

2020-12-07 DM - Metadata duplicate management

Date

07 Dec 2020

Attendees

- [Mandy Chessell](#)

Goals

- Review the proposal for upgrading the existing duplicate detection support in the discovery server to cover

Discussion items

Time	Item	Who	Notes
10min	Introductions	Maryna	<ul style="list-style-type: none">• Overview of 3 days
40mins	Modelling duplicates	Mandy	<ul style="list-style-type: none">• Review open types and their uses
50mins	Engine Services	Mandy	<ul style="list-style-type: none">• Review engine services that replace the Discovery Server and Stewardship Server and provide support for:<ul style="list-style-type: none">◦ running automated metadata discovery◦ monitoring metadata (eg detecting duplicates)◦ triaging issues◦ enacting remediation (fixes)
20 mins	Next Steps	All	Discussion

Notes

De-Duplication:

- Issues with current model:
 - Multiple IGCs synced with Egeria.
 - Objects and their relationships change over time.
 - currently only for assets.
- New approach in stewardship and discovery.
- record fingerprint of metadata, for detection in discovery.
- Use OMASs for previously defined elements.
- Use `KnownDuplicateLink/KnownDuplicate` in repositories will merge relationships.
- With logic for cascade delete/anchors/etc.
- This can be automated with engines.
- Combining `KnownDuplicate` to merge with `KnownDuplicateLinks` with `type: Consolidated` and `status: APPROVED` in resulting-linkage.
- Use stewardship to resolve or invalidate bad information in linked.
- Suggest we don't copy relationships from host repository, but link.
- open metadata types
- same `qualifiedName` values for two assets.

Engine Services:

- consolidate discovery/stewardship into engine host.
- flow:
 - Engine service, engine type, request type, type of service, action.
- Discovery (`Asset Analytics`) for actions from inside of actions.

- WatchDog watches changes in metadata looking for problems
 - de-duplication work
 - classifications
 - creates request-for-actions in Triage
- Triage from WatchDog request for action for automatic or manual interaction.
- Remediation gets issues from WatchDog and receives interactions in automated/manual actions.
- Scheduling for things needed from regulation or automation based on time. request-for-action
- Events listening in each engine and progression.
- Discovery engine work with services such as quality, annotations, etc.
 - Open Discovery Analysis Report
- WatchDog triggered on completed ODAR reports from above.
- OMAS partnered with service invocation.
- Triage uses rules/triggers/actions to move state/events
- There is choice to send directly to remediation based on status.
- Scheduler effective date or actions on dates.
- Asset Provisioning event for external event or action or engine internal or external.
- Metadata: 0461 Governance Action Engines
 - We could create an archive to target specific domains/package to add capability.
- 0462: Governance Action Types
 - GovernanceActionType current to next or define a pre-defined plan based on conditions.
 - SoftwareServerCapability link to engine to take action
- 0463: Governance Actions
 - Audit record of actions take in governance.
 - GovernanceAction
 - asset/referenceable working on, action type, engine, and activity.
 - highlight where the stewardship decisions are made.
- Asset Analysis
 - Engine Host OMAG Server

Review Material



deduplication.drawio



engine-services.drawio

Action items

