USING INTEGRATION CONNECTORS

Mandy Chessell CBE FREng
Egeria Open Source Project Lead
<table>
<thead>
<tr>
<th>Date</th>
<th>time</th>
<th>Title</th>
<th>Description</th>
<th>Presenter</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 2022</td>
<td>n/a</td>
<td>n/a</td>
<td>There is no webinar this month. The team is working on creating a new set of monthly Dojo sessions. The first of these sessions will be on the 17th of January 2022 10AM - 17:00PM (UTC) and will be on <strong>How to setup and run Egeria in different environments.</strong> This will include the Kubenetes environment. <a href="https://zoom.us/j/523629111">Zoom Conference</a></td>
<td></td>
</tr>
<tr>
<td>7th February 2022</td>
<td>15:00 UTC</td>
<td>Using an integration connector</td>
<td>This session covers how to use Integration connectors to connect technologies into Egeria.</td>
<td>Mandy Chessell</td>
</tr>
<tr>
<td>7th March 2022</td>
<td>15:00 UTC</td>
<td>How to build an integration connector</td>
<td>This session covers how to extend Egeria’s automated cataloguing to include metadata from a new technology. It describes how automated cataloguing works and the role of the integration connector. It covers the design of the integration connector using examples to illustrate the different approaches and their benefits and challenges. It shows how to set up a project for a new connector, how to build and package it and finally it shows the new connector running in Egeria.</td>
<td>Mandy Chessell</td>
</tr>
<tr>
<td>4th April 2022</td>
<td>15:00 UTC</td>
<td>Using a repository connector</td>
<td>This session covers how to use <a href="https://egeria-project.org">Repository Connectors</a> to connect technologies into Egeria; focussing on <a href="https://egeria-project.org">XTDB</a> (formerly known as crux). Every wanted to know what the state of your metadata was at some specific time in the past? This session will introduce the XTDB open metadata repository that supports these historical metadata queries.</td>
<td>Chris Grote</td>
</tr>
</tbody>
</table>
Coco Pharmaceuticals persona

https://odpi.github.io/data-governance/coco-pharmaceuticals/personas/
Callie has a database server that she uses to analyze relational data.

She creates a new sandbox for each type of analysis.

However, she often forgets to catalog her sandboxes.
Automated cataloguing of Callie’s Sandboxes

Callie creates a new sandbox

Metadata for new sandbox created automatically
The Integration Daemon

- A type of OMAG Server
- Metadata extraction, capture and delivery
Inside the integration daemon

[Diagram illustrating the integration daemon with various components and connections, including Polling connector, Inbound listening connector, Outbound listening connector, Context, REST API Calls, In Topic, Out Topic, and metadata access server.]
Supporting the metadata needs of different technologies
Integration Connector configuration

- The configuration provides the integration daemon with the information it needs to control the lifecycle and runtime support needed by the connector.

```
integrationServiceId
integrationServiceContextManagerClass
integrationServiceName
integrationServiceFullName
integrationServiceURLMarker
integrationServiceDescription
integrationServiceWiki
integrationServicePartnerOMAS
defaultPermittedSynchronization
integrationServiceOperationalStatus
omagServerPlatformRootURL
omagServerName
integrationServiceOptions
integrationConnectorConfig
```

- Descriptive information about the service.
- Is the service available?
- Location of the metadata access point / metadata server where the access service is running
- Options supported by the integration service.
- List of connectors to run (typically one for each third party technology).

```
connectorId
connectorName
connectorUserId
connection
metadataSourceQualifiedName
refreshTimeInterval
usesBlockingCalls
permittedSynchronization
```
Integration Connector Implementation

IntegrationConnector

- initialize()
- setAuditLog()
- setConnectorName()
- setContext()
- start()
- engage()
- refresh()
- disconnect()

For saving connector instance identifier and connection object.
Called from the ConnectorBroker.

Provides the logging destination.
Provides the name of the connector for logging.

Set up the integration service specific context.

Indicates that the connector is completely configured and can begin processing. This call can be used to register with non-blocking services. For example it can register a listener with the OMAS Out Topic with the context.

Used for blocking calls to wait for new metadata. It is called from its own thread iff the connector is configured to have its own thread. It is recommended that the engage() method returns when each blocking call completes. The integration daemon will pause a second and then call engage() again. This pattern enables the calling thread to detect the shutdown of the integration daemon server.

Requests that the connector does a comparison of the metadata in the third party technology and open metadata repositories. Refresh is called when the integration connector first starts and then at intervals defined in the connector's configuration as well as any external REST API calls to explicitly refresh the connector.

Free up any resources held since the connector is no longer needed.
Metadata Extraction (remote store)
Metadata Distribution
Metadata Synchronization
External Asset Manager

Legend:
- Open Metadata Integration Service (OMIS)
- External Asset Manager
- Integration Connector
Many-to-One Mapping

External Asset Manager: myCatalog

**BusinessTerm**
- id: gt1
- name: car manufacturer

**Example**
- id: ex6
- name: Volvo

Open metadata

**AssetManager**
- myCatalog

**ExternalIdentifier**
- gt1
- ex6

**GlossaryTerm**
- guid: 98400
- qualifiedName: car_manufacturer30307777
- displayName: car manufacturer
- example: Volvo

https://egeria-project.org
One-to-Many Mapping

External Asset Manager: myCatalog

myCatalog

Database
- id: 123
- name: database1
- abbreviation: db1

Table
- id: 456
- name: db1:table1

GlossaryTerm
- guid: 04920
- qualifiedName: database1_supplement
- abbreviation: db1

DeployedDatabaseSchema
- guid: 07275
- qualifiedName: database1

RelationalDBSchemaType
- guid: 409088
- qualifiedName: db1:schemaType

AssetManager
- myCatalog

ExternalIdentifier
- 123

ExternalIdentifier
- 456

RelationalTable
- guid: 203480
- qualifiedName: database1:table1

Open metadata
Onboarding an organization
Steward Notifications
Pushing security tags to security managers
Pushing security tags to a security manager
OpenLineage standard
Receiving OpenLineage events
Direct receipt of OpenLineage events
Lineage Deployment

Direct integration with Egeria

Spark Job

Spark Engine

Data Sources

Integration Daemon

Open Lineage HTTP Endpoint

Lineage Integrator OMIS

Lineage Integration Connector

Choosing what to save

Validating execution against expectations.

Engine Host

Governance Action OMAS

Process Validation Connector

Governance Engine OMAS

Metadata Access Server

Asset Manager OMAS

Data Manager OMAS

Metadata Capture.
Correlation important across the different metadata capture mechanisms

Cohort Topics

https://egeria-project.org
Inside the Lineage Integrator OMIS
<table>
<thead>
<tr>
<th>Date</th>
<th>time</th>
<th>Title</th>
<th>Description</th>
<th>Presenter</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 2022</td>
<td>n/a</td>
<td>n/a</td>
<td>There is no webinar this month. The team is working on creating a new set of monthly Dojo sessions. The first of these sessions will be on the 17th of January 2022 10AM - 17:00PM (UTC) and will be on How to setup and run Egeria in different environments. This will include the Kubenetes environment. Zoom Conference <a href="https://zoom.us/j/523629111">https://zoom.us/j/523629111</a></td>
<td></td>
</tr>
<tr>
<td>7th February 2022</td>
<td>15:00 UTC</td>
<td>Using an integration connector</td>
<td>This session covers how to use Integration connectors to connect technologies into Egeria.</td>
<td>Mandy Chessell</td>
</tr>
<tr>
<td>7th March 2022</td>
<td>15:00 UTC</td>
<td>How to build an integration connector</td>
<td>This session covers how to extend Egeria’s automated cataloguing to include metadata from a new technology. It describes how automated cataloguing works and the role of the integration connector. It covers the design of the integration connector using examples to illustrate the different approaches and their benefits and challenges. It shows how to set up a project for a new connector, how to build and package it and finally it shows the new connector running in Egeria.</td>
<td>Mandy Chessell</td>
</tr>
<tr>
<td>4th April 2022</td>
<td>15:00 UTC</td>
<td>Using a repository connector</td>
<td>This session covers how to use Repository Connectors to connect technologies into Egeria; focussing on XTDB (formerly known as crux). Every wanted to know what the state of your metadata was at some specific time in the past? This session will introduce the XTDB open metadata repository that supports these historical metadata queries.</td>
<td>Chris Grote</td>
</tr>
</tbody>
</table>
Open forum

Questions?
THANK YOU!
Achievements

- 700 linked open metadata types demonstrating how the knowledge from many tools can be linked together.

- Open metadata repository interface proven for table, graph and hierarchical DB stores.

- Enterprise queries and replication across heterogeneous technologies

- Conformance test suite and mark

- Automated configuration of data virtualization technology and security as new data sets are added to a data lake

- Suite of persona-based labs and tutorial using Jupyter Notebooks.

- Virtual graph of metadata maintained across distributed heterogeneous metadata repositories.

- Frameworks, APIs and connectors for minimizing integration cost for different types of technologies

- Virtual repository explorer UI

- Instance based security

- Controlling visibility of assets through zones

- Scalable, secure platform configurable and customizable through connectors

- Purpose-based data access

- Metadata versioning and provenance

- Multi-tenant UI based on carbon

- W3C semantic standards pattern for data model exchange

- Automation of metadata acquisition through templates, daemons, discovery services and stewardship.

- Classification of assets

- Reference data management

- Multi-technology collaboration and feedback

- Multi-domain governance model

- Digital service lifecycle, from business design, development, devOps and use.

- Comprehensive open lineage services.

- Metadata deduplication
Scope of metadata covered

- Governance
- Lineage
- Base Types, Systems and Infrastructure
- Glossary
- Data Assets
- Collaboration
- Models and Reference Data
- Metadata Discovery
Scope of metadata covered

Policy Metadata (Principles, Regulations, Standards, Approaches, Rule Specifications, Roles and Metrics)

Governance Actions and Processes

Campaigns and Projects

Governance Actions and Processes

Rights Management

Implementation

Strategy

Rollout

Classification Schemes

Augmentation

Organization

Business Attributes

Business Objects and Relationships, Taxonomies and Ontologies

Subject Area Definition

Asset Collections (Sets, Typed Sets, Type Organized Sets)

Information Views

Physical Asset Descriptions (Data stores, APIs, models and components)

Connectors

Instrumentation (design lineage)

Information Process

Basic Types, Infrastructure and Systems

Access

Augmentation

Teaming Metadata (people profiles, communities, projects, notebooks, ...)

Feedback Metadata (tags, comments, ratings, ...)

Models and Schemas

Reference Data

Discovery Metadata (profile data, technical classification, data classification, data quality assessment, ...)

https://egeria.odpi.org/open-metadata-publication/website/open-metadata-types/

https://egeria-project.org
Using Egeria …

- Eases the cost of metadata integration through
  - Comprehensive standards and libraries.
  - Active vendor recruitment program.
- Provides direct support to many governance roles, filling the gaps between function offered through commercial tools.
- Provides best practices and content packs to accelerate an organization’s journey to becoming data driven.