Metadata and the Rise of Big Data Governance: Active Open Source Initiatives

October 23, 2018
Today’s speakers

John Mertic, Director of Program Management, Linux Foundation

David Radley, ODPi Egeria maintainer, IBM
Today’s reality
Imagine …

An enterprise data catalog that lists all of your data, where it is located, its origin (lineage), owner, structure, meaning, classification and quality

- Spanning systems both on premises and within cloud providers
- Hosted locally to your data platforms but integrated to provide the enterprise view
Imagine …

New data tools (from any vendor) connect to your data catalog out of the box

No vendor lock-in; nor expensive population of “yet another” proprietary siloed metadata repository
Imagine …

Metadata is added automatically to the catalog as new data is created

Extensible discovery processes characterise and classify the data

Interested parties and processes are notified
Imagine …

Subject matter experts collaborating around the data

Locating the data they need, quickly and efficiently

Feeding back their knowledge about the data and the uses they have made about it to help others and support economic evaluation of data
Imagine …

Automated governance processes protect and manage your data

- Metadata-driven access control
- Auditing, metering and monitoring
- Quality control and exception management
- Rights management
A new manifesto for metadata and governance!

Metadata management must be automated
Metadata management must become ubiquitous
Metadata must become open and remotely accessible
Metadata should be used to drive the governance of data

The discovery, maintenance and use of metadata has to be an integral part of all tools that access, change and move information.
How will this be achieved?
Open metadata management ecosystem

Peer-to-peer network of repositories

Metadata stored and managed close to its source

Each repository/tool brings unique value.

Open, extensible metadata structures for metadata exchange and federation

Open source infrastructure sharing cost of development and maintenance

Collaboration Space Metadata

Cloud SaaS platform Metadata

Analytics Platform Metadata

Application Metadata

Hadoop Platform Metadata
Making Metadata Available to the Enterprise!

Cohort A

Data Lake

Chief Data Office

Systems of Record

Cohort B

Marketing

Data Lake

Mobile Apps

Systems of Record

@ODPiOrg
Open metadata data model

- Governance
- Glossary
- Collaboration
- Models and Reference Data
- Lineage
- Data Assets
- Metadata Discovery
- Base Types, Systems and Infrastructure
Open metadata and governance integration patterns

1. Caller
2. Native
3. Adapter
4. Plug-in
5. Special

Apache Atlas
IBM Information Governance Catalog
Business meaning of the underlying assets

Structural metadata for a data store

Data

Employee

Manager

Employee Name

Employee Id

Job Title

Work Location

Compensation Plan

Hourly Pay Rate

Annual Salary

Sensitive

EMPLOYEE RECORD

EMPNAME

EMPNO

JOBCODE

SALARY

00 3809890 6 7 Lemmie Stage 818928 3082 4 New York 4 27 DataStage Expert 1 45324 300 27 Code St Harlem NY 1 3
Instance representations in Open Metadata Resource Services (OMRS)
Open Source Collaboration through ODPi

ODPi is the vendor neutral home for Egeria.

Governance is open to all in the data governance community

- Vendors
- End-users
- Privacy/governance experts

Code is available under an Apache 2.0 license and documentation under a CC-BY-4.0 license

Join the effort at https://github.com/odpi/egeria

@ODPiOrg
## How this helps

<table>
<thead>
<tr>
<th>Data Governance Professionals</th>
<th>Vendors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your governance program if based on established definitions</td>
<td>Your metadata offerings will deliver value faster as they tap into metadata collected by other vendor’s tools.</td>
</tr>
<tr>
<td>Allow a broader range of tools in your organization</td>
<td>ODPi packages extend your metadata system’s and tools’ capabilities</td>
</tr>
<tr>
<td>Automated governance processes protect and manage your data</td>
<td>Conformance tests minimize your effort in being compliant with key standards and regulations.</td>
</tr>
<tr>
<td>Metadata-driven access control</td>
<td>Customers have increased confidence in your tools and services due to ODPi certification.</td>
</tr>
<tr>
<td>Auditing, metering and monitoring</td>
<td></td>
</tr>
<tr>
<td>Quality control and exception management</td>
<td></td>
</tr>
<tr>
<td>Rights management</td>
<td></td>
</tr>
</tbody>
</table>
ODPi – A neutral home for collaboration

PRE-PRODUCTION

- 29% LAB
- 14% POC
- 29% PILOT

PRODUCTION

- ENTERPRISE-WIDE DEPLOYMENT 28%

Gates / Criteria:

- Data visualization and presentation
- Data storage and management
- Data security and governance
- Data velocity and access methods (SQL, streaming, etc.)

72%
Look to The Linux Foundation

Thankfully, that’s where The Linux Foundation® comes in. For nearly two decades, The Linux Foundation has provided unparalleled support for open source communities through financial and intellectual resources, governance structure, IT infrastructure, services, events, and training.

Dedicated to building sustainable ecosystems around open source projects, The Linux Foundation is working with the global technology community to solve the world’s hardest problems through open source and creating the largest shared technology investment in history.

The Linux Foundation is the umbrella organization for more than 60 open source projects accelerating open technology development and commercial adoption. Some of the game-changing initiatives hosted by The Linux Foundation include:
Get involved!

1. Check out ODPi Data Governance on GitHub
   https://github.com/odpi/data-governance

2. Learn more about Egeria
   https://odpi.github.io/egeria/

3. Follow the discussion
   https://lists.odpi.org/g/odpi-pmc-datagovernance

4. Have your organization support ODPi
   https://lists.odpi.org/about/join
Questions?
THANK YOU!

@ODPi.org - www.odpi.org