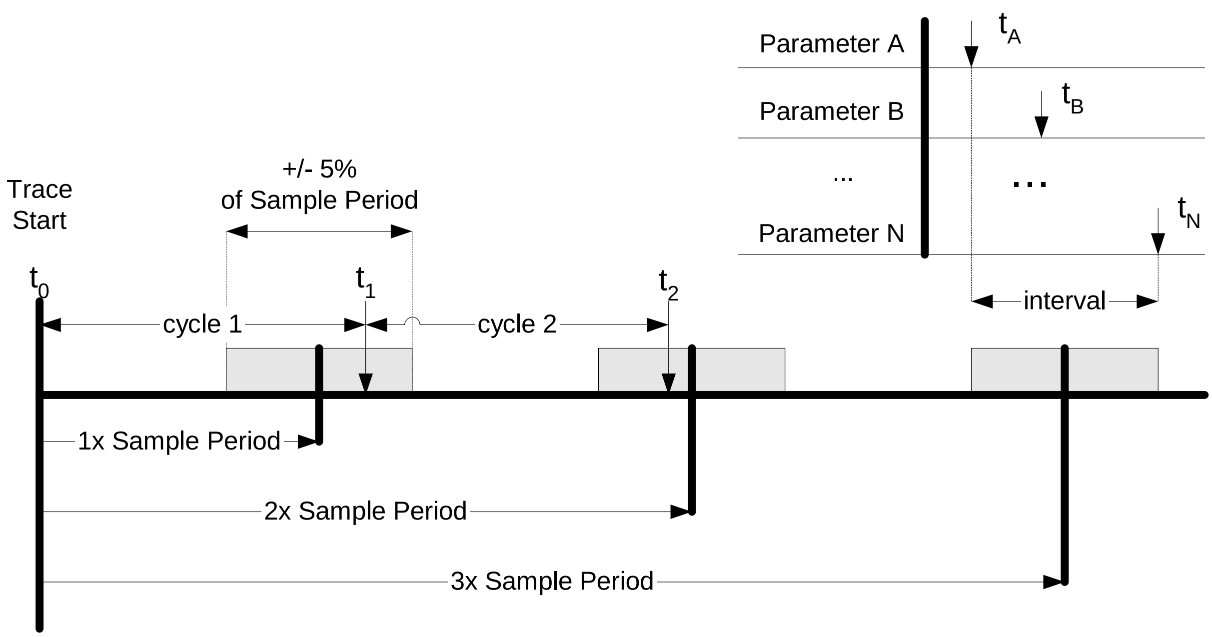

Figure 7 Sampling Uniformity of Trace Data

The overall acquisition interval for the requested sample is shown as the difference in acquisition times between the first and last parameter value acquired. To the extent that the sampling routine cycle is constant and predictable, a phase offset may be applied as an effective correction. For each sample, the time at which data is actually acquired must be within ±5% of the sample period of the nominal sample time such that no data loss is possible.

6.8 Equipment Reported Data Management

Equipment must efficiently manage its resources to produce, consume, and transmit timely and reliable data.

Who Implements: Equipment Suppliers

Who Uses: Equipment Suppliers, Third-Party Suppliers, and IC Makers

Remarks:

It is not practical for the equipment to always send all its available parameters at the highest sample rates possible.

To manage the equipment's limited processing and transmission resources, these resources must be managed in a manner appropriate for each equipment type and potential set of applications. Techniques and solutions for bandwidth management include dynamic data collection, customization, and compression.

For members of a vector whose values change less frequently than the sample rate of the vector itself, the values that have not changed may be backfilled with their previous values. Transient variables must be backfilled with a zero, null, or "no state" value, signifying when the variables are currently out of scope.

Equipment must manage its resources such that data collection reports are not affected by the processing of material, making sure the events are sent in the order they occur.

6.9 Data Timestamping and Resolution

Reported data from equipment must include accurate timestamps that are directly associated with the actual time the parameter is reported or the event occurred. These timestamps must use a factory synchronized time and format as defined in SEMI E5 and/or SEMI E148.

Who Implements: Equipment Suppliers, Third-Party Applications

Who Uses: Equipment Suppliers, Third-Party Applications, and IC Makers

Remarks:

Equipment suppliers must characterize the time domain characteristics of data streams generated during typical fault conditions. Arcing, MFC turn-on transients, and other dynamic characteristics of the equipment should be characterized and the sampling criteria met.

If one or more data variables are reported in a pre-defined vector or list, a timestamp must be included at the beginning of the list or such vectors and reported at intervals according to a pre-defined sample period.

Trace data collection for any trace begins with a start trigger event and continues until a finite number of samples have been gathered, a fixed amount of time has passed, and a stop trigger event has occurred.

6.10 Data Reporting Magnitude

Reported data must use standard engineering units (SI) and use standard data formats

Who Implements: Equipment Suppliers, Third-Party Applications

Who Uses: Equipment Suppliers, Third-Party Applications, and IC Makers

Remarks:

Within a trace, data should be reported in standard engineering units (e.g., V [voltage], cm, Pa [pressure], etc.) and conventional orders of magnitude (e.g., do not report the value “1026” to mean “10.26”).

7 DATA ACCESSIBILITY REQUIREMENTS

The typical data to allow factory applications to effectively use data generated by the equipment must be available on all equipment for efficient manufacturing operations. The required size and specifics of the data vary with the equipment category or relevant process and use. IC makers and third-party applications are users of the data that play an important role in its specification and application.

7.1 Data Related to Equipment Operation Management

The following are general data items that IC makers require for effective equipment management in a production fab environment. This information is critical to factory operation from several perspectives:

- Equipment availability
- Material scheduling
- Manufacturing decision making
- Process control
- Yield analysis

7.1.1 Equipment Operational Data

Equipment operational data is critical to daily factory operations. It is used to schedule material processing as well as to optimize equipment utilization according to factory requirements and the IC maker’s policies. The operational condition of the equipment, modules, loadports, ancillary facilities, etc., needs to be visible. Some (but not all) of the data types required to meet this need are as follows:

- Actuator operational data
- Calibration data, control limits, and sensor settings used to monitor the operation

- Recipe settings and process chamber status
- Valve opening/closing events and number of times operated
- Monitor and report information from the exhaust system, etc.
- Operation of electrodes and susceptor/stage system
- Operation of vacuum chuck (vacuum, static electricity, mechanical, etc.), etc.
- Operational condition of carrier system, loadports, etc.
- Wafer-level tracking and material status
- Process speed and cooling response rate, etc.
- Process results and job status

7.1.2 Process Data

Process data are perhaps the most important information that the factory needs to maintain product yield. These data ares specific to each equipment type. Careful attention must be given to make this data available for equipment setup, qualification, process control, and process monitoring. Some (but not all) process data types are as follows:

- Processing conditions for chamber and associated chamber components
- Recipe used and parameter settings
- Compensation values (variables/parameter)
- Equipment configuration settings
- Offset values
- Processing results (measurements) where applicable
- Wafer flow within equipment
- Context information

7.1.3 Alarm-Related Data

Alarm reporting is critical to the safety of the material, equipment, manufacturing control, users, factory, and the environment. Alarm reports must be accurate and follow standard procedures such that they can be cleared according to the factory policies and operations as well as to the standards that define their behavior. Alarms initiated by the equipment including the associated generation and release must be reported accurately and on time. The data items that must be included are as follows:

- Alarm ID
- Alarm type
- Alarm contents and description
- Set and clear data
- Time of occurrence

7.1.4 Data Related to Inspection/Maintenance

Part of the daily operations of the factory is to inspect and maintain the equipment for optimal performance. These operations include the visual inspection of parts and records of equipment daily operations. They also include the review of processing cycles and daily, monthly, and other scheduled maintenance policies and procedures. The inspection/service records of the equipment are used to schedule future maintenance and to identify potential causes of process inconsistencies or problems. Some (but not all) of the data types needed are as follows:

- Name and description of the actions performed on the equipment component inspected and/or serviced.
- Name of the inspected equipment component (part name [and description]).
- Frequency with which inspection/service is implemented (i.e., number of times of use, usage time, elapsed time, date, etc.).
- Average time required for the inspection/service.
- Name of the person performing the inspection/service.
- Number of equipment/chamber processing cycles.
- Equipment part and/or subsystem mechanical cycles.

7.1.5 Data Related to the Equipment Configuration

The management of equipment configuration is of vital importance because configuration changes can cause differences in process capability and outcomes. Providing access to configuration data therefore allows users, equipment maintainers, and factory systems to proactively monitor equipment configuration and hence control process outcomes.

Equipment configuration data may take several forms, including (but not limited to) the following:

- Paper logs
- Electronic configuration files
- Equipment constants that are settable/readable by the factory system

Equipment suppliers are normally required to provide all data related to equipment configuration when the equipment was shipped from the equipment supplier's factory, as well as the configuration data when it was accepted into the IC maker's factory. Some (but not all) of the configuration data types needed are as follows:

- Recipes
- Recipe parameters
- Recipe parameter units
- Calibration data
- Hardware models and version numbers

- Software models and version numbers
- Hardware and software settings

Particular attention should be paid to the software versions because these tend to change frequently as functional improvements are made during the life of the equipment.

7.1.6 Data Related to Parts

This information is needed to manage the equipment's physical parts, which include durable, consumable, and structural parts. This data should be categorized according to descriptions and requirements as defined in standards like SEMI E120. This description and information must be related to the equipment configuration data described in the previous section.

Required information is as follows:

- Part name and description
- Part number
- Manufacturer
- Usage limitations for the parts
- Usage condition of replacement parts
- Usage of consumables
- Information on the part version

7.1.7 Data Related to Security

To prevent in-plant information leaks and/or recipe and process information leaks during remote diagnosis, etc., a general security system must be provided that makes it safe to handle information related to intellectual property and patents. This requirement is directly related to protecting the user's intellectual property. The information required to establish user authentication before remote diagnostics or to access directories and applications in the tool is as follows:

- Account information
- User name
- Login ID
- Password
- Affiliation/post
- Access level
- Access history
- Login/node
- Connection time, etc.

8 EVALUATING DATA QUALITY

The following list includes items that will be addressed when developing methods and procedures to evaluate data quality. See also Technology Transfer #08074943A-ENG, *Data Quality Evaluation Methods*.

8.1 Equipment-Related Items

- Data must be available through the equipment's communication port(s).
- Collected data values must include sensors, I/O devices, actuators, process events, internal states, etc.
- The communication port will need to support high speed messaging. The minimum data throughput requirement is estimated at 100 data elements per second per process module.
- Data must be accurately time-stamped and the timestamp must be synchronized with the factory’s clock.
- Collected process data must be associated with product-material identifiers or context information (i.e., LotID, SubstrateID, recipe name, recipe step name, wafer orientation, etc.).
- A means to externally stop wafers from processing and to resume processing through remote commands is needed.
- Where applicable, a means to externally and automatically control the process (e.g., endpoint detection, fault detection).
- A means to update process recipe settings based on run-to-run, feed-forward, and feedback data is needed.
- The equipment’s communication port should support spooling during trace data collection.
- Additional sensors should be able to be added and integrated into the equipment.

8.2 Factory and Evaluation Tool Related Items

- The factory must supply equipment with means to time-synchronize with a master clock.
- Support is needed for a method to synchronize information collected from added external sensors with the equipment’s internal state.
- The data quality evaluation tool must be able to detect unavailable messages and variables.
- The data quality evaluation tool must be able to detect duplicate and missing events.
- The data quality evaluation tool must be able to detect data in conflict with the known equipment process flow.
- The data quality evaluation tool must be able to detect missing variables from message.
- The data quality evaluation tool must be able to detect data with incorrect event timestamps.
- The data quality evaluation tool must be able to detect timestamp conflicts.

- The data quality evaluation tool must be able to detect instances of reported completion time being earlier than start time.
- The data quality evaluation tool must be able to detect missing timestamps.
- The data quality evaluation tool must be able to detect alarm message errors.
- The data quality evaluation tool must be able to detect incorrect or missing SEMI standard state transitions.

9 CONCLUSION

These guidelines pave the road for developing the methodologies needed to address the data quality problems raised by ISMI's member companies. This document will continue to evolve as more knowledge is gained and additional guidelines are developed. Data that semiconductor equipment and applications generate are critical to improving equipment productivity.

Data quality is of paramount importance since critical decisions depend on it. As more data become available from process equipment for software applications such as run-to-run control, FDC, SPC, e-Diagnostics, and others data quality *must* be assured. Incomplete, inaccurate, or untimely data will lead to poor decisions impacting the factory's productivity. Data must be provided with sufficient accuracy, resolution, and sampling frequency to allow high fidelity extraction of relevant data features for process/equipment characterization, process control, fault detection, and failure diagnosis.

International SEMATECH Manufacturing Initiative
Technology Transfer
2706 Montopolis Drive
Austin, TX 78741

http://ismi.sematech.org
e-mail: info@sematech.org