Overview of the CADF Event Standard
including Event Model and Format

December 2017

Presented by:

Matt Rutkowski, IBM Cloud Open Tech.
PMC co-chair, Committer, Apache OpenWhisk

Disclaimer

“I don't want to sell anything, buy anything, or process anything as a career. I don't want to sell anything bought or processed, or buy anything sold or processed, or process anything sold, bought, or processed, or repair anything sold, bought, or processed. You know, as a career, I don't want to do that.”

- Lloyd Dobler, Say Anything 1989

Topics

- What is CADF ?
- Supported Use Cases
- Event Model and Format
- Resources
- Summary

The Cloud Auditing Data Federation (CADF) Open Standard

Goals

- **Cloud** *enabled*
  - *Inclusive of all deployment models (SaaS, PaaS, IaaS, and FaaS) and hosting options (Private, Public, Hybrid)*
  - *Supports event driven Workflows, Tool chains, Dev-Ops collaboration & integration*
- **Audit** *ready*
  - *Enables Customized Analysis for customer compliance needs*
    - *adherence to regional, industry or corporate policies, SLAs*
- **Data** *aware*
  - *Clear Guidance on how to Normalize and Categorize event data*
  - *Supports Workflows, Tool chains, Dev-Ops collaboration & integration, change management*
- **Federation** *is a must*
  - *Aggregation and correlation of data from any service provider or any **hybrid** source*

Designed & Supported by Event Experts

- Development supported by DMTF member banks, insurance companies, etc.
- Presented to and endorsed by ISO 27000 SCs & Aligned with IETF standards

Open Source assures conformance

- Libraries available in several languages *(Python, Java and GoLang)*
- *Used by many Cloud open source services (especially OpenStack)*
  - *open source implementations as part of API WSGI frameworks available as well (normalization)*

Original Contributors

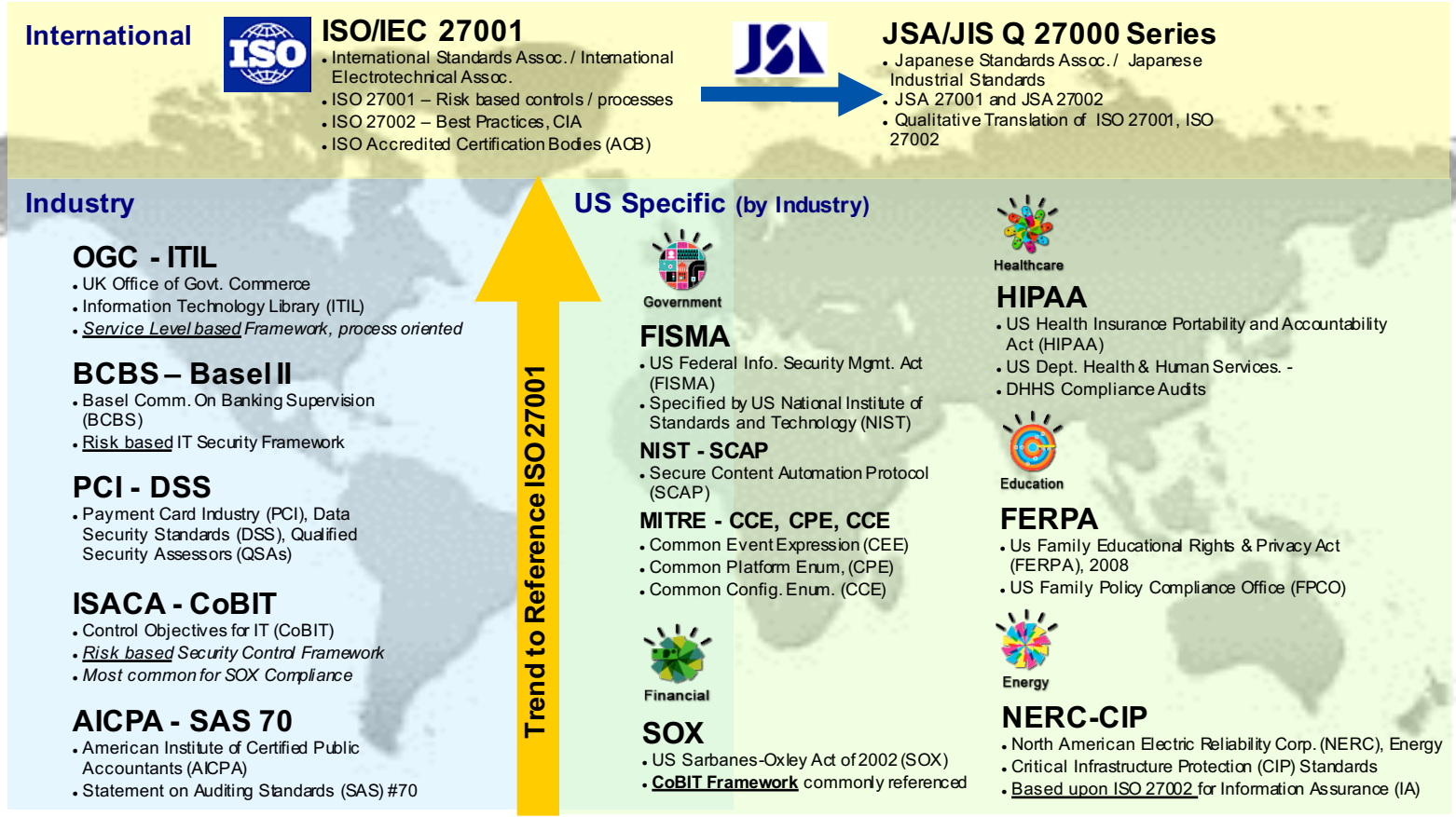
Designed for many event-centric use cases

- **Efficient - High Volume / Minimal footprint**
  - **JSON** format, few required fields
  - designed to accommodate high volume, fast processing (millions of events per minute)
- **Data Compliance**
  - *ISO, HIPPA, FISMA, SOX, PCI, etc.*
- **Data Correlation**
  - based on time (ISO), geolocation (ISO), resources classifications (actors), user-defined tags
  - in Workflows, DevOps tool chains
- **Metric & Measurement Events**
  - describe how your data was measured (e.g., IoT / Sensors / Resource monitors)
- **Considers Complex Events**
  - multiple targets, multiple observers, summarized events, aggregated events
- **Analytics**
  - high value in normalized event data…
  - Search / Query, Real-time
- **More…**

Highly Extensible

- **Define new Event Types**
- **Add additional Keys (maps)**
  - for your type / domain
- **Extend Action, Outcome and Resource Classification Taxonomies**
- **Include Metric & Measurement Data**
- **URI Tagging**
  - Orthogonal classification / search
- **Attachments**
  - Include additional data perhaps in another format
  - include the “original” (origination) event

FaaS Providers need to consider Compliance scenarios

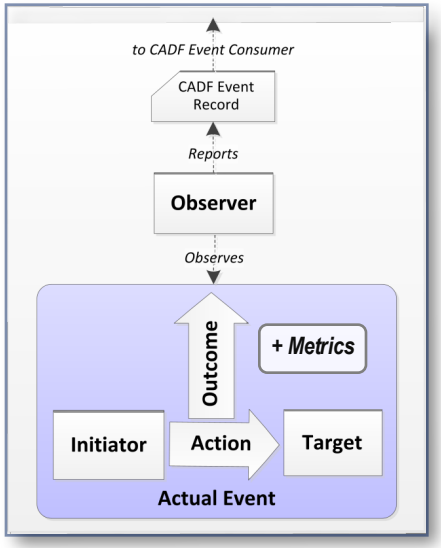


*"as cloud-based services increasingly handle sensitive data as part of cloud-inclusive workflows, the **importance of interoperability and compliance** become more evident."*

Event Model & Format

CADF Event Model – Common to all CADF Event Types

Conceptual Model

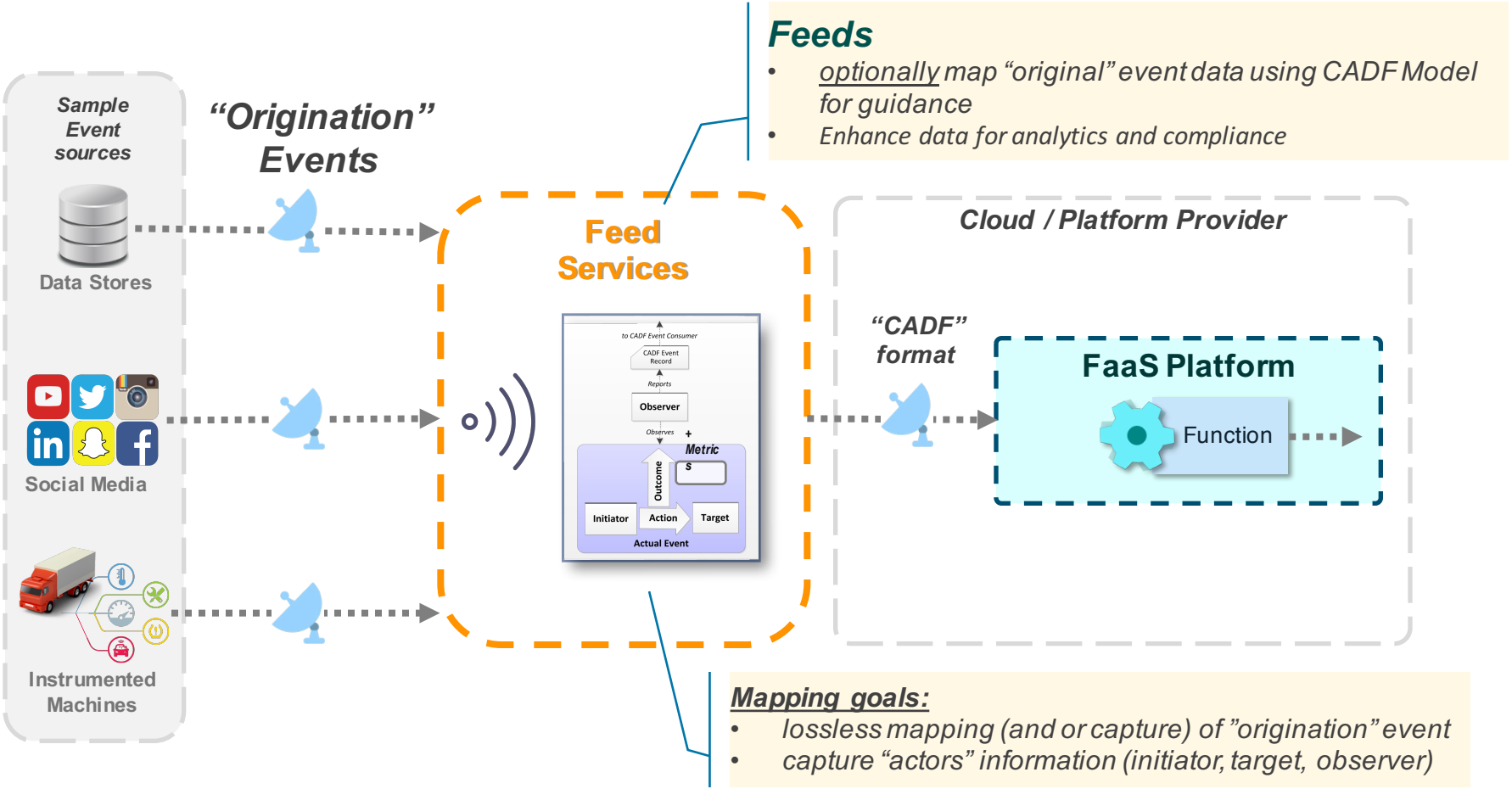


Model Components

Component	Definition
OBSERVER	The RESOURCE that generates the CADF Event Record based on its observation (directly or indirectly) of the Actual Event.
INITIATOR	The RESOURCE that initiated, originated, or instigated the event's ACTION, according to the OBSERVER.
ACTION	The operation or activity the INITIATOR has performed, attempted to perform or has pending against the event's TARGET, according to the OBSERVER
TARGET	The RESOURCE against which the ACTION of a CADF Event Record was performed, was attempted, or is pending, according to the OBSERVER.
OUTCOME	The result or status of the ACTION against the TARGET, according to the OBSERVER.
Metrics	Optionally, include Metrics and Measurements

- **Initiator, Target and Observer "Actors" interact and are represented in the model**
  - **"Categorized"** using extensible CADF Taxonomies, as are Actions & Outcome values
  - **"Geolocation aware"** using ISO 6709, ICANN values
- **Model and Specification are extensible**
  - **Key Maps –** add new keys (maps)
  - **Tagging** - events to create domain-specific views on data
  - **Attachments –** encode / attach larger structure or unstructured data

Model helps map "Origination" Event data to Normalized CADF



CADF Event Model provides the "7W" standard for Clouds

- Supports Activity, Monitoring (Metric) or Control Event Types
  - Clear Guidance to normatively record *Basic*, *Detailed* or *Precise* information for each event component
  - Activity, Monitoring (Metric) or Control

1 What

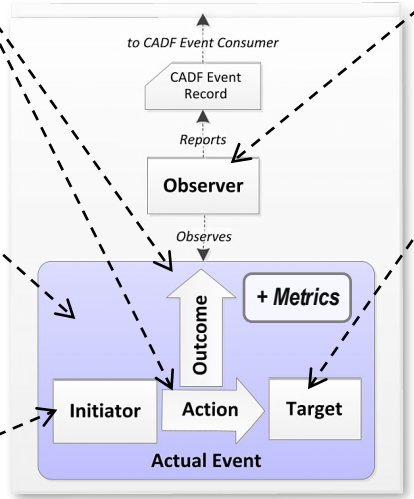
What activity occurred?; What the result?
- Event Type (activity, monitoring, control)
- CADF Action classification (basic)
- CADF Outcome Classification (basic)
- Reason codes (detailed)

2 When

When did the Action happen? When was it observed? How long did it take?
- **ISO 8601 Timestamp** with fractional sections (basic or precise)
- Timezones (detailed)
- Duration for long running events (detailed)

3 Who

User / service that initiated the Action
- Initiator identifer, name (basic)
- CADF Resource classification (basic)
- Credentials (detailed)
- Identity assertions (precise)



4 Where - was the Action observed? What role does the event serve? How was it recorded?
- Observer identifer, name (basic)
- CADF Resource classification (basic)
- Role, Reporting Steps (detailed, precise)

5 On What - resource was the Action targeted?
- Target identifer, name (basic)
- CADF Resource classification (basic)

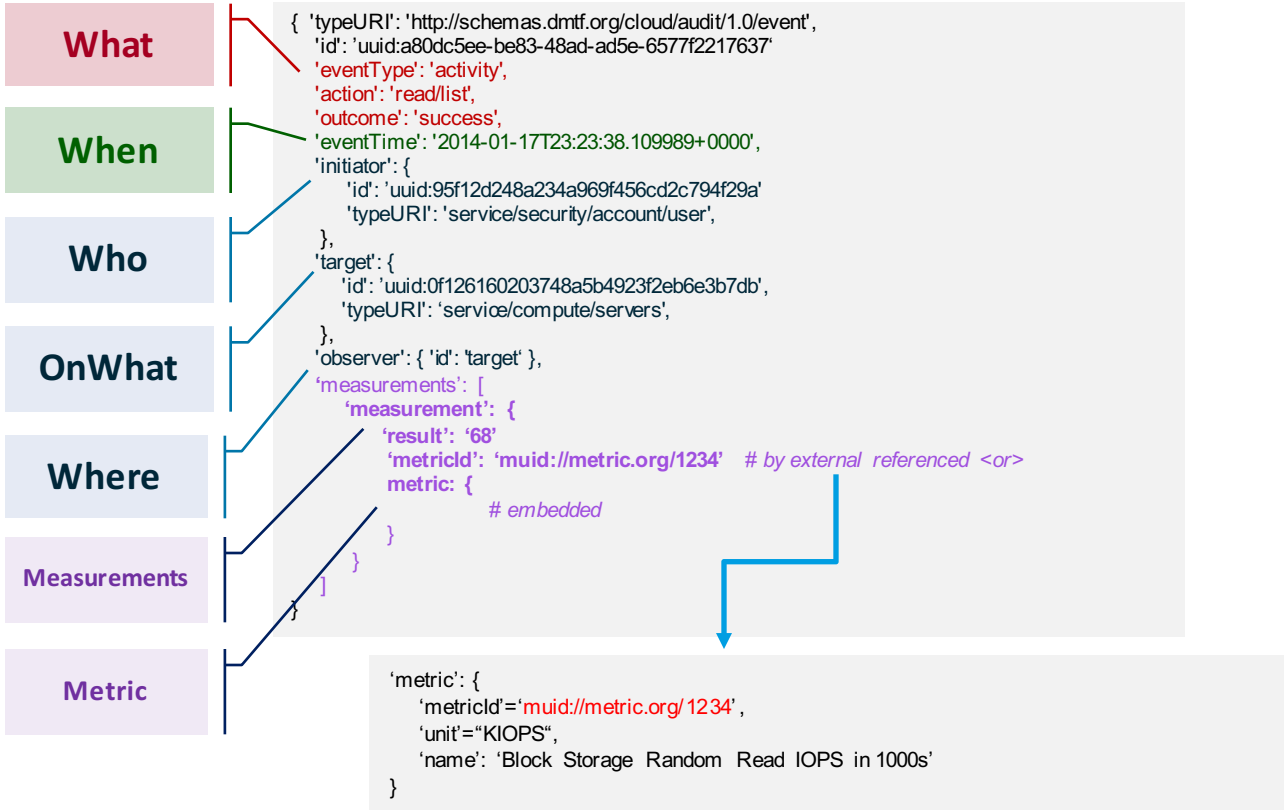
6 FromWhere - was the Activity Initiated?
- Network addresses (basic)
- Host information (agents, platforms, etc.) (detailed)
- **ISO 6709 Geolocation**, ICANN codes (precise)

7 ToWhere – was the Activity targeted?
- detailed information on the target's location.
- Similar to "FromWhere"

** The 7 W's are filled out using the perspective of the **OBSERVER** component resource*

Minimalist CADF Event Record *(only required fields)*

Results in only 5 of 7 Ws for auditing with optional Metric & Measurement data



Resources

Documentation

- **DMTF**: https://www.dmtf.org/standards/cadf
- Specifications, Profiles, White papers, presentations

Libraries *(CADF Event creation & validation)*

pyCADF: **https://pypi.python.org/pypi/pycadf**

- PyPi (pypi.python.org), Python 2 and 3 compatible
- *used by all OpenStack components*
- *JSON format proven to handle millions of events per second*

jCADF: **https://developer.ibm.com/open/jcadf/**

goCADF: **https://github.com/ibmalchemy/cadf**

Summary

- **Open Standard with Open Source**
  - Clear guidance on how to normalize/map data from Origination format
  - Many use cases with examples from various sources
- **Extensible**
  - Define your own **Event Type**
  - Add your own **Keys / Key maps**
  - Use **Tags** and **Annotations**
  - Extend resource taxonomy to describe event actors
- **Compliance Ready**
  - **ISO Compliant**
  - FaaS will have interact with user/customer data with various compliance considerations
- **Please Consider Embracing and Extending CADF**
  - Possibility to publish extensions as a Profile @ DMTF
    - *as "Informational" or "Specification" Profile*
- **Happy to help in any way !**
  - *e.g., work on use cases and show how to map*

Questions ?

Additional Model Components: Measurement & Metric Data

for the "monitor" Event Type

Event Component	CADF Definition
MEASUREMENT	An entity that contains statistical or measurement information for the TARGET resource(s) that are being monitored based upon a well-defined Metric.

Measurement – *The value generated by the application of a Metric*

Property	Description
result	The quantitative or qualitative result of a measurement from applying the associated metric. The measurement value could be boolean, integer, double, a scalar value, etc.
metric	The property defines the in-line metric used in generating the measurement result.
metricId	This property identifies an existing CADF Metric definition whose definition exists outside the event record itself.
calculatedBy	An optional description of the resource that calculated the measurement

Metric – *Description of the rules and processes for measuring some activity or resource*

Property	Description
metricId	The identifier for the metric (allows reuse) Metric data is designed so that it can be described once, for example in the context of a CADF Log, and referenced by the multiple CADF Event (records) the log contains..
unit	The metrics unit (e.g., "msec.", "Hz", "GB", etc.)
name	A descriptive name for metric (e.g., "Response Time in Milliseconds", "Storage Capacity in Gigabytes", etc.)
annotations	User-defined metric information.

Optionally, metrics may be provided on "activity" or "control" types

Additional Model Components: Geolocation Data

Unambiguous, standardized Geolocation of event location

Property	Description
id	Optional identifier for a geolocation.
latitude	Indicates the latitude of a geolocation. Geolocation MAY be provided in a pair of latitude and longitude. Latitude values adhere to the format based on **ISO 6709:2008**
longitude	Indicates the longitude of a geolocation. Geolocation MAY be provided in a pair of latitude and longitude. Longitude values adhere to the format based on **ISO 6709:2008**
elevation	Indicates the elevation of a geolocation in meters. Elevation at or above the sea level shall be designated using a plus sign (+), or no sign. Elevation below the sea level shall be designated using a minus sign (−).
accuracy	Indicates the accuracy of a geolocation in meters. Geolocation expresses the resource location to a reasonable degree of accuracy.
city	Indicates the city of a geolocation.
state	Indicates the state/province of a geolocation
regionICANN	Indicates a region (e.g., a country, a sovereign state, a dependent territory or a special area of geographical interest) of a geolocation.
annotations	Indicates user-defined geolocation information (e.g., building name, room number).

CADF's Powerful Path-Based, Extensible Taxonomies

CADF defines three taxonomies designed to provide the basis for a domain extensible, path-based mechanism to name resources, actions and that appear in audit events in order to enable normative classification and query of events data.

1. **CADF Resource Taxonomy**
   - Normalized classification type "names" for the resource types that participate on an event (e.g. INITIATOR, TARGET, OBSERVER)
   - *Enables Resource-based Query* by type of resource

2. **CADF Action Taxonomy**
   - Normalized names used to describe actions or activities performed on resources
   - *Enables Activity-based Query*

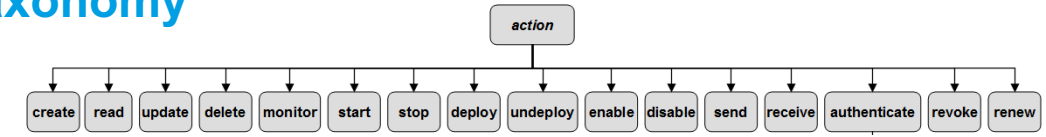
3. **CADF Outcome Taxonomy**
   - Normalize the names used to describe outcomes of activities
   - *Enables Outcome-based Query*

CADF Taxonomies Absolute, Versioned URIs:

Taxonomy Name	Taxonomy URI
resource	http://schemas.dmtf.org/cloud/audit/1.0/taxonomy/resource/
action	http://schemas.dmtf.org/cloud/audit/1.0/taxonomy/action/
outcome	http://schemas.dmtf.org/cloud/audit/1.0/taxonomy/outcome/

All taxonomies are extensible for any domain-specific compliance framework

CADF Action Taxonomy



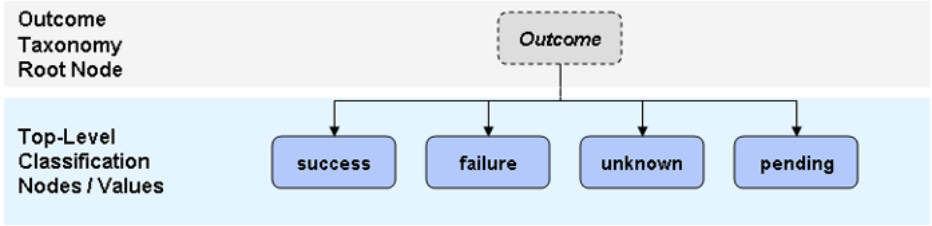
Value	Description
create	The target resource described in the event was created (or an attempt was made to do so) by the initiator resource.
read	Data was read from the target resource by the initiating resource (or an attempt was made to do so).
update	One or more of the target resource's properties were modified or changed by the initiator resource.
delete	The target resource described in the event was deleted (or an attempt was made to do so) by the initiator resource.
backup	The target resource described in the event is being persisted to storage without regard to environment, context or state at the time of storage.
capture	The target resource described in the event is being persisted to storage along with relevant environment and state information (e.g. program settings, network state, memory/cache, etc.). Conceptually, a "snapshot" of the resource is being captured at a moment in time.
configure	The target resource described in the event is being set-up to enable it to run on a particular environment or for a particular application or use.
deploy	The target resource is being positioned or made available for use by the initiator resource, but not yet started.
disable	The initiator resource is causing the target resource [that has been started] to disallow or block some set of functions.
enable	The target resource (that has been started) is being changed by the initiator resource to allow or permit some set of functions.
monitor	The target resource is the subject of a monitoring action from the initiating resource.
restore	The initiator is requesting the target resource (or some portion of it) be restored from persistent storage.
start	The target resource is being made functional by the initiator resource and able to perform or execute operations.
stop	The initiator resource is causing the target resource to no longer be functional or able to perform or execute operations.
undeploy	The initiator resource is causing the target resource to no longer be positioned or available for use.
receive	The initiator resource is receiving a message or data from the target resource.
send	The initiator resource is transmitting a message or data to the target resource.
authenticate	A security request used to establish an initiator's identity and/or credentials to the target resource against a trusted authority.
renew	A security request from the initiator resource to renew a resource's identity, credentials, or related attributes or privileges sent to the target resource (an authority).
revoke	A security request from the initiator resource to remove entitlements or privileges from a resource's identity and/or credentials sent to the target resource.
allow	Indicates that the initiating resource has allowed access to the target resource.
deny	Indicates that the initiating resource has denied access to the target resource.
evaluate	The evaluation or application of a policy, rule, or algorithm to a set of inputs.
notify	Indicates that the initiating resource has sent a notification based on some policy or algorithm application – perhaps an alert to indicate a system problem.
unknown	Indicates that the OBSERVER of the event is not, to the best of its ability, able to classify the exact action for the actual event it is reporting.

Color Key

General resource management (e.g. CRUD)
Workload and data management
Messaging
Security - Identity
Security - Policy

19

CADF Outcome Taxonomy (*with optional Reason data)*

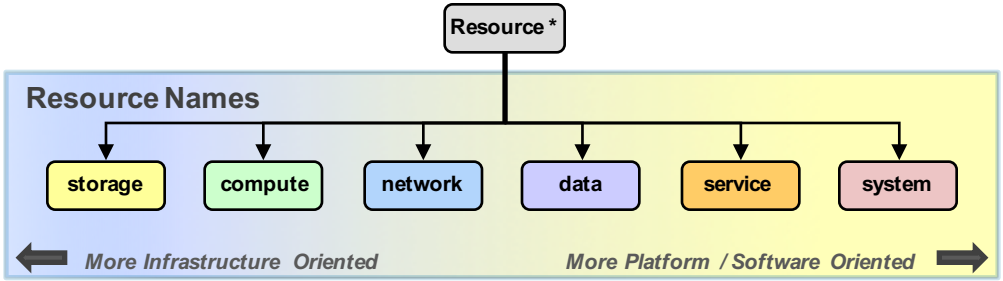


Value	Description
success	The attempted action completed successfully with the expected results.
failure	The attempted action failed due to some form of operational system failure or because the action was denied, blocked or refused in some way.
unknown	The outcome of the attempted action is unknown and it is not expected that it will ever be known.
pending	The outcome of the attempted action is unknown, but it is expected that it will be known at some point in the future. A future event correlated with the current event may provide additional detail.

Reason data type properties *For additional domain specific information, reasons, or codes that enable further diagnostics*

Property	Description
reasonType	The domain URI that defines the "reasonCode" property's value. See examples below.
reasonCode	An optional detailed result code as described by the domain identified in the "reasonType" property. Note: The “reasonCode” should in general indicate what type of policy was violated for its associated domain.
policyType	The domain URI that defines the “policyId” property’s value. See examples below.
policyId	An optional identifier that indicates which policy or algorithm was applied in order to achieve the described OUTCOME.

CADF Resource Taxonomy : Top-Level Logical Resource Classifications



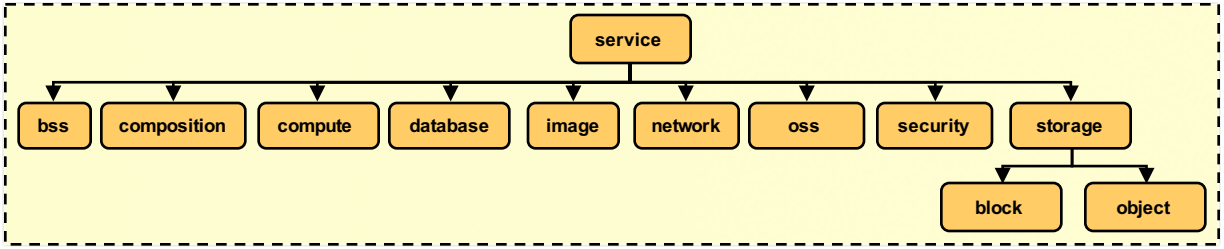
Name	Description
storage	Logical resources that represent storage containers
compute	Logical resources that are used to perform logical operations or calculations on data
network	Logical resources that interconnect computer systems, terminals, and other equipment allowing information to be exchanged.
data	Logical named sets of information (objectified data) that are referenced and managed by services.
service	Logical set of operations, packaged into a single entity, that provides access to and management of cloud resources (for a given domain).
system	Logical resources that are a combination of several other [cloud] resources that operate as a functional whole, this combination being manageable (created, operated, audited, etc.) as a unit i.e. offering some operations that could activate lower-level operations over each of the sub-resources.
unknown	Indicates that the OBSERVER of the event is not, to the best of its ability, able to classify a resource that contributed to the actual event it is reporting on using any other valid resource taxonomy value. Note: This value SHOULD only be used as a last resort, and when using another classification value from the CADF Resource Taxonomy is not possible.

* The name value "resource" (tree root) implies the absolute absolute name:

- **"http://schemas.dmtf.org/cloud/audit/1.0/taxonomy/resource/"**

21

CADF Resource Taxonomy - Service subtree (example)



Name	Descriptive Name	Description
bss	*Business Support Services (BSS)*	The logical classification grouping for services that are identified to support business activities.
composition	*N/A*	The logical classification grouping for services that supports the compositing of independent services into a new service offering
compute	*N/A*	Infrastructure services for managing computing (fabric).
database	*Database Services (or DB-as-a-Service)*	Database services that permit substitutability to various provider implementations.
image	*N/A*	Infrastructure services for managing virtual machine images and associated metadata.
network	*N/A*	Infrastructure services for managing networking (fabric).
oss	*Operational Support Services (OSS)*	The logical classification grouping for services that are identified to support operations including communication, control, analysis, etc.
security	*Security Services (or Sec-as-a-Service)*	The logical classification grouping for security services including Identity Mgmt., Policy Mgmt., Authentication, Authorization, Access Mgmt., etc. (a.k.a. “Security-as-a-Service”)
storage	*N/A*	Infrastructure services for managing storage (fabric).
storage/block	*N/A*	Infrastructure services for managing Block storage.
storage/object	*N/A*	Infrastructure services for managing Object storage.