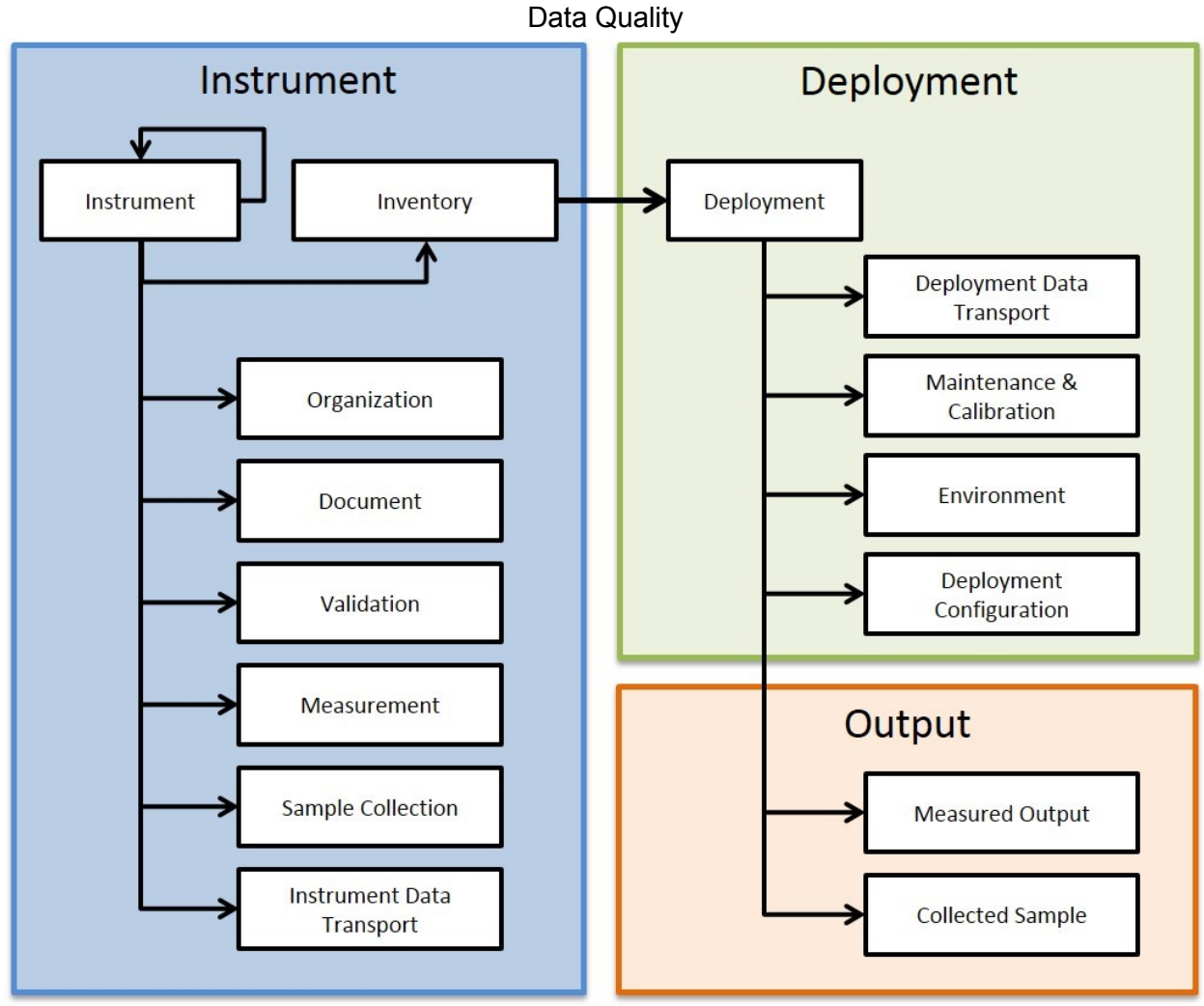
Sensor Common Metadata Specification

Version 1.0

October 2016



MIT License

Acknowledgement

This effort was supported by U54EB021973, National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health.

Table of Contents

List of Figures

List of Tables

1. Introduction

Sensors, especially personal and mobile sensors have provide methods for measuring environmental exposures of individuals and populations. But sensors use different methods and technologies for measuring different environmental species and output their measurements in different formats and specifications. In addition, sensors have differences in their performances and uncertainties associate with their measurements. It is there necessary to describe sensors in a generalized and sharable manner to support their proper use.

The Sensor Common Metadata Specification (SCMS) is designed to support the conduct of research utilizing personalized and environmental sensors. The scope of the specification ranges from nano-sensors up satellites. Sensor measurements may include physical, chemical, and biological species. In addition, sensors includes that instantaneously (or with a transient storage) measure these species or those that collect physical samples for later analysis. Sensors may be deployed in various environments, including personal (i.e. implanted & mobile), immediate (i.e. indoor), and general environment (i.e. external environmental protection agency monitors).

This business specification document may serve as a guide for sensor data modeling within any data management technology as required for your specific implementation. For example, you can use these specifications to develop relational, graphical or document stores of your sensor data. In similar line, we plan to develop a separate data modeling document for our data platform based on this specifications document.

A *sensor* in this document is used interchangeably with *instrument*. The reason being a sensor may be comprised of one or more observing instruments, where a hardware device (sensor or instrument) may contain other devices in a hierarchical manner. This terminology attempts to reduce the confusion that a sensor may contain other sensors.

1.1 Project Background

Understanding the effects of the modern environment on pediatric asthma requires generation of a complete picture of environmental exposures, clinical, biological and socio-behavioral factors. Such an exposome require integration of data from wearable and stationary sensors, environmental monitors, physiology, medication use and other clinical data.

This work is being undertaken as a part of the he Pediatric Research using Integrated Sensor Monitoring Systems (PRISMS) program (https://www.nibib.nih.gov/research-funding/prisms) for performing exposomic studies of pediatric asthma and other chronic diseases. These specifications will be used in the developed of informatics platform for data exposomic data collection, harmonization, semantic integration and provisioning of the data for different research study analyses and visualizations. These specifications will be used at the Utah Informatics platform to develop a logical data model to store and harmonize metadata from sensors and load it into OpenFurther's metadata repository to support metadata driven semantically consistent integration of all data.

1.2 Purpose of the Common Metadata Specification

The purpose of the SCMS is to:

1. Establish a library of instruments: Investigators can use this library to select appropriate instruments for different studies and acquire information necessary to contact the organizations owning or manufacturing these with instruments.
2. Describe and document deployments of sensors: Store sensor environmental and deployment attributes that are useful when using the measurements for analysis.
3. Assess quality of data collected by different instruments within its deployment environments: Use descriptive metadata to compare sensors and check if measurements are as expected.
4. Support harmonization and integration of data collected from various sensors
5. Guide for structuring and storing sensor output data

The scope of the specification is to support a diverse set of exposomic research questions and studies (*Table 1*) including observational, epidemiological and prospective studies (*Figure 1*).

Table 1: Research questions supported by these specifications.

1. *Mobile Instrument Models that can measure PM2.5.*
2. *Mobile Instrument Models that have been deployed to measure PM2.5.*
3. *Serial Number of all Instruments deployed supporting REST Data Transport Protocols and capturing output of PM2.5 (Or PM10, or Ozone).*
4. *Instrument Models that were manufactured by the AirMetrics.*
5. *Deployed Instruments owned University of Utah and currently measuring Ozone.*
6. *Organizations the collected personal exposures of PM2.5 in indoor and outdoor environments.*
7. *Calibration procedures used for MiniVol when deployed in an area with tall buildings.*
8. *Reference detection limit of MiniVol to assess quality of data quality in a study.*
9. *Number of sensors deployed by University of Utah in Salt Lake County that are less than 100 meters of I-15 in April 2016, and give the geolocation of each sensor and species measured by each sensor.*

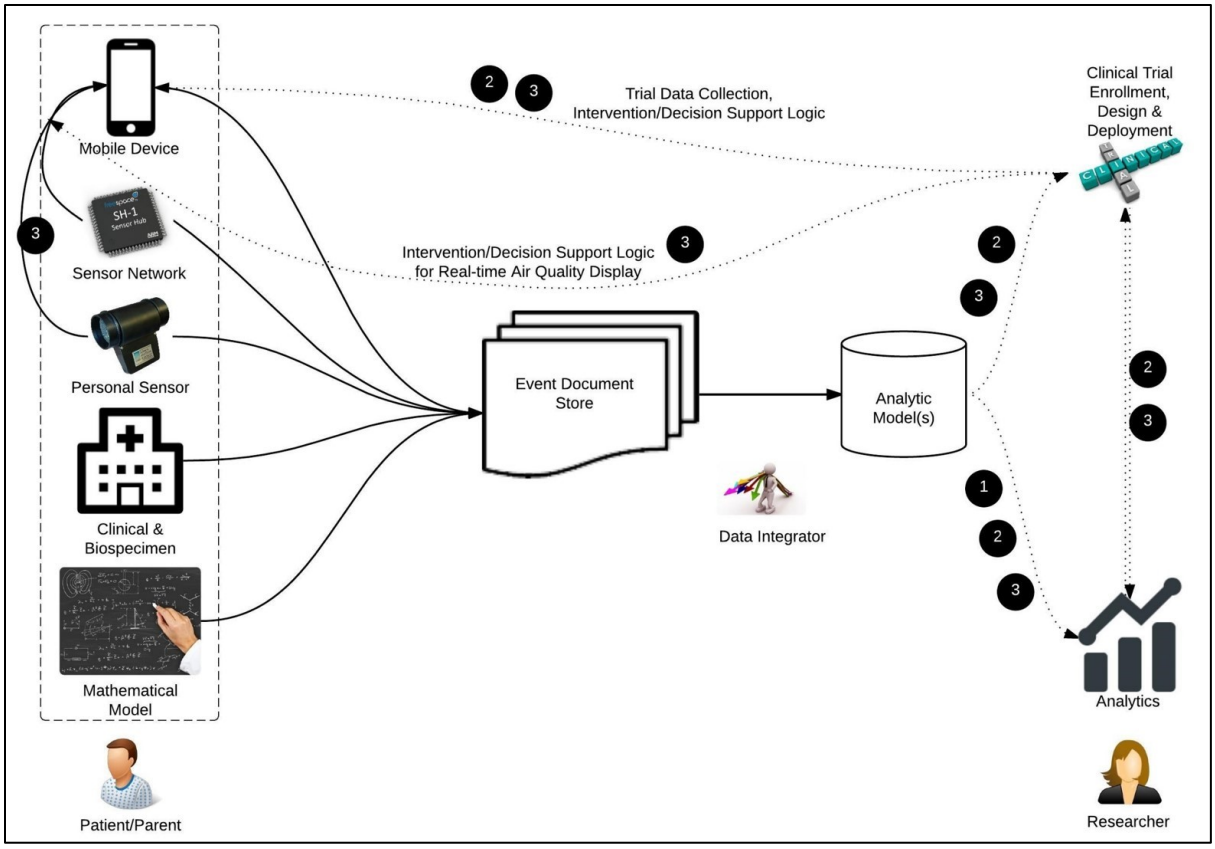


Figure 1: Diverse use-cases supported by the specifications.

1.3 Development Process

We are following a four step process in developing the Sensor Common Metadata Specification (SCMC).

Figure 2: Four step specifications development process

1. Literature Review: We performed a literature using PubMed with the search criterion “Pediatric Asthma Sensor Studies.” This returned 231 journal articles from August 1985 - December 2015, of which 40 full texts were read. Sensor types found in this literature corpus included stationary sensors (e.g. EPA), personal sensors (mobile), and indoor sensors. A list of metadata elements were manually extracted from this literature corpus, and the first conceptual model was established.
2. Preliminary mapping with sample data: To further establish the model, we did a preliminary mapping of sample data with the conceptual model. Sample data included data from the Environmental Protection Agency (EPA), Utah Department of Air Quality (UDAQ), MesoWest (http://synopticlabs.org/), West Valley Study (UDAQ), Asthma Triggers (Dr. Rima Habre), Wood Burning (Dr. Kerry Kelly), Purple Air (Mr. Adrian Dybwad,

http://www.purpleair.org/), Modeled Air Quality Data (1999 to 2007, 6 km grid, Dr. Jeffrey Yanosky), Measured Air Quality Data with Altitude (Dr. Geoff Silcox), and Hierarchical Bayesian Modeled Air Quality Data (EPA). Existing fields found in the data, but not present in the model, were added to the model.

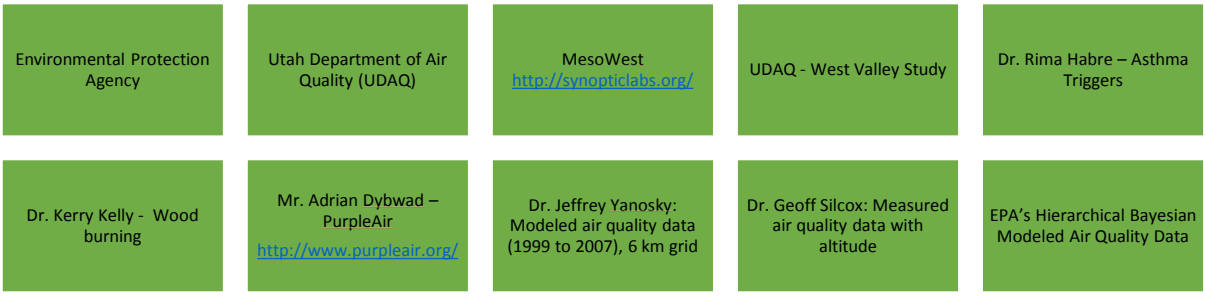


Figure 3: Sample data used for developing the specifications.

3. Utah Expert Review: We then reviewed the model with air quality experts in Utah Experts included: Dr. Kerry E. Kelly, Assistant Professor, Chemical Engineering, University of Utah; Dr. John D. Horel, Professor, Atmospheric Sciences, University of Utah; Dr. Scott C. Collingwood, Research Assistant Professor, Pediatrics, University of Utah; Mr. Adrian Dybwad, Purple Air; and Dr. Neal Patwari, Associate Professor, Electrical Engineering, University of Utah. We modified the model further based on their inputs
4. Community review of version 1.0: We share the SCMS with the PRISMS community and with help of assisted surveys review the model.

1.4 Glossary of Terms

Table 2: Glossary of Terms used in the specifications.

Term	Description
Calibration	Operation that, under specified conditions, in a first step, establishes a relation between the quantity values with measurement uncertainties provided by measurement standards and corresponding indications with associated measurement uncertainties (of the calibrated instrument or secondary standard) and, in a second step, uses this information to establish a relation for obtaining a measurement result from an indication. *(JCGM 200:2008 International vocabulary of metrology — Basic and general concepts and associated terms (VIM))*
Concept Identifier (ID)	An Identifier that uniquely identifies a Concept.
Concept	Represents a set or class of entities or things within a domain. Also called concept name.
Namespace	An abstract container to hold a logical grouping of unique concepts or identifiers.
Data	In this specifications data refers to all measurement values.

Deployment	The event that the physical instrument is utilized and brought to effective action.
Device	A composite set of one or more sensors each of which captures a specific measurement. Used interchangeably with instrument.
Graph Database	A database that uses graph structures of nodes, edges and properties to represent and store data.
Instrument	A composite set of one or more sensors each of which captures a specific measurement. Used interchangeably with instrument and monitor.
Metadata	Information that provides describes measured data from sensors, the deployment of sensors, and the sensors itself
Metadata Repository	A store for metadata that can be leveraged in computational platforms.
Relational Database	A digital database whose organization is based on the relational model of data
Sensor	A thing that is capable of making a specific observation or measurements of the real world and contains one or more instruments.
Species	An entity that is subject to measurement.
Validation	The process to test and evaluate that an instrument have the capacity to measure what it is supposed to measure.

1.5 Design Overview

The SCMS consists of three domains (*Figure 4*).

1. Instrument: The instrument domain contains data elements that describe a physical inventory of models, along with its documentation, data transport, validation tests, measurement features, and owning and manufacturing organizations. It can be used to maintain a library of instruments using which researchers can make informed selections of instruments for different research purposes.
2. Deployment: The deployment domain contains data elements that describe how a physical instrument is deployed in real world and includes characteristics such as the instruments deployment environment, setting, data transport, and calibration.
3. Output: The output domain contains data elements that describe the measurement of the sensors or the physically collected samples of different species.

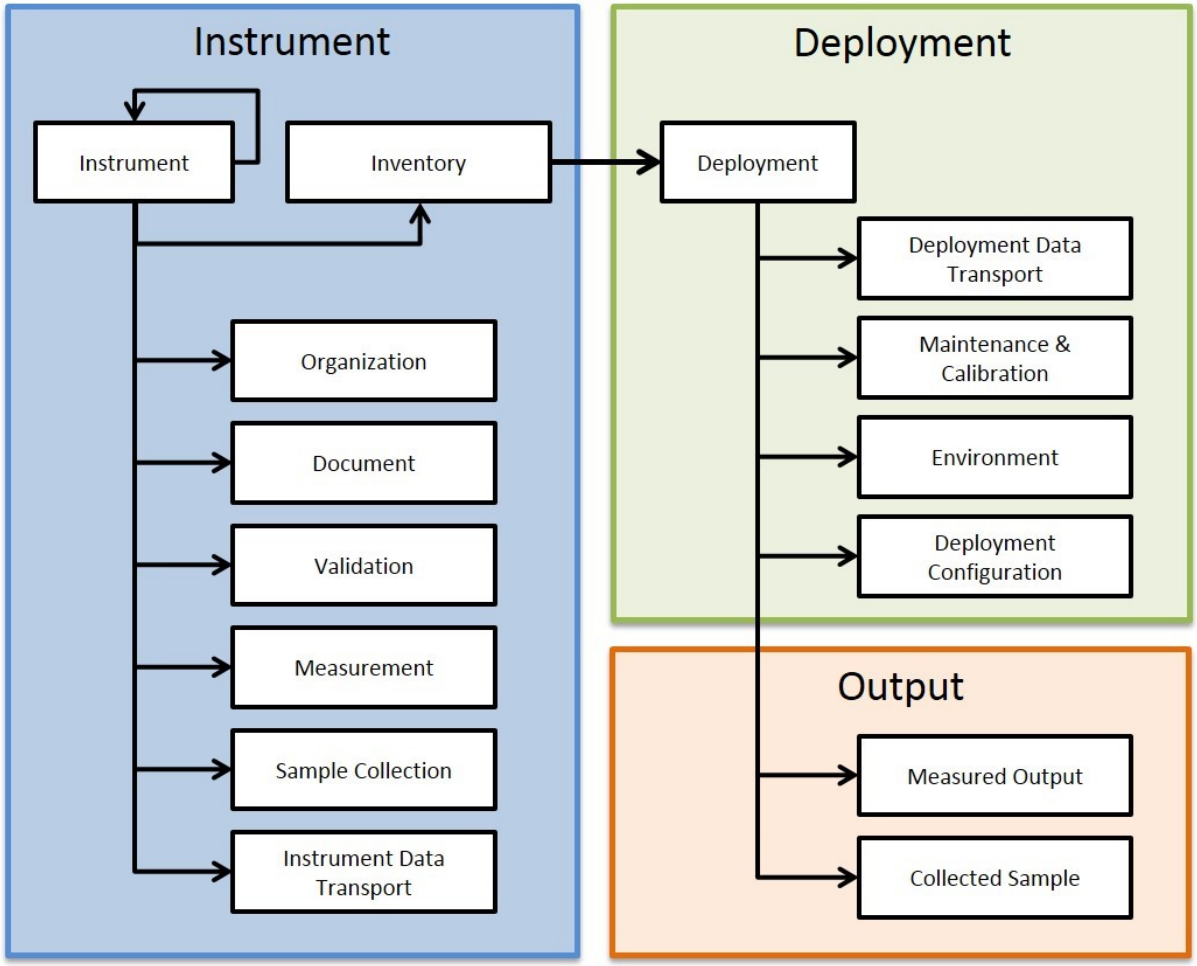


Figure 4: Domains and entities of the SCMS.

1.6 Implementation

The three domains of the Sensor Common Model Specification (SCMS) may be implemented with various database technologies (e.g. relational, graph, or document databases. Here are some examples of how you may implement the 3 domains.

1. **Instrument**: This is a functional description of each sensor including its ownership and manufactory specifications. These metadata can be stored as a library for supporting investigator selection of appropriate sensors and deploying for different studies.
   These metadata may be best as a relational or a graph metadata repository may be best implemented as a. Document databases are not recommended here since this Library portion of the model is highly relational. If you plan to directly store large amounts of binary objects such as PDF documents as a part of the metadata, a relational database would be better suited since graph databases have very limited support for binary large objects. On the other hand, if you plan to only store file paths to external files then a graph database may be more suitable as graphs have better support for hierarchical structures such as instruments containing a hierarchy of sensors. Graph databases also have better support

for web-linked data. In other words, graphs provide better support for ternary or more degrees of relationship types, along with many-to-many cardinality as in the case of the Instrument's self-referencing relationships and the ternary degree relation between the Instrument, Organization, and Inventory entities. We at Utah will are implement this as a graph.

2. **Deployment**: This is the metadata regarding how each physical instrument or sensor is deployed. Deployment provides information on how the device was configured and setup and inform the investigator on using the Output for specific study analysis.
   The deployment model is fairly simple with one-to-many relationships surrounding the Deployment entity type. Therefore, a relational or graph database may be equally well suited.
3. **Output**: The actual output received from each deployed instrument.
   A document store database may be better suited for this purpose. Sensor output may be generated with a high frequency and generally sensors output a type of file such as JSON or XML or text file, which fits well with the nature of a document data store. The high throughput of data may also require the use of Big Data technologies such as a distributed file system and a framework for parallel data processing. Document databases are generally well suited and designed for Big Data Technologies.
   For example, imagine 100 sensors collecting data every minute, over a period of one year. That would mean 100 x 1440 x 365 = 52,560,000 Records. Although this may sound reasonable for a relational DB, if we scale this up to 1000 sensors collecting data every 10 seconds, for ten years would mean; this would mean 1000 x (1440 x 6) x 3650 = 31,536,000,000 records. This would become much more difficult to transform, process and store in a single large relational table.

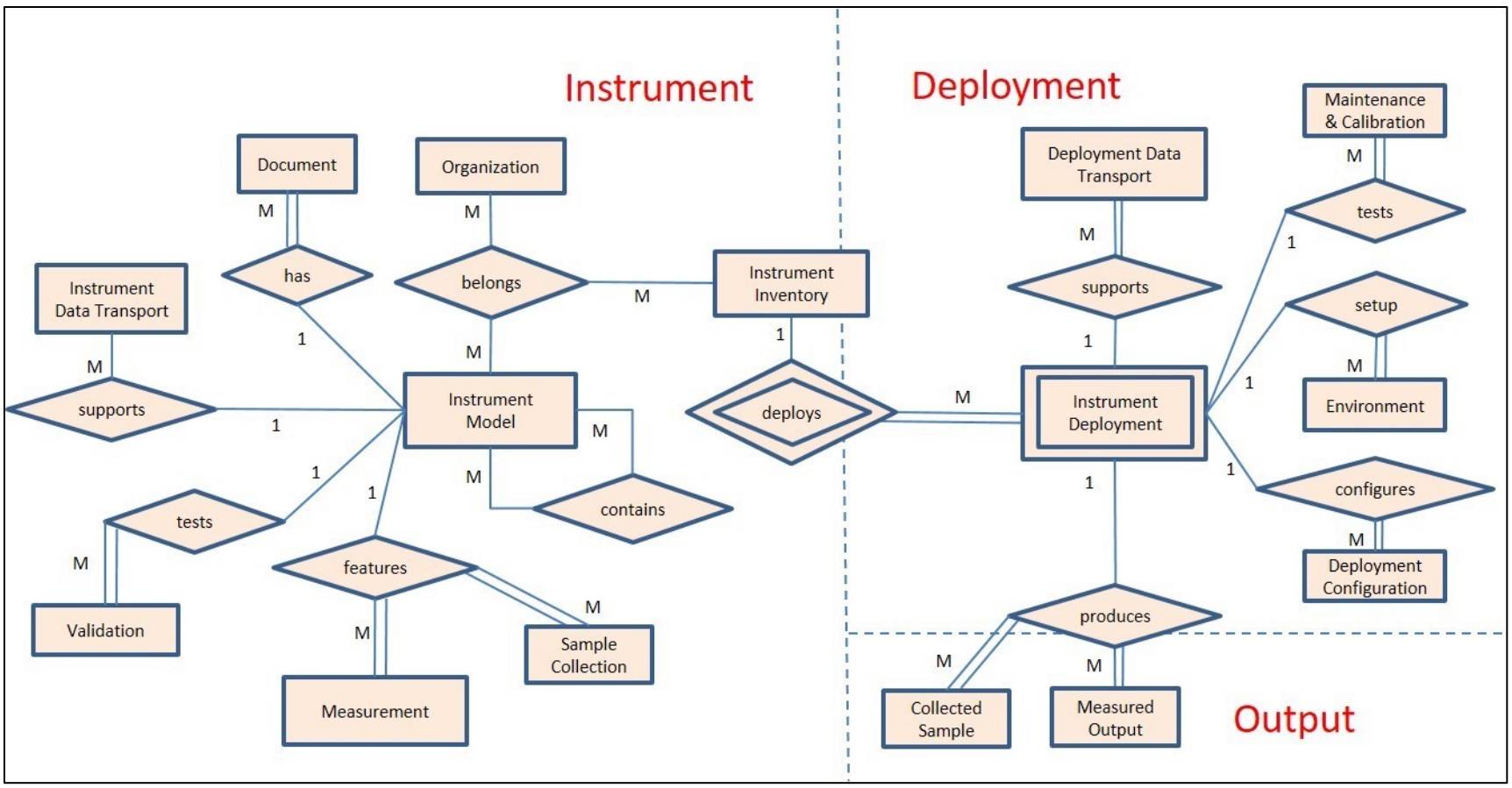


Figure 5: Entity relationship diagram of SCMS.

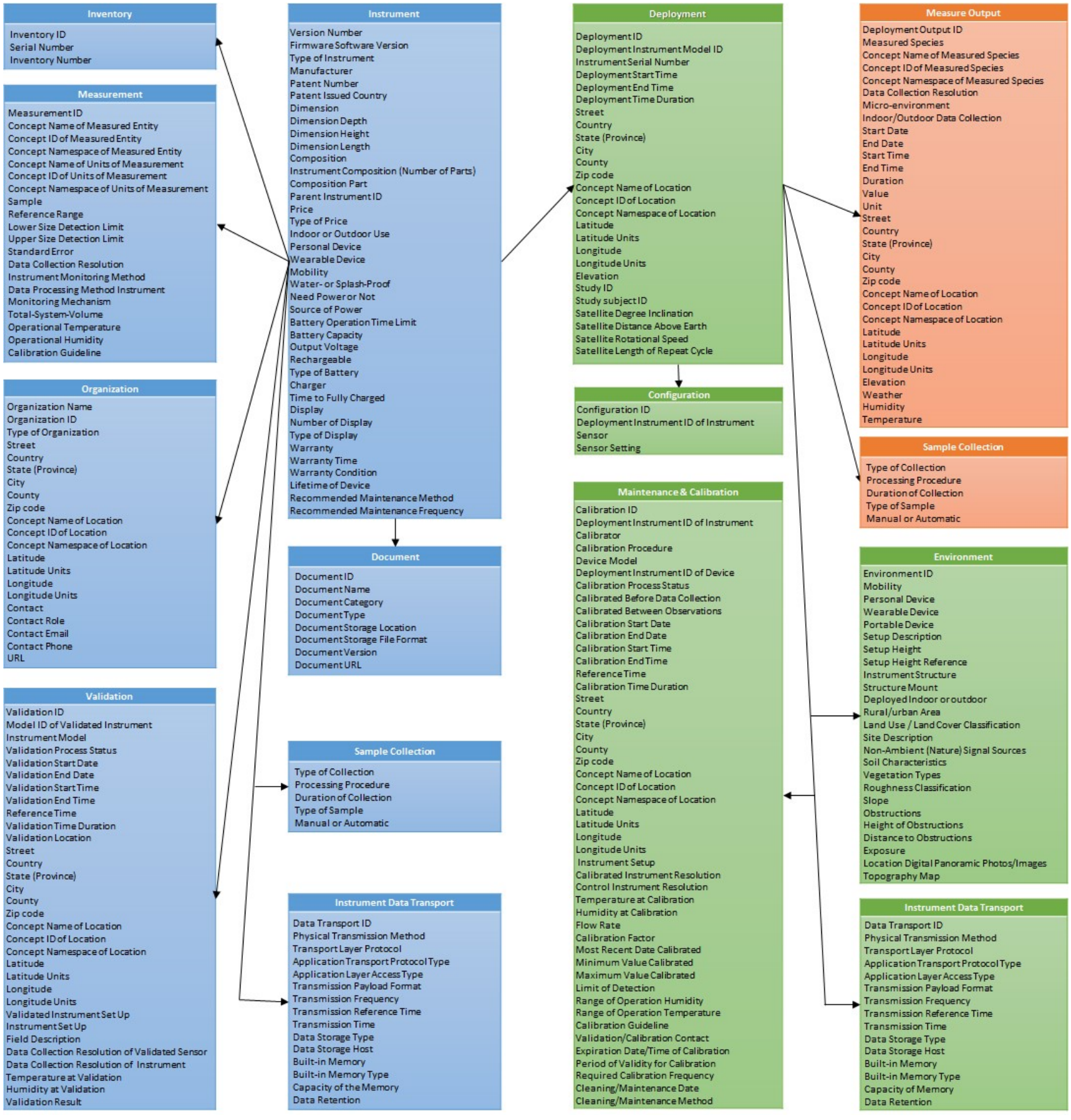


Figure 6: Entities and attributes of SCMS.

2. Instrument

2.1. Instrument model

The Instrument model data element is a list of metadata elements used to describe general information about the instrument, such as the model, capacity, version, content, power, display, manufacturer, price and species it measures.

Table 3: Instrument model entity details.

Data Element	Required	Description	Data Type	Example Value
Model Name	Y	The term by which the instrument is known. This could be a trade name or an alias.	String	AirU, MiniVol™ TAS, FooBot, Dylos 1700, personal UFP (PUFP) sensor
Model ID	Y	The unique identifier used to differentiate each model of an instrument made by certain manufacturer.	String (Numerical)	DC 1700, DC 1100
Version Number	N	The current version of the instrument model. It differentiates instruments within the same model. The usually refers to a version of the hardware.	String (Numerical)	1,2,3, alpha, beta
Firmware Software Version	N	Current firmware or software version of the model.	String (Numerical)	1,2,3
Type of Instrument	N	The category of instrument based on the species measured by the instrument.	Category	Gas sensor, particle sensor
Manufacturer	N	The person, group, or organization that develops or produces the instrument.	String	University of Utah, Airmetrics
Patent Number	N	The serial number of the patent, if the instrument is patented.	String	Patent # US 844965
Patent Issued Country	N	The country issuing the patent.	Category	US
Dimension	N	The size of the instrument in physical space. The dimension could have attributes of depth, height, and length. Each dimension includes a value and unit. (Use this if the dimensions aren't available discretely, else use the below fields.)	Value and Unit	10 x 10 x 5 cm; 10 in * 20 in * 15 in; depth: 10 cm, height: 10 cm, width: 10 cm
Dimension Depth	N	The depth or thickness of the instrument.	Value and Unit	10 cm

Dimension Height	N	The vertical height of the instrument.	Value and Unit	20 mm
Dimension Length	N	The horizontal length of the instrument.	Value and Unit	3 nm
Composition	N	The description of the composition of combining parts or elements making up of the instrument.	String	Comprised of an evaporation–condensation-tube, a miniature diaphragm air pump, an optical detection module, a flow regulator, water tank, GPS, and battery pack in a plastic shell body. The instrument includes 3 parts, and they are PM2.5 sensor, GPS module, and a backpack
Parent Instrument ID	N	Foreign Key. The Instrument Model ID representing the parent instrument containing this instrument.	String (Numerical)	12345.6
Price	N	The cost of the instrument. This could be a potential price or price range of the instrument, such as the manufacturer recommended price, actual price, or price range to purchase the instrument.	Value and Unit	$30
Type of Price	N	Whether the price/price range is the potential price or actual price to purchase the instrument.	Category	Manufacturer recommended price
Indoor or Outdoor Use	N	Whether the instrument is intended to be used inside a building or structure that is protected from the natural environment. Or, if the instrument can be used outdoors and can tolerate exposure to the natural environment.	Category	Indoor, outdoor, indoor/outdoor, indoor and outdoor
Personal Device	N	Whether or not the instrument is intended to be used to and track information for individuals.	Yes, No	Yes, No

Wearable Device	N	Whether or not the instrument can be worn by individuals and track information.	Yes, No	Yes, No
Mobility	N	Whether the instrument can be moved around for measuring the species.	Category	Mobile, fixed
Water- or Splash-Proof	N	Whether or not the instrument can tolerate exposure to water.	Yes, No	Yes, No
Need Power or Not	N	Whether or not the instrument needs a source of power for its normal function. If power is needed, the type of power should be listed. See "Source of Power."	Yes, No	Yes, No
Source of Power	N	The type of power that supports the instrument for its normal function/s.	Category	Battery, AC, solar, wind
Battery Operation Time Limit	N	The duration of battery life, if the "source of power" is battery.	Value and Unit	12 hours
Battery Capacity	N	The amount of electric charge the battery can deliver at the rated voltage.	Value and Unit	2200mAh
Output Voltage	N	The voltage released by the battery.	Value and Unit	14.8V
Rechargeable	N	Whether or not the battery's electric charge can be restored by connecting the battery to a recharging device.	Yes, No	No
Type of Battery	N	The category of battery, based on the chemical used in the battery's electrochemical cells.	String	Lithium Ion batteries, Nickel–cadmium battery
Charger	N	If the battery is rechargeable, this element is used to describe the charger.	String	1.2 amp external battery charger
Time to Fully Charged	N	The time taken to recharge the battery.	Value and Unit	Full recharge in less than 6 hours
Display	N	Whether or not the instrument is capable of displaying information. If yes, more information can be recorded in the following data element, such as how many monitors, and what type of monitors does it have.	Yes, No	Yes

Number of Display	N	The number of displays with the instrument.	String (Numerical)	2
Type of Display	N	The category of the monitor used to display information.	Category	LCD screen, LED monitor
Warranty	N	Whether or not the instrument comes with a warranty. If Yes, more information can be provided, such as the warranty time and warranty condition.	Yes, No	Yes
Warranty Time	N	The length of time covered by the instrument's warranty.	Value and Unit	1 year
Warranty Condition	N	The facts or conditions under which the warranty is valid.	String	The period of warranty shall start from the date of delivery of the product to the customer and shall cover a period of 2 years
Lifetime of Device	N	The duration of time during which the instrument is expected to function properly according to the manufacturer.	String	Re-usable; If a problem must replace (not repairable)
Recommended Maintenance Method	N	The method suggested for maintaining the instrument.	String	Clean with compressed air
Recommended Maintenance Frequency	N	The frequency at which the maintenance should be repeated	String	At least once a month

2.1.1. Conventions

- Each type of instrument with a certain model made by a certain manufacturer will have an instrument model ID.
- The instrument is uniquely identified by its instrument model and the version of the hardware/software. The Instrument Modeled ID is the unique identifier to differential certain type of instrument with certain model and specific hardware/software version.

2.2. Instrument Inventory

The Instrument Inventory is used to register the physical instrument of a specific model with certain version in a library of sensors.

Table 4: Instrument inventory entity details.

Data Element	Required	Description	Data Type	Example Value
Library ID	Y	This is a foreign key to Instrument Model	Number	123

Serial Number	Y	Serial number of physical Instrument	String	A1B234567
Inventory Number	N	Internal inventory number used to uniquely identify each instrument.	String	A1B234567

2.3. Measurement

The measurement data element is a list of metadata elements used to describe the characteristics of a species t the instrument measures.

Table 5: Measurement entity details.

Data Element	Required	Description	Data Type	Example Values
Measurement ID	Y	An identifier generated by the system that identifies each type of measurement.	String (Numerical)	12345
Concept Name of Measured Entity	N	A name representing the measured entity which is assigned by the system or taken from a reference terminology or ontology.	String	Fine particles, motion, temperature
Concept ID of Measured Entity	N	An ID representing the measured entity which is assigned by the system or taken from a reference terminology or ontology.	String (Numerical)	86257
Concept Namespace of Measured Entity	N	The standard terminology/reference to be used as a reference for measured entity data.	Category	chEBI (https://www.ebi.ac.uk/chebi/)
Concept Name of Units of Measurement	N	A name representing the units of measurement which is assigned by the system or taken from a reference terminology or ontology.	String	mg/m3
Concept ID of Units of Measurement	N	An ID representing the units of measurement which is assigned by the system or taken from a reference terminology or ontology.	String (Numerical)	12345
Concept Namespace of Units of Measurement	N	The standard terminology/reference to be used as a reference for the units of measurement data.	Category	QUDT (http://www.linkedmodel.org/catalog/qudt/1.1/index.html), Units of Measurement Ontology (https://bioportal.bioont

				ology.org/ontologies/UO)
Sample	N	A category referring to the nature of the sample analyzed to measure an entity.	Category	Air, dust
Reference Range	N	The reference range used by the instrument for the limit of detection. This could include the upper value and the lower value.	Range (Value and Unit)	0mg/m3 to 1.6 mg/m3
Lower Size Detection Limit	N	A manufacturer reference limit value distinguishing the lower size detection limit of the instrument. The format could be of a value and unit.	Value and Unit	10 um
Upper Size Detection Limit	N	A manufacturer reference limit value distinguishing the upper size detection limit of the instrument. The format could consists of a value and unit.	Value and Unit	100 um
Standard Error	N	An indication used to represent the precision of instrument.	Value and Unit	±16%
Data Collection Resolution	N	Time step used for data collection.	Value and Unit	1 second
Instrument Monitoring Method	N	The description of the method utilized by the instrument for monitoring.	String	Resistor for humidity sensor and a thermostat for temperature; particle counter based on light-scattering technology
Data Processing Method Instrument Monitoring Mechanism	N	The description of the mechanism by which the signal and data are generated and processed	String	The two central processing units on a board convert analog laser particle scattering signature to digital counting data along with the global positioning system (GPS).
Total-System-Volume	N	This value represents the capacity of the instrument to collect the sample by volume.	Value and Unit	1500 cm3
Operational Temperature	N	The reference range of temperature used for the instrument under normal working conditions. This could be a range of	Value and Unit	50 F to 80 F

		temperature with upper and lower values.		
Operational Humidity	N	The reference range of humidity used for the instrument under normal working condition. This could be a range of humidity with upper and lower values.	Value and Unit	80% to 90%
Calibration Guideline	N	A reference to whether or not a calibration guideline exists for the instrument sensor. If yes, the text, document, or links of the manufacturer's recommended calibration should be listed.	Boolean	Yes or no, if yes, the text, document, or links of the manufacturer's recommended calibration should be listed.

2.3.1. Conventions

- Each measurement will have a measurement ID.

2.4. Validation

The validation is the process to test and evaluate the instrument's capability of measuring what it supposed to measure. The Validation data element is a list of metadata elements used to describe the process, settings and results of the validation process.

Table 6: Validation entity details.

Data Element	Required	Description	Data Type	Example Value
Validation ID	Y	A unique identifier used to differentiate each validation event.	String (Numerical)	123456
Model ID of Validated Instrument	Y	The identifier that identifies the instrument validation. This ID is a foreign key linking the Validation entity to the Instrument Model entity which documents the information about the validated instrument.	String (Numerical)	123
Instrument Model	Y	The device used as a reference to validate the instrument. The data generated from the tested instrument is compared with the data of control device.	String	PMS 1003

		If the controlled device is registered as an instrument model entity, the Versioned Instrument ID of the controlled device should be provided. Control device used for validation. This could be an "instrument" entity. Refer to "instrument administration data"		
Validation Process Status	N	The current state or condition status of the validation process.	Category	In process, done
Validation Start Date	N	The point in time as month, day, year, where validation begins.	Date	05May2016
Validation End Date	N	The point in time as month, day, year where validation terminates.	Date	05May2016
Validation Start Time	N	The point in time as hour, minute, seconds where validation begins.	Time	01:22:16
Validation End Time	N	The point in time as hour, minute, seconds where validation terminates.	Time	02:22:16
Reference Time	N	Time zone of the validation.	Category	Mountain Standard
Validation Time Duration	N	The amount of time it took to validate the instrument.	Value and Unit	3 months
Validation Location	N	The name of the location where the validation took place.	String	In the lab
Street	N	The street where the instrument was validated.	String	545 South 700 East
Country	N	The country where the instrument was validated.	String	USA
State (Province)	N	The state (province) where the instrument was validated.	String	UT

City	N	The city where the instrument was validated.	String	Salt Lake City
County	N	The county where the instrument was validated.	String	Salt Lake
Zip code	N	The zip code where the instrument was validated.	String (Numerical)	84102
Concept Name of Location	N	A name representing the location where the instrument was validated which is assigned by the system or taken from a reference terminology or ontology. This name can incorporate rows 14-18 combined as a single entity or examined independently.	String	USA_UT_SLC_UT_84102
Concept ID of Location	N	An ID representing the location where the instrument was validated which is assigned by the system or taken from a reference terminology or ontology. This name can incorporate rows 14-18 combined as a single entity or examined independently.	String (Numerical)	12345
Concept Namespace of Location	N	The standard terminology/reference to be used as a reference for location data.	Category	GeoNames Ontology (http://www.geonames.org/)
Latitude	N	The latitude in which the organization resides.	String (Numerical)	40.76
Latitude Units	N	The direction in which the latitude is running.	String	Degrees North
Longitude	N	The longitude in which the organization resides.	String (Numerical)	-111.863, 111.863

Longitude Units	N	The direction in which the longitude is running.	String	Degrees West, West
Validated Instrument Set Up	N	The description of how the validated instrument is fastened securely in position fixed or set up during validation.	String	Set up on a tripod located on top of the cabin
Instrument Set Up	N	The description of how the instrument is fastened securely in position or set up during validation.	String	Co-located with Hawthorn FRM monitors
Field Description	N	The description of the site where the instrument is set up.	String	In an open field near the I-15; In an open field surrounded by the trees
Data Collection Resolution of Validated Sensor	N	Time step for data collection of the tested instrument.	Value and Unit	1s, 1min
Data Collection Resolution of Instrument	N	Time step for data collection of the instrument.	Value and Unit	1s, 1min
Temperature at Validation	N	Ambient temperature during validation.	Value and Unit	75F, 30C
Humidity at Validation	N	Ambient humidity during calibration.	Value and Unit	70%
Validation Result	Y	The performance of the tested instrument when compared with the instrument.	String	The correlation of the PM2.5 from the two sensor is 0.9; passed the validation

2.4.1. Conventions

- Each Validation event should have a unique identifier, the Validation ID.
- The instrument that is used for validation is called “Validation Instrument”, whereas the instrument that is used as a reference device for validation is called “Control Instrument”.
- The Versioned Model ID of the Validation Instrument is used as a foreign key to link the Validation event with the information of the instrument that is evaluated.
- The Versioned Model ID of the Control Instrument is used as a foreign key to link the Validation event with the information of the instrument that is used as a reference device.

2.5. Instrument Data Transport

The Instrument Data Transport data element is a list of metadata elements that describe the capacity of the data community between the instrument and the data storage center.

Table 7: Instrument data transport entity details.

Data Element	Required	Description	Data Type	Example Value
Data Transport ID	Y	The unique identifier used to differentiate data transport from others.	String (Numerical)	123
Physical Transmission Method	N	Means of transmission of data from field to initial central collection point.	category	Point to Point RF, LAN, satellite downlink, blue tooth, Wi-Fi, cellular network, landline telephony, other
Transport Layer Protocol	N	Type of transport layer implemented by the physical transmission method.	category	TCP/IP, UDP,PROPRIETARY
Application Transport Protocol Type	N	Type of protocol used for connectivity implementation of data transmission.	category	HTTP,HTTPS,SMTP,JMS,SSH,PROPRIETARY
Application Layer Access Type	N	Access method to the data stored from the sensor.	category	API-REST,API-SOAP,JDBC, SQL
Transmission Payload Format	N	Type of message used for data transmission.	category	csv, xml, json, binary
Transmission Frequency	N	Frequency of data transmission.	string	real time, twice a day, every 30 minutes
Transmission	N	Reference time used.	category	Greenwich Mean Time (GMT)

Reference Time				
Transmission Time	N	Time duration of the data transmission.	value and unit	12 hours
Data Storage Type	N	Category of physical data storage mechanism.	category	cloud storage, data warehouse, direct broadcast station, localhost
Data Storage Host	N	Description of the hostname and location of the data storage.	string	data.proxyhost.somewhere.com, 127.0.0.1
Built-in Memory	N	Does the instrument have built in memory or not.	Yes, No	Yes
Built-in Memory Type	N	The kind of memory the instrument uses for the built-in memory.	category	volatile, non-volatile
Capacity of the Memory	N	The storage space of the built-in memory.	Value and unit	12MB
Data Retention	N	The action to be taken when out of memory.	category/String	overwriting oldest data

2.5.1. Conventions

- Each Instrument Data Transport will have a unique Data Transport ID.

2.6. Organization

The Organization data element is a list of metadata elements used to describe an organization (i.e., owner, manufacturer, etc.) tied to the instrument.

Table 8: Organization entity details.

Data Element	Required	Description	Data Type	Example Values
Organization Name	Y	A name representing the organization.	String	University of Utah, EPA
Organization ID	Y	An identifier generated by the system that identifies	String (Numerical)	12345

		a unique organization in our database.		
Type of Organization	N	A representative category for which the organization belongs.	Category	Manufacturer, owner, research institute, government
Street	N	The street in which the organization resides.	String	545 South 700 East
Country	N	The country in which the organization resides.	String	USA
State (Province)	N	The state (province) in which the organization resides.	String	UT
City	N	The city in which the organization resides.	String	Salt Lake City
County	N	The county in which the organization resides.	String	Salt Lake
Zip code	N	The zip code in which the organization resides.	String (Numerical)	84102
Concept Name of the Location	N	A name representing the location of the organization which is assigned by the system or taken from a reference terminology or ontology. This name can incorporate rows 8-12 combined as a single entity or examined independently.	String	USA_UT_SLC_UT_84102
Concept ID of the Location	N	An ID representing the location of the organization which is assigned by the system or taken from a reference terminology or ontology. This name can incorporate rows 8-12 combined as a single entity or examined independently.	String (Numerical)	12345
Concept Namespace of the Location	N	The terminology/referenc e to be used as a	Category	GeoNames Ontology (http://www.geonames.org/)

		reference for location data.		
Latitude	N	The latitude in which the organization resides.	String (Numerical)	40.76
Latitude Units	N	The direction in which the latitude is running.	String	Degrees North
Longitude	N	The longitude in which the organization resides.	String (Numerical)	-111.863, 111.863
Longitude Units	N	The direction in which the longitude is running.	String	Degrees West, West
Contact	N	A name for the contact within the organization.	String	John Doe
Contact Role	N	The role/s the contact possesses within the organization.	Category	Manager, staff
Contact Email	N	An email address to contact the organization's contact.	String	JohnDoe@gmail.com
Contact Phone	N	A phone number to contact the organization's contact.	String (Numerical)	(800) 123-4567
URL	N	An address to a resource on the Internet that contains information regarding the organization.	String	www.organization.com

2.6.1. Conventions

- Each organization will have an organization ID.

2.7. Document

The Document data element is a list of metadata elements used to describe any documentation supporting instrument models.

Table 9: Document entity details.

Data Element	Required	Description	Data Type	Example Value
Document ID	Y	The unique identifier that is used to differential the document with the other documents.	String (Numerical)	doc.1
Document Name	N	The name of the document.	String	User manual for MiniVol
Document Category	N	The category of the document.	Category	User manual, calibration guideline,

				topography map, location photo
Document Type	N	The type of the document.	Category	Hardcopy, online, electronic copy
Document Storage Location	N	The location where the document is stored. This could be a physical place, on the hardware, or online. If online, the URL might be provided.	String	In the shipping box, in the database, in the hardware
Document Storage File Format	N	If electronic, the document file format.	Category	pdf, MS world, HTML, ASCII
Document Version	N	The version of the document.	String (Numerical)	1.0.0;1.0.1, alpha, beta
Document URL	N	URL of the document, if online.	String	http://www.airmetrics.com/products/minivol/

2.7.1. Conventions

- Each Document will have a unique Document ID.

2.8. Sample Collection

The Sample data element is a list of metadata elements that describe the characteristics of samples collected by the instrument and the procedure used for sample collection.

Table 10: Sample Collection entity details.

Data Element	Required	Description	Data Type	Example Values
Type of Collection	N	How the sensor collects data	Category	Filter, bag, cartridge
Processing Procedure	N	How to find the value of the measurement	Text	Weight, Send to Lab
Duration of Collection	N	How long the sensor is open to collect data	String (Numerical)	3 days
Type of Sample	N	The entity the sensor captures.	Category	Air
Manual or Automatic	N	Does the sensor require a human to measure the value output?	Category	Manual, Automatic

3. Deployment

The Deployment is the event that the physical instrument is utilized and brought to effective action.

3.1. Deployment

The Deployment data element is a list of metadata that is used to describe the details of how, when, where and for what the instrument is deployed into action. For instruments with different types, specific metadata element might be listed to document the deployment. For example, specific data elements, such as Satellite Degree Inclination, is from the deployment of satellite.

Table 11: Deployment entity details.

Data Element	Required	Description	Data type	Example Values
Deployment ID	Y	A unique identifier used to distinguish the deployment.	String (Numerical)	DEP.1
Deployment Instrument Model ID	Y	An identifier generated by the system that identifies very unique deployed instrument.	String	12345
Instrument Serial Number	Y	A number provided by a manufacturer showing the position of an instrument in a series for the purposes of identification.	String	12345
Deployment Start Time	N	The time at which the instrument was deployed. Note this could be different from the time when the instrument starts measuring.	Date-Time	
Deployment End Time	N	The point in time at which the instrument deployment terminated. Note this could be different from the time when the instrument ends measuring.	Date-Time	
Deployment Time Duration	N	The duration of time an instrument has been deployed, if Deployment Start and	Time	6 weeks, 21 months

		End Times are available.		
Street	N	The street in which the instrument is deployed.	String	545 South 700 East
Country	N	The country in which the instrument is deployed.	String	USA
State (Province)	N	The state (province) in which the instrument is deployed.	String	UT
City	N	The city in which the instrument is deployed.	String	Salt Lake City
County	N	The county in which the instrument is deployed.	String	Salt Lake
Zip code	N	The zip code in which the instrument is deployed.	String (Numerical)	84102
Concept Name of the Location	N	A name representing the location where the instrument is deployed which is assigned by the system or taken from a reference terminology or ontology. This name can incorporate rows 15-19 combined as a single entity or examined independently.	String	USA_UT_SLC_UT_84102
Concept ID of the Location	N	An ID representing the location where the instrument is deployed which is assigned by the system or taken from a reference terminology or ontology. This name can incorporate rows 15-19 combined as a single entity or examined independently.	String (Numerical)	12345

Concept Namespace of the Location	N	The standard terminology/reference to be used as a reference for location data.	Category	GeoNames Ontology (http://www.geonames.org/)
Latitude	N	The latitude in which the organization resides.	String (Numerical)	40.76
Latitude Units	N	The direction of the parallels of latitude.	String	Degrees North
Longitude	N	The longitude in which the organization resides.	String (Numerical)	-111.863, 111.863
Longitude Units	N	The direction of the longitude meridian.	String	Degrees West, West
Elevation	N	The height above sea level at which the instrument is deployed.	Value and Unit	
Study ID	N	This element is used to link the instrument to the clinical study if possible.	String	12345
Study subject ID	N	This element is used to link the instrument to the study subject, e.g. the person.	String	12345
Satellite Degree Inclination	N	The satellite's degree of orbit.	String (Numerical)	98.2 degree inclination
Satellite Distance Above Earth	N	The distance at which the satellite orbits above Earth.	String (Numerical)	705 km, 438 miles
Satellite Rotational Speed	N	The number of revolutions encountered by the satellite per cycle.	String (Numerical)	233 revolutions per cycle
Satellite Length of Repeat Cycle	N	The number of days until the satellite repeats its cycle.	String (Numerical)	16 days

3.1.1. Conventions

- Each Deployment has a unique Deployment ID.
- Instrument Versioned Model ID is an internal identifier (foreign key) linking the physical deployed instrument to the general information of the instrument (Instrument Model).

3.2. Maintenance and Calibration

Calibration is the action or process of calibrating an instrument readings. The instrument that is being calibrated is called Calibrated Instrument, whereas the instrument used as a standard is called Controlled Instrument. The Calibration data element is a list of metadata used to describe the process of calibration. The maintenance data element is a list of metadata used to describe the process of maintenance during a certain deployment.

Table 12: Maintenance and Calibration entity details.

Data Element	Required	Description	Data Type	Example Value
Calibration ID	N	A unique identifier used to differential each calibration event with others.	String (Numerical)	123
Deployment Instrument ID of the Calibrated Instrument	N	An identifier generated by the system that identifies the Calibrated Instrument.		
Calibrator	N	The person/group who calibrated the instrument.	Category	manufactory, institute, owner
Calibration Procedure	Y	Description of the Calibration Procedure.	Text	Press the calibration button; Weigh each individual filter
Instrument Model	Y	The instrument used as a reference for calibration. If the device is registered as a "Deployed Instrument", the Deployed Instrument ID of the controlled device should be provided as a reference. instrument used for Calibration	String	PMS 1003
Deployment Instrument ID of the Instrument	N	An identifier generated by the system that identifies the instrument.		
Calibration Process Status	N	The current/latest status of the calibration.	Category	In Process, Done
Calibrated Before Data Collection	N	Was the instrument calibrated before	yes, no	yes

		the first recording took place?		
Calibrated Between Observations	N	Was the instrument calibrated between different observations?	yes, no	yes
Calibration Start Date	N	The point in time as month, day, year, where calibration begins.	Date	05May2016
Calibration End Date	N	The point in time as month, day, year where calibration terminates.	Date	05May2016
Calibration Start Time	N	The point in time as hour, minute, seconds where calibration begins.	Time	01:22:16
Calibration End Time	N	The point in time as hour, minute, seconds where calibration terminates.	Time	01:30:15
Reference Time	N	Time Zone referenced for the calibration times.	Category	Mountain Standard
Calibration Time Duration	N	The amount of time that has passed for calibration.	Value and Unit	1 month
Street	N	The street where the instrument was calibrated.	String	545 South 700 East
Country	N	The country where the instrument was calibrated.	String	USA
State (Province)	N	The state (province) where the instrument was calibrated.	String	UT
City	N	The city where the instrument was calibrated.	String	Salt Lake City
County	N	The county where the instrument was calibrated.	String	Salt Lake
Zip code	N	The zip code where the instrument was calibrated.	String (Numerical)	84102
Concept Name of the Location	N	A name representing the location where the	String	USA_UT_SLC_UT_84102

		instrument was calibrated which is assigned by the system or taken from a reference terminology or ontology. This name can incorporate rows 18-22 combined as a single entity or examined independently.		
Concept ID of the Location	N	An ID representing the location where the instrument was calibrated which is assigned by the system or taken from a reference terminology or ontology. This name can incorporate rows 18-22 combined as a single entity or examined independently.	String (Numerical)	12345
Concept Namespace of the Location	N	The standard terminology/reference to be used as a reference for location data.	Category	GeoNames Ontology (http://www.geonames.org/)
Latitude	N	The latitude in which the organization resides.	String (Numerical)	40.76
Latitude Units	N	The direction of the parallels of latitude	String	Degrees North
Longitude	N	The longitude in which the organization resides.	String (Numerical)	-111.863, 111.863
Longitude Units	N	The direction of the longitude meridian	String	Degrees West, West
Instrument Setup	N	The description of how the Instrument is arranged during the calibration. This could also include set up parameters.	String	set up on a tripod directly adjacent to the cabin housing the TEOM-FDMS

Calibrated Instrument Resolution	N	Time step for data collection in the real situation of the calibrated instrument.	Value and Unit	1s
Control Instrument Resolution	N	Time step for data collection in the real situation of the instrument.	Value and Unit	1s
Temperature at Calibration	N	Ambient temperature during calibration.	Value and Unit	75F, 30C
Humidity at Calibration	N	Ambient humidity during calibration	Value and Unit	50%
Flow Rate	N	The rate at which air flows through the instrument.	Value and Unit	4 Liters/minute
Calibration Factor	Y	The value calculated based on things such as temperature, humidity, wind speed, min, max, etc. Each calibration or response factor represents the slope of the line between the response for a given standard and the origin. The same as correction factor.	Value	1/1/1900
Most Recent Date Calibrated	N	The date of the last calibration.	Value and Unit	5/7/2015
Minimum Value Calibrated	N	The minimum value the manufacturer found.	Value and Unit	.5 micrograms/cubic meters
Maximum Value Calibrated	N	The maximum value the manufacturer found.	Value and Unit	13 micrograms/cubic meters
Limit of Detection	N	Range in which PM sensors detect particles. Could include upper and lower limit.	Value and Unit/Range	.5-13 micrograms/cubic meter
Range of Operation Humidity	N	Humidity limits for gas sensor detection. Could include upper and lower limit.	Value and Unit	0-50%

Range of Operation Temperature	N	Temperature limits for gas sensor detection. Could include upper and lower limit.	Value and Unit	0-50 Celsius
Calibration Guideline	N	The guideline used in the calibration process.	Document Entity	
Calibration Contact	N	The contact who did the calibration.	Text	Kerry Kelly
Expiration Date/Time of Calibration	N	The expire date/time of this calibration	Date and Time	5/5/2017
Period of Validity for Calibration	N	The length of time the calibration is validated.	Value and Unit	3 months
Required Calibration Frequency	N	The frequency of calibration.	Value and Unit	every 3 months
Cleaning/Maintenance Date	N	The time the instrument is cleaned/maintained.	String	06/24/2016
Cleaning/Maintenance Method	N	The method with which the instrument is cleaned/maintained.	String	Clean with compressed air

3.2.1. Conventions

- Each Calibration will have a unique Calibration ID.
- The instrument that is being calibrated is called Calibrated Instrument, whereas the instrument used as a standard is called Controlled Device.
- The Deployment Instrument ID of the Calibrated Instrument is an internal identifier (foreign key) referring the instrument being calibrated. This ID can be used to link the calibration with the information of the instrument being calibrated.
- The Deployment Instrument ID of the Controlled Device is an internal identifier (foreign key) referring the Controlled Device. This ID can be used to find the information of the Controlled Device.

3.3. Environment

The Environment data element is a list of metadata elements used to describe how the instrument is deployed within a particular environment and settings.

Table 13: Environment entity details.

Data Element	Required	Description	Data Type	Example Value
Environment ID	Y	The unique ID to identify the setting of	ID	123456

		the setup and the field.		
Mobility	N	The instrument is movable or fixed to certain site when deployed.	category	mobile, fixed
Personal Device	N	Is the instrument used by individuals as a personal device	yes, no	Yes, No
Wearable Device	N	Is the instrument worn by individuals as a wearable device to track information?	yes, no	Yes, No
Portable Device	N	Is instrument deployed as a portable device that can easily be carried?	yes, no	Yes, No
Setup Description	N	A description of how the instrument is arranged	string	the sensor is set up on a tripod; the sensor is worn by the study subject
Setup Height	N	If the instrument is fixed in a site, what is the height above certain reference level? This element should be used with "setup height reference" element.	value and unit	5m; 12 feet
Setup Height Reference	N	The reference level to which the setup height is measured.	category	sea level, roof, the ground
Instrument Structure	N	Type of structure the instrument is installed on.	category	Mast; Tower; Tripod; Freestanding; Other (specify)
Structure Mount	N	Type of mount the structure uses/is affixed to.	category	Building/Rooftop (penetrating, Non-penetrating), Freestanding (concrete pad, guy wires)
Deployed Indoor or outdoor	N	The instrument is deployed in the building or out of doors.	category	indoor, outdoor, indoor/outdoor, indoor and outdoor
Rural/urban Area	N	The type of area the instrument is deployed.	category	rural, urban, urbanized

Land Use / Land Cover Classification	N	The land use/cover of the surrounding area within which the instrument is located.	category	Urban or Built-up Land Residential; Commercial and Services; Industrial
Site Description	N	The description of the site in which the instrument is used. It may include environmental, topographic, soil and/or vegetation information, or relationship of site to roadway surface (e.g., distance from the road).	string	The instrument is setup 100 feet away from the I-15 in an area with tall trees surrounded.
Non-Ambient (Nature) Signal Sources	N	Non-natural signal sources that might affect the instrument signal.	category	Air conditioner, heat pump, vent, south facing reflective wall (north of sensor), generator, diesel engine, manmade surfaces (asphalt, concrete)
Soil Characteristics	N	Texture, description and quartz content of soil.	category	Texture / Description / Quartz Content:1: Coarse / Loamy Sand / (0.82) ;2: Medium / Silty Clay Loam / (0.10) ;3: Fine / Light Clay / (0.25) ;4: Coarse Medium / Sandy Loam / (0.60) ;5: Coarse Fine / Sandy Clay / (0.52);6: Medium Fine / Clay Loam / (0.35) ;7: CoarseMedFine / Sandy Clay

				Loam / (0.60) ;8: Organic / / (0.40) ;9: Glacial Land Ice / Loamy Sand / (0.82)
Vegetation Types	N	Type of vegetation at station installation site.	category	Broadleaf – Evergreen (Tropical Forest); 2: Broadleaf – Deciduous Trees; 3: Broadleaf and Needle leaf Trees (Mixed Forest); 4: Needle leaf – Evergreen Trees ; 5: Needle leaf – Deciduous Trees (Larch); 6: Broadleaf Trees with Groundcover (Savanna) ;7: Groundcover Only (perennial) ;8: Broadleaf Shrubs with Perennial Groundcover ;9: Broadleaf Shrubs with Bare Soil ;10: Dwarf Trees and Shrubs with Groundcover (Tundra) ; 11: Bare Soil ;12: Cultivations (same parameters as for type 7) ;13: Glacial (same parameters as for types 11) ; other

Roughness Classification	N	Classification of effective terrain roughness.	category	Davenport classification-1: Sea 2: Smooth 3: Open 4: Roughly Open 5: Rough 6: Very Rough 7: Skimming 8: Chaotic
Slope	N	General slope (inclination from horizontal) of area surrounding station.	category	Slope Class: Percent Slope :1: 08 ;2: 830 ;3: > 30 ;4: 030 ;5: 08 & > 30 ;6: 830 & > 30 ;7: 08, 830, > 30 ;8: Glacial Ice ;9: Ocean/Sea
Obstructions	N	Obstructions around the instrument. The obstructions can be described using width, height, and distance to the sensor. See elements below.	category	tree, building, tower, fence, other
Height of Obstructions	N	Height of obstruction above reference level.	value and unit	10 feet, 100 meters
Distance to the Obstructions	N	Linear distance to obstructions	value and unit	10 feet
Exposure	N	Description of the exposure of the site/instrument in terms of obstructions to wind and sun and artificial temperature/ moisture sources.	string	The site is exposed to moisture
Location Digital Panoramic Photos and Drawings	N	Photos and graphic drawings that display the exposure and surrounding environment.	image	
Topography Map	N	Map image/file of the area surrounding the station.	image	

3.3.1. Conventions

- Each Setup and Field Description will have a unique Setup and Field Description ID.
- The Event ID is an internal identifier (foreign key) referring the event to which the Setup and Field Description is attached. The event can be the deployment or the calibration.

3.4. Deployment data transport

The Deployment Data Transport data element is a list of metadata elements that describe the data community between the instrument and the data storage center in certain deployment situation. The data elements are the same as shown in “Instrument Data Transportation” part. See “Instrument Data Transportation” part for reference.

3.4.1. Conventions

- Each Data Transport will have a unique Data Transport ID.

4. Output

4.1 Measured Output

The output of an instrument with certain deployment is the data/signal collected in specific time and location. The deployment output data element is the list of metadata of the data/signal generated/collected by the instrument.

Table 14: Measured Output entity details.

Data Element	Required	Description	Data Type	Possible/Example Value
Deployment Output ID	Y	The unique identifier used to differential the output of certain deployment.	String (Numerical)	123
Measured Species	Y	The species the instrument measures.	String	PM 2.5, PM 10, Ozone
Concept Name of the Measured Species	N	A name representing the measured species which is assigned by the system or taken from a reference terminology or ontology.	String	Fine particles, motion, temperature
Concept ID of the Measured Species	N	An ID representing the measured species which is assigned by the system or taken from a reference terminology or ontology, .standard universal term for our measured entity, from sources such as CHeBI	String	O3 for ozone or PM_2.5 for PM 2.5
Concept Namespace of the Measured Species	N	The standard terminology/reference used as a reference for identifying of the measured species. The Standard Ontology of Measured Entity	Category	CHeBI ontologies of ammonium are applications, biological roles and chemical roles
Data Collection Resolution	N	Time step for data collection in the real situation	Value and unit	1 min, 10 second
Micro-environment	N	The immediate small-scale environment where the data is collected.	Category	work, school, transportation
Indoor/Outdoor Data Collection	N	Is the data collected indoor or outdoor? Is this an indoor or outdoor sensor?	Category	Indoor, Outdoor sensor
Start Date	N	The start date of data capture. This could be different from the instrument deployment start	Date	05May2016

		date. The date of the recording		
End Date	N	The end date of data capture. This could be different from the instrument deployment end date.	Date	05-23-2016
Start Time	N	The point in time the data capture began. This could be different from the instrument deployment start time.	Time	01:16:33
End Time	N	The point in time the data capture terminated. This could be different from the instrument deployment start time.	Time	02:16:33
Duration	N	The length of time of data collection.	Time	60 minutes
Value	Y	Resulting output value from the data capture session.	Value	25
Unit	N	Unit of measure used for the output value.	Unit	Micrograms per cubic meter, ppm
Street	N	The street in which the deployed instrument collected data.	String	545 South 700 East
Country	N	The country in which the deployed instrument collected data.	String	USA
State (Province)	N	The state (province) in which the deployed instrument collected data.	String	UT
City	N	The city in which the deployed instrument collected data.	String	Salt Lake City
County	N	The county in which the deployed instrument collected data.	String	Salt Lake
Zip code	N	The zip code in which the deployed instrument collected data.	String (Numerical)	84102
Concept Name of the Location	N	A name representing the location where the deployed instrument collected data which is assigned by the system or taken from a reference terminology or ontology. This name can incorporate rows 18-22 combined as a single entity or examined independently.	String	USA_UT_SLC_UT_84102

Concept ID of the Location	N	An ID representing the location where the deployed instrument collected data which is assigned by the system or taken from a reference terminology or ontology. This name can incorporate rows 18-22 combined as a single entity or examined independently.	String (Numerical)	12345
Concept Namespace of the Location	N	The standard terminology/reference to be used as a reference for location data.	Category	GeoNames Ontology (http://www.geonames.org/)
Latitude	N	The latitude in which the organization resides.	String (Numerical)	40.76
Latitude Units	N	The direction of the parallels of latitude.	String	Degrees North
Longitude	N	The longitude in which the organization resides.	String (Numerical)	-111.863, 111.863
Longitude Units	N	The direction of the longitude meridian	String	Degrees West, West
Elevation	N	The height above sea level at which the data is collected	Value and Unit	3600 feet
Weather	N	The type of weather at the time when the data is collected.	Category	Sunny, Cloudy, Windy
Humidity	N	The humidity at the time when the data is collected.	Value and Unit	65%
Temperature	N	The temperature at the time when the data is collected	Value and Unit	70F, 30C

4.1.1 Conventions

- Each Deployment Output will have a unique Deployment Output ID.

4.1 Collected Sample

The Sample data element is a list of metadata elements that describe the characteristics of samples collected by the instrument and the procedure used for sample collection. The data elements are the same as the Sample Collection part. See “Sample collection” part for reference. The collected sample can be associated with existing biospecimen data models such as the OpenFurther’s biospecimen integration model and others (https://github.com/biobanking).