Talend Metadata Manager

Reduce Risk and Friction in your Information Supply Chain
Talend Metadata Manager

Talend Metadata Manager provides a comprehensive set of capabilities for all facets of metadata management. At the heart of Talend Metadata Manager is a repository which contains repository objects, such as models and mappings that are organized into folders. Models can be harvested from Talend Data Integration models, Data Modeling tools, Data Warehouses, external metadata repositories for relational databases (RDBMS), and Data Integration and Business Intelligence tools. A particular type of repository object called *Configuration*, can connect “metadata stitching” models and mappings together to represent an Enterprise Architecture, including full support for data flow lineage and impact analysis, as well as semantic lineage definitions.

Talend Metadata Manager consists of four major components:

- Metadata Bridge (metadata import)
- Metadata Manager
- Data Governance
- Metadata Authoring with Forward Engineering (metadata export)
Metadata Bridge

*Metadata is everywhere*. Data warehousing, business intelligence, CASE and ETL tools all have their own repositories. Just about every application has its own data dictionary. XML carries the metadata with it in the message or document, and enterprise application integration environments have their own repositories and metadata mapping and integration facilities.

In order to succeed, one must have a good enterprise repository integration environment that can integrate the different format of metadata from all tools. The Talend Metadata Manager repository bridges the technical and non-technical aspects of metadata, while simultaneously addressing the chasm between the different metadata source and target systems that constitute any modern information management environment.

The Metadata Bridge imports all metadata via “bridges” (metadata import components), including Extract, Transformation and Load (ETL)/ Data Integration tools, Business Intelligence tools, Data Modeling tools, databases, most all metadata exchange standards, and numerous data formats including XML.

*Importing metadata from Talend Studio with Talend Metadata Manager*
Metadata Manager (MM)

Version and Configuration Management
Not only must the repository be able to import on demand in any format and to any tool or import metadata many times as needed, it must be able to manage the versions created by this continuous activity. It must also be fundamental to the repository organization for administrators to then organize, publish and selectively present the information in appropriate configurations of metadata, as is required for the correct and precise answers to a wide range of “cuts” across this metadata.

Talend Metadata Manager was designed from the ground up with version and configuration management as a key capability.

Metadata Comparison
All metadata is represented by an integrated metamodel in Talend Metadata Manager. This feature provides comparisons across metadata from data source formats supported, including design tools, databases, etc., not simply among versions of a given model.

Comparing models or model versions with Talend Metadata Manager
Data Mapping Specifications

Once imported, metadata can be mapped in a myriad of ways to any other metadata within Talend Metadata Manager. This ability is critical to the success of any metadata management solution. In particular, you can define data flow mappings describing data movement type relationships, e.g. when a database is read and the results written to another database, as well as semantic mappings which identify semantic relationships between elements, oftentimes conceptual or logical in nature, such as for a data dictionary or conceptual model such as a UML model.

Metadata Stitching

Metadata stitching is fundamental to the correct and automated analysis of the data flow and semantic lineage of metadata in the repository. It also supports version management across the constant rate of updates and changes in a repository.

Talend Metadata Manager keeps complete versions of all imported metadata in self-contained “models”, which are then related via stitching’s (simple connection mappings). In this way, version management and configuration management is not only entirely clean and isolated from the definition and maintenance of mappings, it also automatically supports updates and changes into the future.

Getting a high level view of information flows across systems with metadata stitching
In this way, the enterprise architecture is correctly modeled, and data flow lineage is completely and accurately derivable.

<table>
<thead>
<tr>
<th>Population in the org.</th>
<th>Roles with respect of Data and IT</th>
<th>Need for impact analysis</th>
<th>Need for Lineage</th>
</tr>
</thead>
<tbody>
<tr>
<td>IT Architect &amp; Developer</td>
<td>Develop and design, Maintain and evolve, Administer and control</td>
<td>Business process improvements</td>
<td>Support business users</td>
</tr>
<tr>
<td>Business User</td>
<td>Enterprise app. user, Data entry, Reporting &amp; analysis</td>
<td>Not applicable</td>
<td>Reporting &amp; analysis</td>
</tr>
<tr>
<td>Business User</td>
<td>Define Enterprise terms, Author &amp; resolve errors, Specify requirements</td>
<td>Business process improvements</td>
<td>Risk &amp; compliance</td>
</tr>
</tbody>
</table>

The different roles and their needs with respect to data and related metadata

Lineage and Impact Analysis
Once metadata is managed, metadata is then available for detailed technical and business analysis. Talend Metadata Manager supports full technical and business level lineage and impact analysis providing you new insight across all the connected metadata sources.

Business User – Lineage

*Reporting analysis* is the typical use case, with questions such as:
- Given an item on a report, what data entry system fields impact these results?
- Why are the numbers on this report the way they are?
- How do I change the system data to correct the results of this report?
Technical User – Impact Analysis
Of high interest to the technical user are questions like:

• If I must change these elements (data type, code sets, etc.) in my operational data store, what is the downstream impact?
• This new ETL process is populating my staging warehouse in new ways, how does this impact the OLAP model in my reporting services?

Technical User – Lineage
Reverse lineage type questions may also be asked by more technical users, such as:

• How many systems are required to determine the dimensions for this portion of the OLAP model?
• A business report use case is asking the lineage for particular values on a report, so where does the data come from and how is it manipulated?

Business Users – Impact Analysis
Finally, business users may ask the forward lineage or impact analysis questions, such as:

• If I make a change to this field, what reports will be impacted?
• How is this identity information merged with the personnel system information on these other reports?

Impact analysis with Talend Metadata Manager
Data Governance (DG)

Critical to the development and management of a complete data architecture is a Business Glossary. Talend Metadata Manager provides an ISO 11179-based Business Glossary to capture, define, maintain and implement an enterprise Business Glossary of terminology, data definitions, code sets, domains, validation rules, etc. In addition, semantic mappings describe how elements in a source Model (more conceptual like the Business Glossary) define elements in a destination Model (closer to an implementation or representation).

The Business Glossary helps an enterprise reach agreement between all stakeholders on their business assets (e.g. terms) and how they relate to data assets (e.g. database tables) and technology assets (e.g. ETL mappings). The Business Glossary can be used to document logical/physical data entities and attributes across IT collaboratively. Again, it involves tracing dependencies between business and technical assets.

In Talend Metadata Manager, a Business Glossary is a self-contained collection of categories and the terms sub-categories contained within each category. In turn, the terms may be semantically mapped to objects throughout the rest of the repository, such as tables and columns in a data model. Once mapped, one may perform semantic lineage traces such as definition lookups and term semantic usage across any configurations containing the Business Glossary, mappings and mapped objects.

![Business Glossary](image)

*Authoring the common business terms used in the organization with the Business Glossary*
Bootstrapping a Business Glossary

Building a Business Glossary can be as simple as dragging in an existing well-documented data model, via import from other sources (a CSV file format), or can be populated directly via the user interface during the process of classifying objects in other data store models. In general, a combination of such methods are employed in conjunction with one another.

Workflow

In order to ensure that the Business Glossary is accurate, up-to-date, available to all who need access to it, and integrated properly with the rest of the metadata in the repository, Talend Metadata Manager also provides a robust collection of Data Governance tools and methodologies. The Business Glossary provides a very flexible workflow and publication process that can address both basic and complex needs. In addition, one may maintain any number of business glossaries, each with different workflow and publication characteristics.

The Business Glossary may be part of your lineage. It will appear in the repository panel and when you open a Business Glossary, you will be presented with a different UI than other (imported) Models.

Workflow-driven search criteria are available allowing one to efficiently organize terms and identify what actions are required at any given time. When working with individual terms, which are at some point in the workflow process, workflow transition buttons prompt you with possible actions.

Semantic Mapping

A Semantic Mapping describes how elements in a source model (more conceptual) define elements in a destination model (closer to an implementation or representation). Put another way, elements in the destination model are representations or implementations of the associated element in the source model.

They are three primary uses for semantic mapping:

- Data Standardization and Compliance
- Multi Level Modeling of semantic relationships from conceptual to logical, and to physical data model with a few sub cases
- Business Glossary term classification
Metadata Authoring (MA) with Forward Engineering (Metadata Export)

*Note: The following features only come with Talend Metadata Manager with Authoring.*

**RDBMS and Big Data Documenter and Physical Data Modeler**

The Talend Metadata Manager Data Documenter allows users to document existing data stores, like databases, big data sources, and imported models, and publish the resulting documented data stores to the enterprise.

The Data Documenter offers a different approach than traditional data modeling tools:

- The Business Glossary-driven Data Documenter methodology allows for immediate reuse and creation of terms and naming standards on the fly, fast tracking the data store documentation process ensuring complete semantic synchronization among your data models and data governance environment.
- Web-enabled Data Documenter offers better access to users than desktop tools
- Data Modeling and diagramming capabilities of the Data Documenter are similar to conventional data modeling tools.
- Full integration (import/export) to most popular data modeling tools is provided.

*Visualizing Data Models with Talend Metadata Manager*
Logical Data Modeler

Talend Metadata Manager provides a completely web-enabled logical data modeling environment for producing logical and conceptual models:

- The Business Glossary-driven methodology allows for immediate reuse (creating of entities, attributes and domains) and creation of terms and naming standards on the fly, fast tracking the modeling process and ensuring complete semantic synchronization among your models and data governance environment.
- The Web-enabled modeler offers better access to users than desktop tools.
- The Data Modeling capabilities are competitive with conventional data modeling tools.
- Full integration (import/export) with most popular data modeling tools is provided.

Data Mapping Designer

Data Mapping Designs represents data integration process designs containing all the necessary data movement design details, such as lookups, filters, joins and transformation expressions. These Data Mapping Designs are complete enough that they may be forward engineered into Talend Data Integration using the Metadata Bridge. In this way, Talend Metadata Manager provides a completely web-based data mapping design tool that can reuse and be synchronized with all other metadata artifacts in the repository and your complete data governance environment.

Defining the mappings directly in Talend Metadata Manager
Visualizing the end to end information flows with Talend Metadata Manager