# Background

When evaluating a services SAML entity metadata for publishing in the InCommon metadata, the InCommon Federation enforces several validation rules on the entity's entity ID:

1. The entity ID meets the syntax requirements described in SAML specification
2. The entity ID meets the syntax requirements described in SAML Deployment Profile for Federation Interoperability
3. Uniqueness - the entity ID is not already registered in the InCommon metadata
4. (do we check this?) Uniqueness - the entity ID is not already published in the eduGAIN metadata
5. The entity ID is a Uniform Resource Locator (URL)
6. The registrant must demonstrate control of the domain used in the entity ID's URL.

Of these rules, 5 and 6 are InCommon-specific and are not required by SAML or other international standards. These additional rules are meant to ensure an entity ID is unique across our federation space (InCommon, eduGAIN), and that the entity ID doesn't create a perception that it represents a different organization when it shouldn't (e.g., Stanford University registering an entity ID with a domain name of arizona.edu, creating the perception that the entity represents University of Arizona).

## Challenge

As Participants increasingly adopt cloud-based, commercial IAM solutions, we are encountering products that issue SAML entity ID programmatically (i.e., the customer cannot customize the entity ID) under the vendor's domain, typically in the form of <vendor domain>/<customer-id>. These IDs meet SAML specifications and are globally unique. However, they cannot meet InCommon's proof of domain control requirements. We know at least 3 major vendors who programmatically assign entity IDs: Microsoft (Azure AD), Oracle (Oracle Identity Manager), and Amazon (Amazon Cognito).

Cloud IAM solutions frequently have other integration gaps preventing them from seamlessly interoperating in Federation. Thus far, we have used our entity ID requirements as a shield of sorts to prevent services using these products from automatically registering. However, the conversation that follows frequently gets stuck on entity ID choices rather than the other gaps. We don't get the chance to really address those gaps with Participants. Worse, given this is often the first real technical interaction a new Participant has with InCommon, we risk creating a perception that InCommon is enforcing arbitrary, non-standard rules and being unnecessarily difficult. If that is their first impression of us, we face an uphill battle when we try to convince them to meet any other federation-required interoperability standards.

## Change

There is a way to accommodate products that programmatically assign entity IDs, meet our goal to curb entity ID misuse and to create opportunities to help Participants understand integration gaps early in their onboarding. The rest of this document describes a new entity ID validation process:

# Proposed Entity ID Validation Rules

## Create Opportunity to Alert Registrant of Additional Integration Gaps

*This is new.*

Maintain documentation of known integration gaps for known commercial IAM solutions.

On metadata entry, if the registrant enters a known programmatically created entity ID (e.g., one issued by Azure AD):

- Display warning message to alert the registrant that the integration requires additional work beyond metadata registration; point them to relevant documentation; suggest alternatives
- Alert them of potential consequences of using a vendor specific entity ID (portability, etc)
- Require acknowledgement from registrant
- Present follow up support discussion opportunities.

## SAML and SAML Deployment Profile Conformance

*Continue current practice.*

On metadata entry in Federation Manager, perform syntax check to make sure the submitted entity ID conforms to the syntax requirements defined in **[SAMLMetadata]** (section 2.2.1), **[SAMLCore]** (section 8.3.6), and **[SAML2Int]** ([SDP-G04]).

Note: **[SAML2Int]** specifies a shorter string size than **[SAMLMetadata]**

If entered value does not meet requirements, display appropriate error message explaining how the entered value fails to meet requirement. Disable metadata submission until error is corrected.

## Entity ID Uniqueness

*This may be new.*

On metadata entry:

- Compare entered value against all published entity IDs in InCommon metadata

- Compare entered value against all published entity IDs in eduGAIN metadata

If the entered value is found in InCommon metadata, display appropriate error message explaining value is already in use. Disable metadata submission until error is corrected.

If the entered value is found in eduGAIN metadata (other than InCommon published metadata), display appropriate error message explaining value is already in use. (Would we block an entity from being registered if the entity ID appears in another federation's metadata?)

## Accept URI

This is a change from current behavior.

Accept all valid entityID as defined by **[SAMLMetadata]** and **[SAML2Int]**. Neither specifies the entity ID needs to be a URL.

## Domain Validation

*This is a change from current behavior.*

On metadata entry:

Compare the domain used in entered entity ID against all published, registered Scope in InCommon and eduGAIN metadata.

If the domain used matches a Scope registered by a different organization other than the registrant's organization:

- Display a warning message indicating a potential domain use violation;
- Route the metadata submission to RA for manual triage / validation

TODO: we have work to do to define what a "domain" is in entity ID, e.g.,

https://dmv.ca.gov/idp
https://www.ucla.edu/idp

What is the "domain" in each?

## Follow Up Domain Validation

To intercept a party who may have registered an entity using a domain previously unknown in federation space, on IdP metadata registration and successful completion of domain control validation of the Scope(s) listed in that metadata:

Scan published (and potentially any unpublished ones in edit) InCommon metadata to determine whether the domain is referenced in an entity ID not registered by the Scope owner.

If found, alert the Scope owner of the use; seek authorization. If Scope owner does not consent, notify the entity ID registrant of domain use violation; require remediation. Eventually remove metadata from Federation if registrant fails to remediate.

NOTE: might need to handle violations across eduGAIN space as well… TBD

## References


**[SAMLMetadata]** Metadata for the OASIS Security Assertion Markup Language (SAML) V2.0, https://docs.oasis-open.org/security/saml/v2.0/saml-metadata-2.0-os.pdf

**[SAMLCore]** Assertions and Protocols for hte OASIS Security Assertion Markup Language (SAML) V2.0, https://docs.oasis-open.org/security/saml/v2.0/saml-core-2.0-os.pdf

**[SAML2Int]** SAML V2.0 Deployment Profile for Federation Interoperability, https://kantarainitiative.github.io/SAMLprofiles/saml2int.html

**[EntityID]** InCommon Federation Wiki: Entity ID, https://spaces.at.internet2.edu/display/federation/saml-metadata-entityid

The road to hell is paved with SAML Assertions
https://www.economyofmechanism.com/office365-authbypass.html