Smithsonian Institution
3D Metadata Overview, v0.6

A product of the Smithsonian's Digitization Program Advisory Committee's 3D Sub-Committee's Metadata Working Group

Working group members:
Jon Blundell | Lynda Schmitz Fuhrig | Holly Little | Suzanne Pilsk | Vince Rossi
Rebecca Snyder | Beth Stern | Ben Sullivan | Melinda Jane Tomerlin | Keats Webb

Vocabulary

- **Array** - A set of capture devices fixed to an armature, usually used to improve the resulting 3D model by leveraging the geometric relationships between the sensors as fixed by the armature or for instantaneous capture from multiple sensors.
- **Capture** - Another way to say this would be a 3D scan. It is the result of a single action of recording an object or scene into a digital form for processing into a 3D model.
- **Capture Data Element** - A capture data element is what is created from a single capture. For example: In photogrammetry, this would be an image (one photograph); in a CT scan, this would be a single slice; and for spherical laser scans, this would be a single point cloud.
- **Capture Dataset** - A logical collection of Captures which were collected during a during Capture Event which are intended to be processed as a group. A photogrammetry image set is an example of a type of Capture Dataset.
- **Capture Event** - Another way to say this would be a 3D scan session. It is the action of recording an object or scene into a digital form for processing into a 3D model
- **DPO** - Acronym for the Smithsonian's Digitization Program Office
- **Field ID** - A simple integer ID that is human readable and can be assigned in the field at the point of capture, usually starting at 1 and incrementing up from there. These IDs each a have a ‘resets’ parameter which denotes how to make a Field ID give a unique value within a project. For example, the ‘capture_sequence_number’ element "resets per Capture Device Configuration used with in a given Capture Dataset” This means with in a photogrammetry image set (Capture Dataset), each different camera (Capture Device Configuration) used should start incrementing ‘capture_sequence_number’ from 1.
- **Holding entity** - The person or organization which owns or is in charge of an object.
- **Item** - An object or collection of objects as that is digitized during a Capture Event.
- **Master Model** - Being the highest resolution output from processing, this model is a definitive version from which other downstream derivatives can be produced. When citing a model, the Master Model should always be used.
- **Record** - A set of metadata that describes a specific conceptual or physical item thing in the scan to mesh workflow.
- **Repository** - The “repository” as mentioned in the spreadsheet is the central 3D data Content Management System (CMS)/repository system which is being developed at the Smithsonian.
- **Subject** - An object or collection of objects as described in the system of record for that object(s)

Styling Key

This Document

- Capitalized - Denotes a specific use of a term, as defined above
- **Bold and capitalized** - Denotes a metadata record type

Spreadsheet Color codes

	Not fully defined, placeholder for a term that needs to be defined or vetted by working group
	Up for removal, information might need to be integrated somewhere else
	Something that needs attention, has an issue or uncertainty
	Heading for a separate record type

Spreadsheet columns

Name
Name of the element

Record type
The record type to which the element belongs

Category
A grouping of elements within a record type

Definition
Concise definition of the element

Metadata Type

Metadata Type	Definitions
Administrative	Facilitates both short-term and long-term management and processing of digital collections ● includes technical data on creation and quality control ● includes rights and reproduction management, access control and use requirements Pulled from: http://preservationtutorial.library.cornell.edu/metadata/table5-1.html
Descriptive	Describing and identifying information resources ● Gives context to a record ● Enables searching, discovering, and retrieving (e.g., searching for individuals or institutions involved in the creation of a record) Pulled from: http://preservationtutorial.library.cornell.edu/metadata/table5-1.html
Technical	Facilitates the manual and automated processing of datasets ● Hardware and software documentation ● Information surrounding capture event ● Technical digitization information, e.g., formats, compression ratios, scaling routines ● Documentation of post processing performed
Structural	Facilitates navigation and presentation of electronic resources ● provides information about the internal structure of resources, how different record types relate ● describes relationship among records (e.g., calibration image B was included in dataset A) ● binds the related files and scripts (e.g., File A is the JPEG format of the archival image File B) Pulled from: http://preservationtutorial.library.cornell.edu/metadata/table5-1.html

Source
Where does the information that populates an element originate from? Who is the source of the data?

- **Producer** - The producer is an entity responsible for capturing the scan data. The producer could refer to the organization or technician responsible for the capture.
- **Repository** - When the repository is the source, it means that the data element is an internal tracking field the repository creates and uses to link different Records together.
- **Subject Authority** - The entity which is the system of record for a Subject. This could be an organization or an individual, and could refer to a person or an authoritative CIS.
- **Device** - The element is populated from metadata taken directly from the data file created by the capture device.

Mandatory/Optional

- Mandatory
- Mandatory if Applicable
- Recommended
- Optional

Repeatable
The element can be used multiple time in a single record.

Variable Type
Specific allowable data type(s)

Data Type	Definitions
DateTime	Defines both a date and a time
Float	A number which can have a fractional value
Integer	A whole number (no fractions)
String	One or more characters

Element Type
Specific allowable data type(s)

Data Type	Definitions
Arbitrary text	Any arbitrary text
Controlled vocabulary, internally defined	Predefined terms that can be the only values of an element. "..." denotes a controlled vocabulary that is expected to expanded in the future
DateTime	Defines both a date and a time
External ID	An ID used by an external system of record
GUID	Persistent resolvable globally unique ID (Example ARC with a UUID tail)
Integer id	A human readable ID that can be assigned at the time of capture and built into the filename, so it can be applied in the field before entry into repository
Repository UID	ID unique to the database/repository. An internal reference.

Controlled vocabulary values:
The predefined terms that can be the only values of an element. "..." denotes a controlled vocabulary that is expected to expanded in the future

Example values:
Examples of values an element could contain.

Record Types

Project Record

Project Records are one of the two top level record types in this data model, along with **Subject Records**.

A **Project Record** represents a digitization effort with a clearly defined scope and set of goals. For the DPO, this is usually a collaborative effort between the DPO and one of the Smithsonian museums, galleries, or research centers. A project can be organized around the scanning of a single object, a collection of objects, a research site, etc. A **Project Record** is a way to provide context around the collection of datasets and models.

Examples of projects undertaken by the DPO would include:

- The NMNH Paleo pilot project where 118 items were scanned over a one week period
- The NMNH Nation's T. rex project where each bone was scanned individually processed over the course of 3 months, and then the full articulated skeleton was later digitized as a single object
- The FSG Cosmological Buddha project where a single sculpture was scanned

A **Project Record** must have one or more **Capture Dataset Records** and/or **Model Records** pointing to it.

Subject Record

A **Subject Record is** one of the two top level record types in this data model, along with **Project Records**.

A **Subject Record** is a reference to a conceptual thing that is digitized during the course of a Project. A **Subject Record** can describe a discrete object, collection of objects, or an environment. A **Subject Record** is mostly a pointer to the collection/catalogue record in the collection information system (CIS) of the entity which holds the Subject. For environments, a Subject is one discrete, identifiable area. A **Project Record** is a way to connect datasets and models with the entry in that Subject's system of record.

Examples of Subjects from DPO projects include:

- An articulated woolly mammoth skeleton
- A porcelain vase with lid
- The Cosmological Buddha statue
- The Apollo 11 command module
- A segment of the Qorikancha wall in Cuzco wall in Peru

A **Subject Record** must have one or more **Item Records** pointing to it.

** This schema is designed from an institutional perspective and assumes that there is a separate system of record for any Subject scanned*

Item Record

An **Item Record** describes a discrete physical component of a Subject that was scanned during the course of a Project. While a **Subject Record** represents the concept of a thing, an **Item Record** represents a physical manifestation of that thing as it was scanned. A **Subject Record** might only have one **Item Record** associated with it, or it might have many **Item Records** associated with it. An Item could be the single object that is the Subject, a single object that is part of a Subject, or a particular and intentional arrangement of multiple objects that make up a Subject.

For environments, an **Item Record** is a discrete logical section of the environment, which could encompass the entire area of the Subject or just a portion.

Examples of **Subject Records** with only one **Item Record** from DPO projects include:

- The Cosmological Buddha statue
- One of Lincoln's life masks

Examples of **Subject Records** with multiple **Item Records** from DPO projects include:

- Nation's T. rex.
    - Each bone of the Nation's T. rex was scanned individually, and each would be represented by an **Item Record**
    - The fully articulated skeleton was scanned, this would be represented by its own separate **Item Record**
- A porcelain vase with lid. Each of the following could be an Item:
    - The lid
    - The jar
    - The jar with lid on
- An example of a location with multiple **Item Records** would be the the Jamestown chancel burials
    - Each individual burial would have an **Item Record**
    - The overall excavation would have an **Item Record**
    - The overall site scan would have an **Item Record**

**the line that separates a Subject and Item is clearly subjective, and mostly relies on how the Subject is defined in its system of record*

An **Item Record** must always point to only one **Subject Record. Item Records** are required to have at least one **Capture Dataset Record** or **Model Record** point to them.

Capture Dataset Rights Record

Capture Dataset Rights Records relate to the rights of a particular Dataset as represented by a **Capture Dataset Record** and not the actual Subject or Item.

Examples of a **Capture Dataset Rights Record**:

- Copyright
- Culturally sensitive
- Embargoed for research

A **Capture Dataset Rights Record** must be linked to only one **Capture Dataset Record**. One **Capture Dataset Record** can have multiple **Capture Dataset Rights Records** associated with it.

Capture Dataset Record

A **Capture Dataset Record** gives context and technical information around the capture event during which capture data was gathered, allowing all information contained within this record to be filled out at time of capture.

A **Capture Dataset Record** represents a logical collection of Capture Data Elements which share the following criteria: collected during a given Project, pertaining to usually one Item, collected using the same equipment/method, and are intended to be processed together. One way to define what constitutes a 'logical collection' of Capture Data Elements is a grouping where all values in the containing **Capture Dataset Record** must be true for the Capture Data Elements contained within. This logical grouping can be further divided into sets for logistical processing and other logistical reasons, which should be clearly demonstrated/explained in the 'image_set_name' and 'image_set_description' elements. A common example would be splitting up photogrammetry images into different sets based on the area of a Subject they cover (ie the left wing of an airplane, a wall in a room, etc).

A photogrammetry image set is an example of a type of Capture Dataset. Thus, for the 'logical collection' of Capture Data Elements mentioned above, a photogrammetry Capture Dataset would comprise images which only provide 3D data in relation to each other and are meant to be processed together in photogrammetry software with minimal effort.

A **Capture Dataset Record** must point to one **Item Record** and one **Project Record**. It must have one or more **Capture Data Elements** point to it. A **Capture Dataset Record** may also link to one or more other **Capture Dataset Record** via the linking records 'resource_capture_dataset' field**.**

Context Category

The context category gives broad frame of reference information about a Capture Dataset.

Positioning Relationships Category

The positioning relationships category is used to determine what Datasets can be processed together, and to some extent how they should be processed. The crux of this category is that, for Datasets to be processed together, the geometric arrangement (shape and position) of an object or environment being captured cannot change. If there has been a change, extra processing considerations must be taken into account. Another way to say this is, if an object moves or deforms during capture, files from before and after the change must be part of a different dataset, and cannot easily be processed together.

It should be noted, for processing relationships to be applicable between two photogrammetry image sets, both sets must have the same parent Project and parent Item.

Photogrammetry Category

The photogrammetry category contains technical information that is specifically relevant to photogrammetry or image based captures.

Cluster Information Category

Cluster information describes any armature/apparatus/motion control system used to position or manipulate one or more capture devices.

Resources Category

This points to resources that can be used to assist in processing of a given Dataset.

Capture Data Element Record

A **Dataset Element Record** stores capture event information much like a **Capture Dataset Record**, except the metadata applies only to an individual capture instead of a grouping of captures. Thus, all information contained within this record should be able to be filled out at time of capture.

For example: In a photogrammetry image set, a Capture Data Element would be an image; in a CT scan, this would be a single slice; and for spherical laser scans, this would be a single point cloud.

*In some types of capture with complex, proprietary data structures, it might not be possible to break out individual Data Elements to the degree above. In this case, the entire project file or folder would be represented by the **Dataset Element Record** with more limited metadata elements.

A **Capture Data Element Record** must point to one **Capture Dataset Record**. It must have one or more **Capture Data File Records** point to it. A **Capture Dataset Element Record** may also link to one or more **Capture Datasets Records** via the linking records 'resource_capture_data_element' field**.**

Position Information Category

Position information is used to record the relative position of a sensor/capture devices. It can be used to relate Capture Data Elements together when they share a geometrically significant relationship.

Capture Data File Record

A **Capture Data File Record** is the lowest level of granularity for recording capture event metadata and represents the individual files which make up a Capture Data Element.

For example: a photogrammetry image (the Capture Data Element) could be represented by a TIF, JPG, DNG, or other raw file format; a CT scan slice could be represented by a DICOM or TIF; and a point cloud could be represented by a PTX, PTS, FLV, etc...
*In some types of capture with complex, proprietary data structures, it might not be advisable to break out individual files; in these cases the entire project folder would be placed in a single archive file for storage.

A **Capture Data File Record** must point to one **Capture Data Element Record**.

Model Record

A **Model Record** contains information related to an individual geometry file. This record stores descriptive information as well as provenance information.

A **Model Record** should point to one or more **Capture Dataset Records** or other **Model Records**. If the model has color or other information stored in a UV map, it should have one or more **UV Map Records** pointing to it.

Point Cloud Attributes Category

These are attributes that apply specifically to point clouds.

Mesh Attributes Category

These are attributes that apply specifically to meshes.

UV Map Record

A **UV Map Record** contains information related to an individual UV map file and is used to associate UV maps with the correct model.

A **UV Map Record** must point to one **Model Record.**

Processing Action Record

Processing Action Records are used to capture processing events that contribute to the creation of a given model.

*This record type is still being worked out and is likely to change in the future.

A **Processing Action Record** must point to one **Model Record**.

Capture Device Configuration Record

A **Capture Device Record** represents a specific configuration of the equipment that was used to capture Data Elements/Data Files during a specific project.

*This record type is still being worked out and might change in the future.

A **Capture Device Record** should have one of more **Capture Data Element Records** pointing to it.

Capture Device Component Record

A **Capture Device Component Record** represents the components that make up a capture device. For example, in photogrammetry, Capture Device Components would most likely be camera bodies and camera lenses.

A **Capture Device Component Record** should have one or more **Capture Device Records** pointing to it.

Actor Record

An **Actor Record** contains information related to an individual or organization which took part in a capture or model creation.

An **Actor Record** should have one or more **Capture Dataset Records** or **Processing Action Records** pointing to it.

Physical Scale Bar Record

A Physical Scale Bar Record is used to associate a scale bar with the Capture Datasets it was used in.

A **Physical Scale Bar Record** should have one of more **Capture Dataset Records** pointing to it and one or more **Scale Bar Target Pair Records** pointing to it.

Scale Bar Target Pair Record

A **Scale Bar Target Pair Record** contains the information that describes a pair of targets on a physical scale bar. This includes information on how to identify the targets and the distance between the targets.

A **Scale Bar Target Pair Record** must point to one **Physical Scale Bar Record**.