Open Metadata Repository Services (OMRS)

Overview

The Open Metadata Repository Services (OMRS) enable metadata repositories to exchange metadata. Traditional metadata management technology tends to centralize metadata into a single repository. An organization often begins with a single metadata repository, typically deployed to support a single project or initiative. However, over time, depending on the tools they buy, the projects they run or the political structures within the organization, the number of deployed metadata repositories grows, creating multiple metadata silos. So for example, an organization may have:

- a metadata repository and tools for its governance team. This metadata repository may host the canonical glossary, and the governance policies, rules and classifications.
- a metadata repository for its data lake. This metadata repository has the details of the data repositories in the data lake and the movement of data between them.
- a metadata repository for its data integration tools that continuously extract data sets from the operational systems and send them to the data lake.

The role of the OMRS is to bring these metadata repositories together so this metadata can be linked and used together across the organization. It enables these metadata repositories to act as a aggregated source of metadata. The metadata repositories using OMRS may be instances of Apache Atlas and they may include a mixture of repositories from different vendors that support the OMRS integration interfaces.

Home Metadata Repositories

The open metadata philosophy states that metadata should be managed as close to its source as possible but it should also be easily accessible through standard open APIs and notifications. This means new metadata entities are created in the metadata repository that is connected to the source of the metadata, such as:

- A tool that a person uses to create metadata;
- An engine that is processing data and recording lineage; or
- A specific data platform that is cataloguing new data sources.

The metadata repository where a metadata entity or relationship is created is called its **home metadata repository**. OMRS then ensures this metadata is sharable.

Open Metadata Repository Cohort

An **open metadata repository cohort** is a collection of metadata repositories sharing metadata using OMRS.

The metadata integration capability of OMRS needs to be flexible to support different non-functional requirements. For example, where metadata is changing rapidly (such as in a data lake), this metadata should be dynamically queried from its home metadata repository because the rate of updates mean it would cost a lot of network traffic to keep a copy of this metadata up to date. On the other hand, governance classifications (such as confidentially) and glossary terms rarely change. They are often administered centrally by the governance team and then linked to all metadata that describes the organization's data resources. Thus it makes sense for this metadata to be replicated across the repositories within the cohort. These copies are called reference copies of the metadata entities/relationships and they are read-only.

To join an open metadata repository cohort, a metadata repository must support the following OMRS integration methods. Each method is optimized for specific use cases and so the metadata repository can only play a full role in the open metadata use cases if it supports all integration methods. These are:

- Support for an **OMRS repository connector** to allow open metadata API calls to the repository to create, query, update and delete metadata stored in the repository. The OMRS connectors support the Open Connector Framework (OCF) to provide a call interface to the metadata repositories. The **OMRS Repository Connector API** is a standard interface for all metadata repositories. This enables services such as the **Enterprise OMRS Repository Connector** to interact with 1 or many metadata repositories through the same interface. The connection configuration it passes to the OCF determines which type of OMRS connector is returned by the OCF.
- Support for the **OMRS event notifications** that are used to synchronize selective metadata between the metadata repositories.
- Support for **OSLC linked data relationships** to allow relationships between metadata entities that happen to reside in different metadata repositories.

The OMRS protocols are peer-to-peer. Each repository in the cohort has an **OMRS Cohort Registry** that supports the registration of the metadata repositories across the cohort. Through the registration process, each OMRS Cohort Registry assembles a list of all of the members of the cohort. This is saved in the OMRS Cohort Registry Store. The list of connections to the remote members of the cohort are passed to the **OMRS Enterprise Connector Manager** that in turn manages the configuration of the **Enterprise OMRS Repository Connectors**.

The Enterprise OMRS Connector provides federated query support across the metadata cohort for the **Open Metadata Access Services (OMAS)**.

When a metadata repository registers with the OMRS Cohort Registry, the administrator may either supply a unique server identifier, or ask the OMRS to generate one. This server identifier (called the **metadata collection identifier**) is used in the OMRS event notifications, and on OMRS repository connector calls to identify the location of the home copy of the metadata entities and to identify which repository is requesting service or supports a particular function.
Once the metadata repository has registered with the OMRS Cohort Registry, it is a member of the metadata repository cohort and it can synchronize and share metadata with other repositories in the cohort through the OMRS Topic.

**Note:** A single metadata repository can register with multiple metadata cohorts as long as its server identifier is unique across all cohorts that it joins and it manages the posting of events to the appropriate OMRS Topic. This pattern is useful for a metadata repository that is aggregating reference copies of metadata from multiple metadata repository cohorts.

**Summary**

In summary, the OMRS enables the integration of metadata that is distributed amongst a number of metadata repositories either:

- Through a call interface which is provided by an OMRS connector.
- Using notifications that broadcast changes to metadata in a repository that other repositories can subscribe to in order to maintain reference copies of specific metadata entities.
- Via linked data URLs that enable a metadata entity to have a relationship with a metadata entity in a different repository.

The OMRS has a OMRS Cohort Registry to manage the membership of the metadata repository cohort.