Building an Open Software Ecosystem

Mandy Chessell
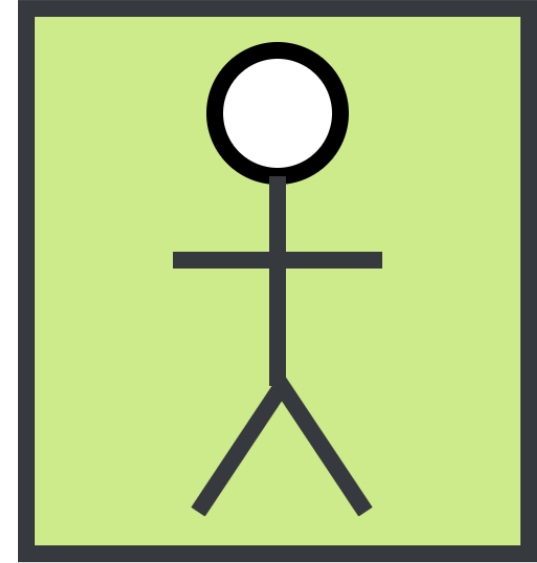Pragmatic Data Research Ltd

June 2022

Brief Introduction

- Mandy Chessell CBE FREng CEng FBCS
  - Project Leader of the Egeria open-source project
  - Based in the UK
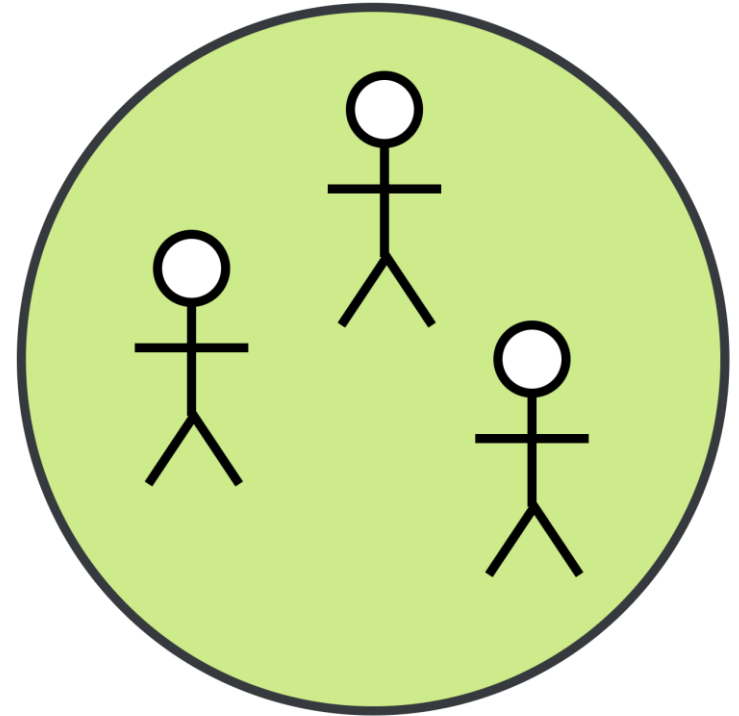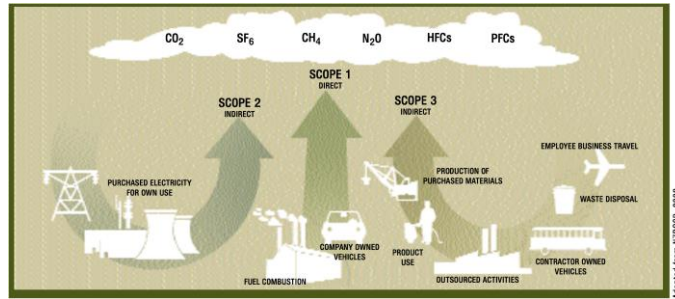 - 30+ years experience in the software industry

Individual action on sustainability …

- How much can an individual do?
    - Recycling
    - Reusing
    - Voting
    - Careful choice of products …

Acting together

- Real change requires concerted action
  - Throughout supply chain and product lifecycle
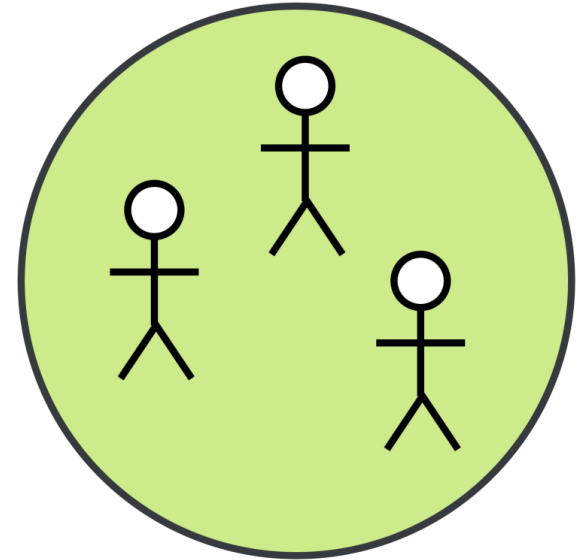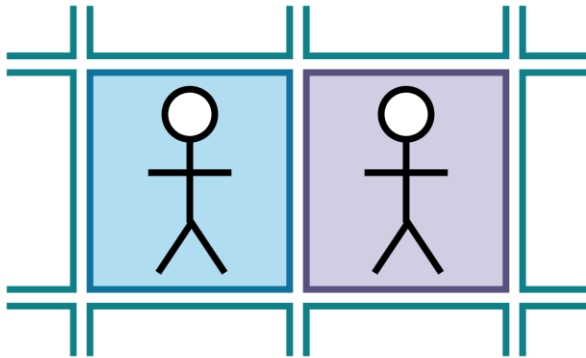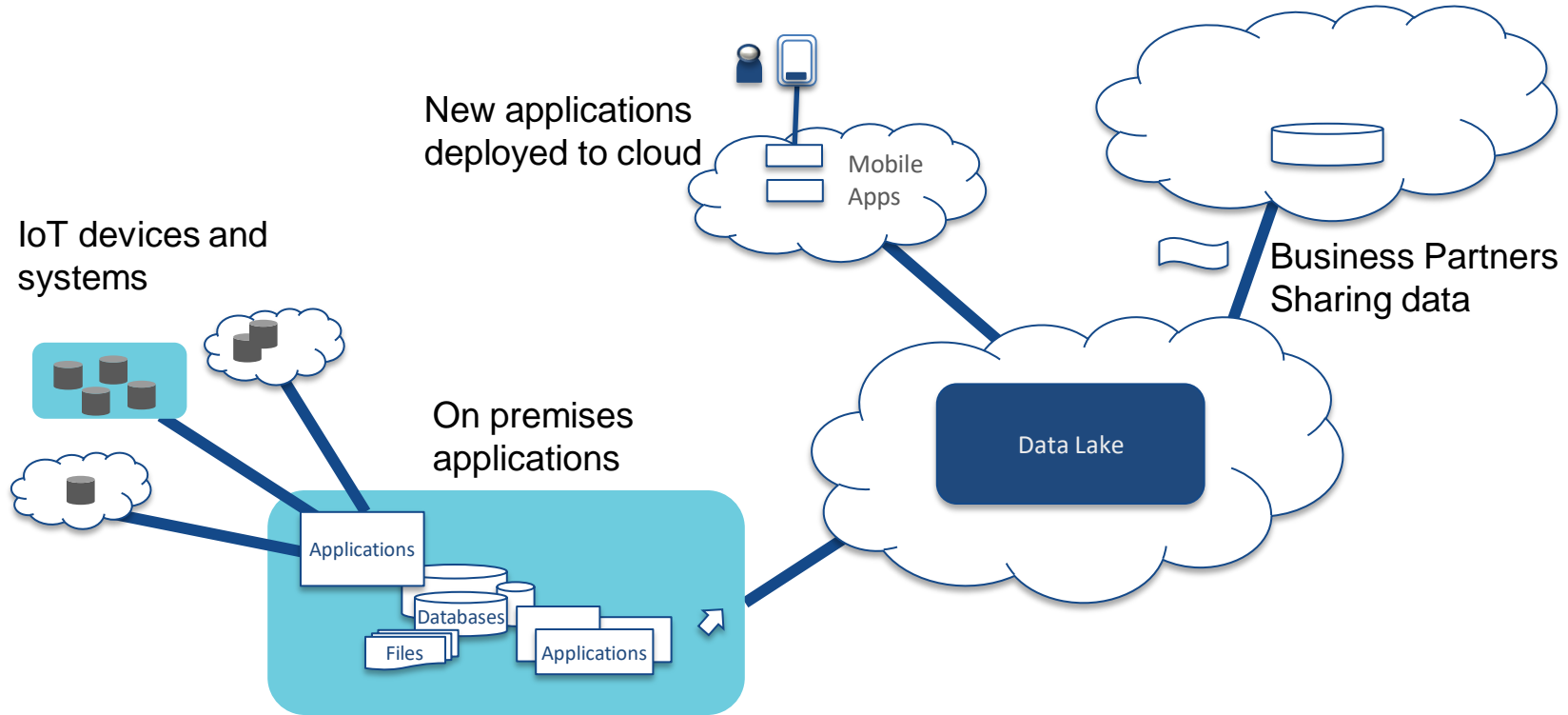 - In all aspects of our planet-wide society …

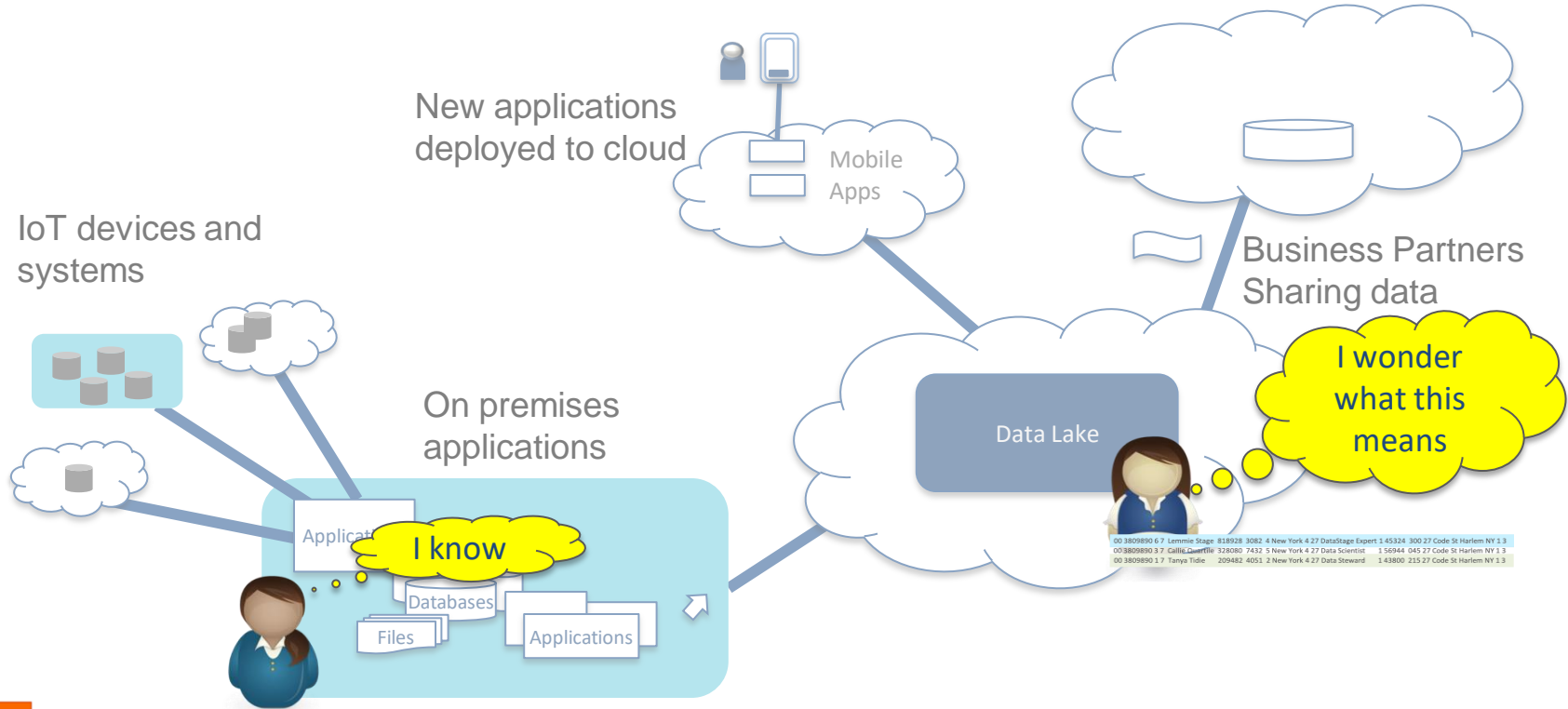# Organizations and Change

Challenge for the organisation

- To understand and report on current sustainability status

- To make changes to improve sustainability of operations

- To measure and report the effects of the change

- … and repeat (whilst still running the business …)

Finding the data you need to report on sustainability


New applications deployed to cloud

Mobile Apps

Business Partners Sharing data

IoT devices and systems

On premises applications

Data Lake

Applications

Databases

Files

Applications

Knowledge is separated in space and time

If these are the computer systems for one business process …

- How can you be sure which data set to use?

- How can you make the integrations more efficient and less error prone?

- If a link or a system goes down, how long do you have before data corruption occurs?

- If an invalid value was added to one of the systems yesterday, how far has the contamination spread?

Knowledge is dispersed within technologies

- Most technologies retain knowledge of the resources that they are working with.

- This knowledge is fragmented; locked behind proprietary formats into small nuggets of value.

- Nothing binds this knowledge together.

- Gathering the facts needed to act is slow and error prone.

Organizations need a solution that links this knowledge together so that they can cross between the organizational, professional and technological gaps in their digital landscape.
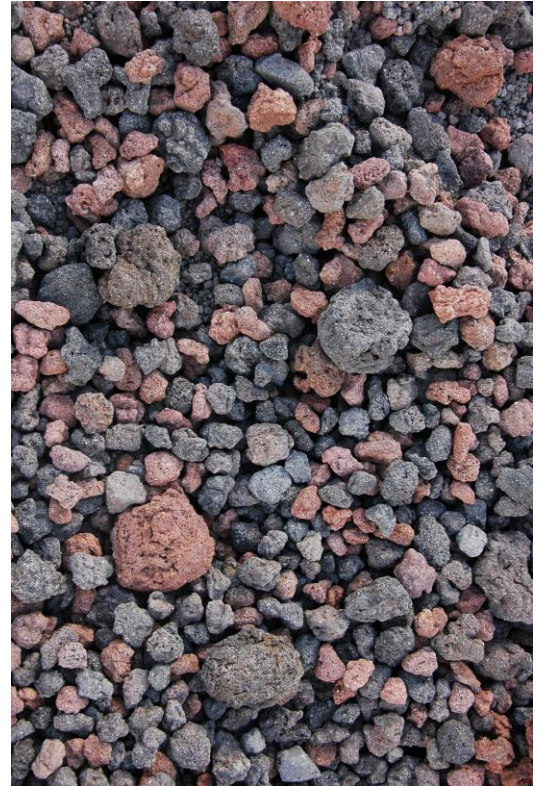
Imagine a data catalogue that brings all of the information about the data sets that is collectively known by the organisation to Callie's data science.

Data Set Name: Employee Directory

Description:
Core attributes describing all employees of Coco Pharmaceuticals created from a daily extract from Kenexa.

Owner: Penny Payer

Classification Ranges:
 Confidentiality: Public, Confidential, Sensitive
 Confidence: Authoritative
 Retention: Indefinitely

Status:
 Last accessed: 6th May 2016
 Records: 3488
 Last Update: 1st May 2016

Contents:
Structure …
Contents …
Lineage …

X

Column:
Band

Description   Characteristics   Lineage

Position reference number for non-exempt employees. The value ranges from 01 to 06 where 01 is the most senior and 06 is the most junior.
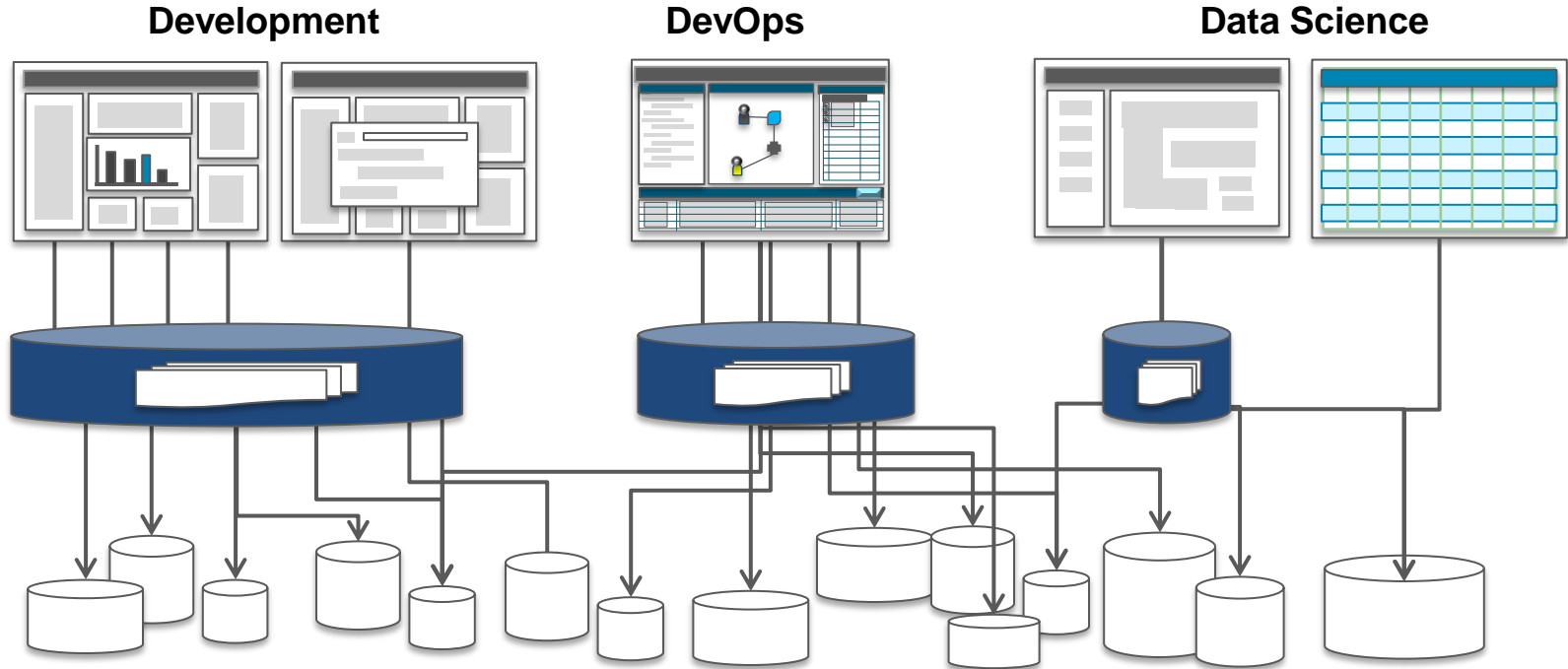
Type: String
Classification: Public

X

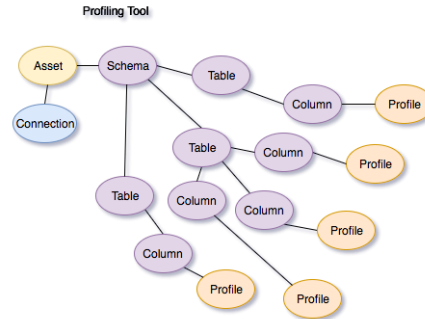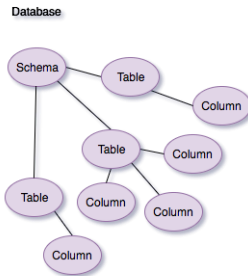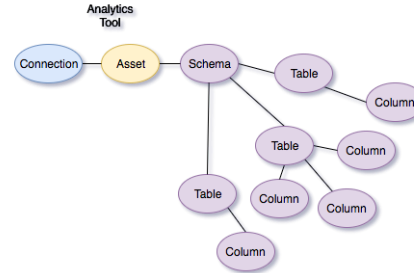IBM Data Science Experience

Employee Directory

Name	Band	Job Title

Add to Project
Connection
Data Set
Notebook
Data Lake Asset

Today's reality – organizations buy lots of tools

Development **DevOps** **Data Science**

Observations on data-driven projects
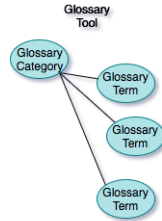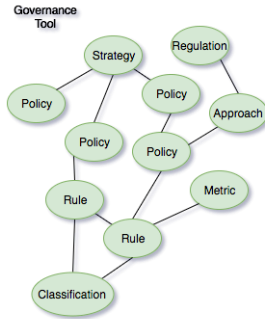
Observations on data-driven projects
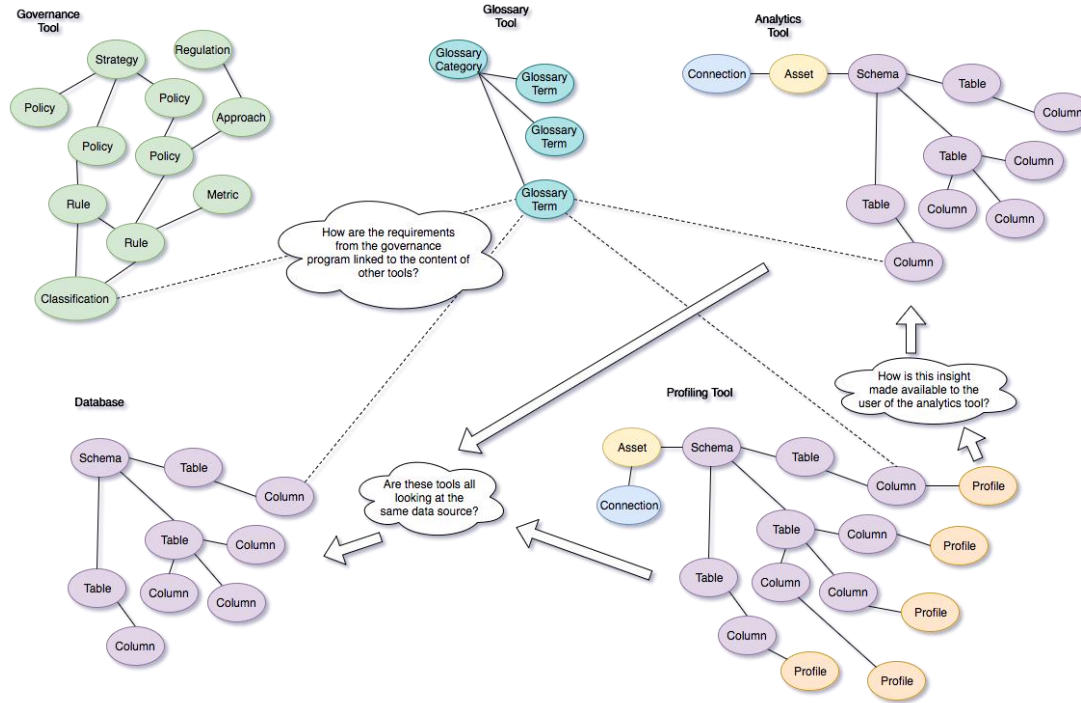
This is not a new problem …

- Over the years, our industry has gained plenty of experience discovering what does not work
    - Centralized metadata repository
    - "Just use our tools" – Single vendor lock in
    - Paper standards that each vendor implements
    - Bridges that copy metadata between tools
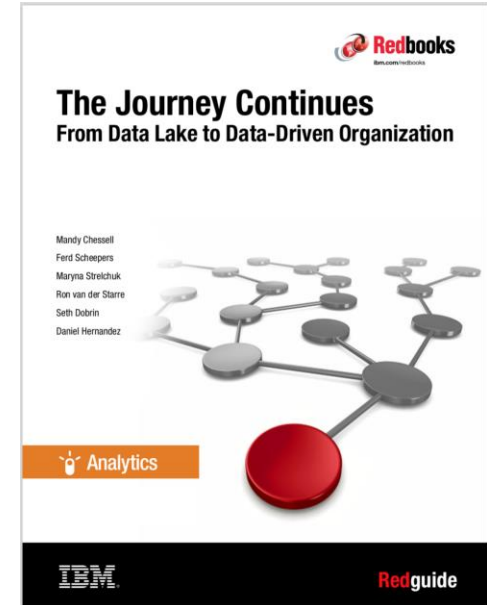   - Bespoke point-to-point integrations between tools

First publication on need for Open Metadata and Governance

http://www.redbooks.ibm.com/abstracts/redp5486.html?Open

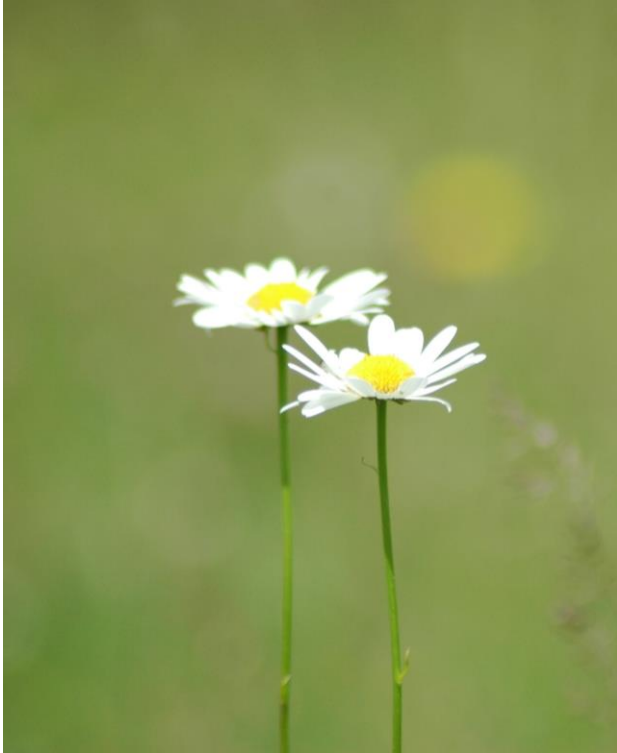https://www.youtube.com/watch?v=LMmm74t-BBA

Egeria's value

- Egeria enables the sharing of knowledge between people and processes that use different tools and technologies
    - Increasing agility, collaboration and the availability of relevant data when making decisions.
    - Breaking down silos between traditional technologies such as data and applications; applications and infrastructure management; security and data.

What is different about Egeria

- Open source, multi-vendor and client collaboration

- Peer-to-peer architecture; retaining the value of each technology both current and future

- Self-configuring, distributed, real-time exchange of knowledge

- Scales from Raspberry Pi to multi-instance HA cloud deployment

- Multi-tenant

- Instance based security
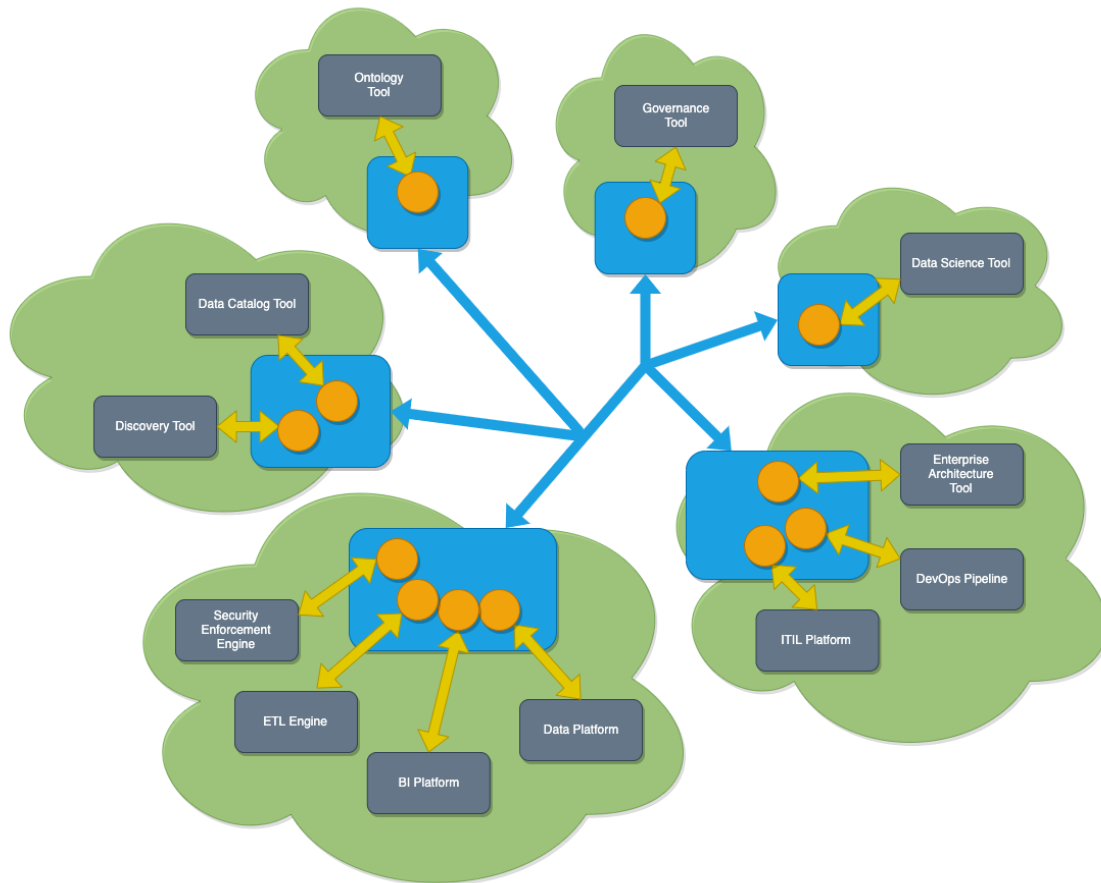
- Customizable through connectors
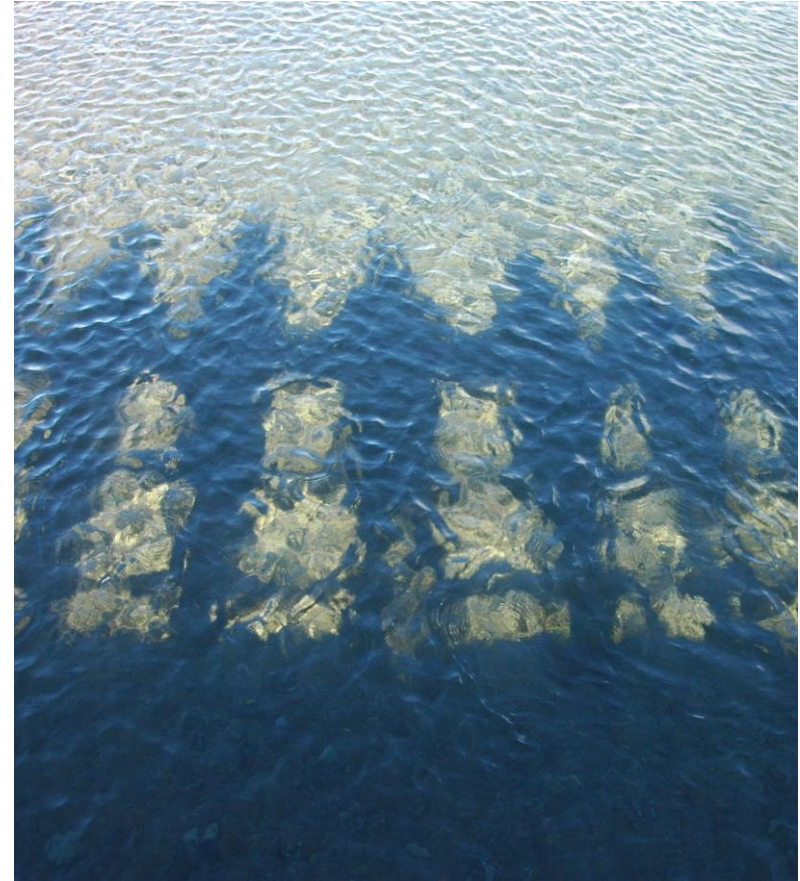
Egeria enables exchange of metadata between tools from different vendors

Development

DevOps

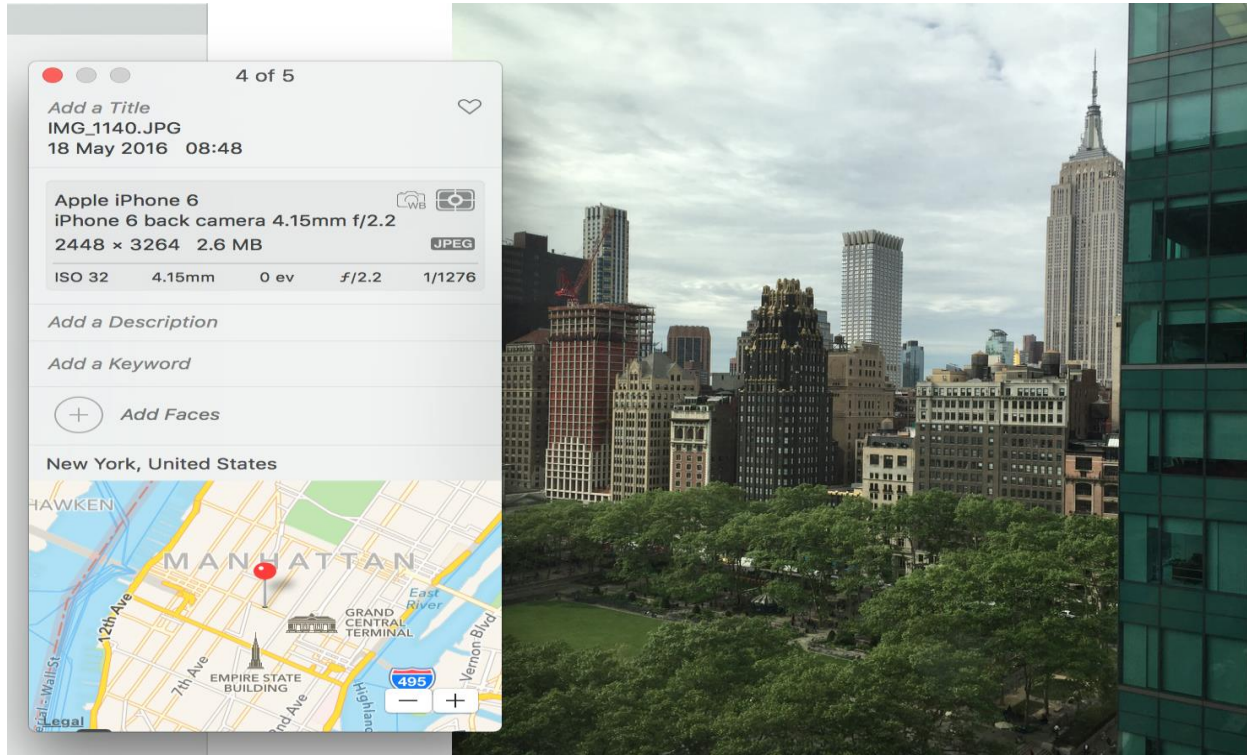**Data Science**

Open and
Unified Metadata
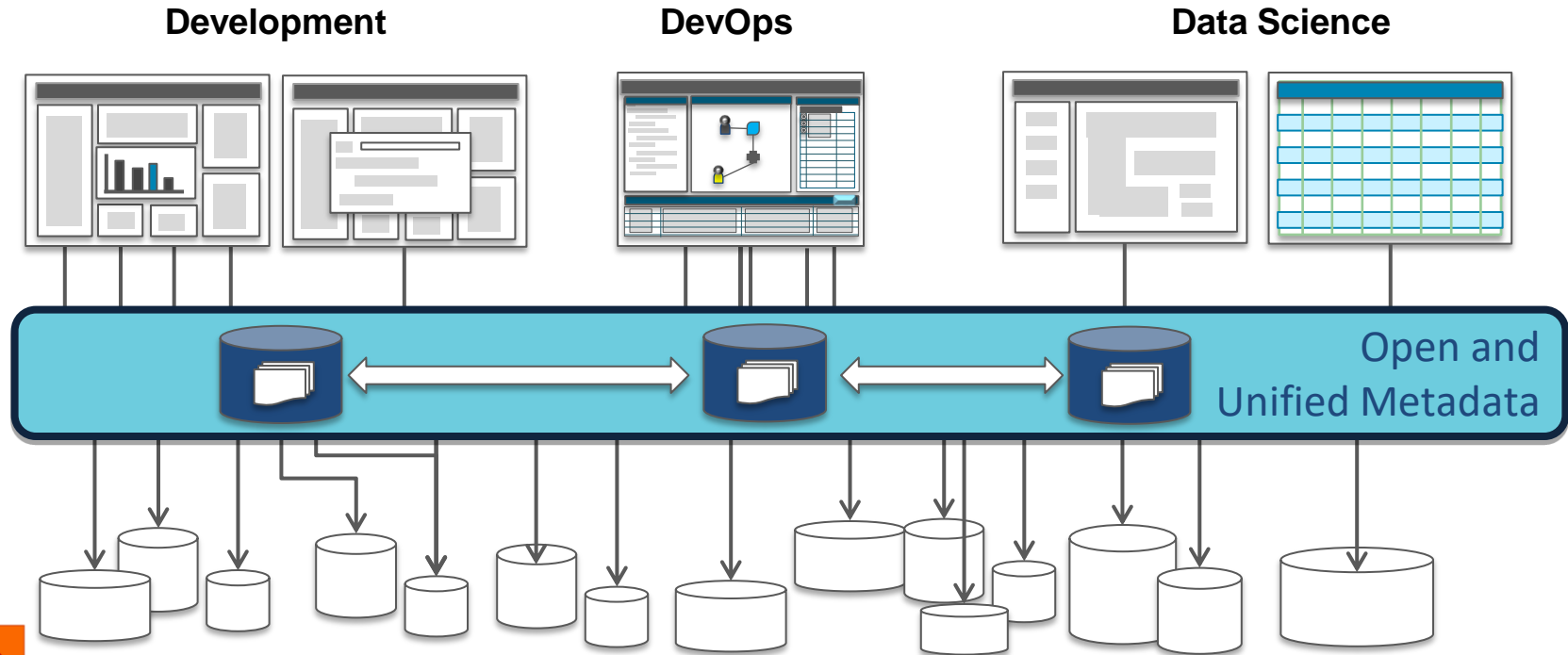
Egeria in operation

- Egeria's platform (in blue) runs in each deployment environment (cloud or on premises)

- Egeria's servers (orange circles) are deployed on the platform. Each server is specialized to support the metadata needs of specific technologies.

- The interaction of the servers organizes the metadata exchange

https://egeria-project.org/guides/planning/overview

Automatic metadata capture

Manual metadata capture

BUILDING EGERIA

How does the Egeria community work

- Iteratively perform a series of practical experiments that ignore the boundaries of traditional approaches.
  - Holistic approach
  - Open, public collaboration
  - Developing interest and partnership as we go
  - Agile development with feedback loop
  - Enterprise grade software

GitHub is the heart of the project

- Code repository

- Issues

- Code Changes

- Build

- Shop window

Design philosophy

Open Metadata Access Services

Open Metadata Repository Services
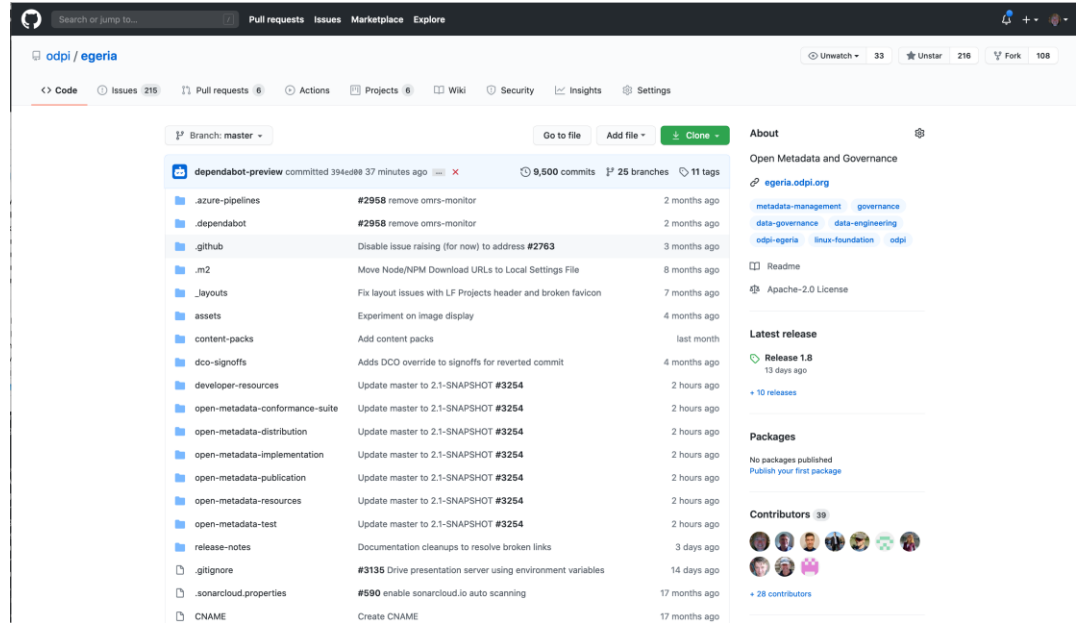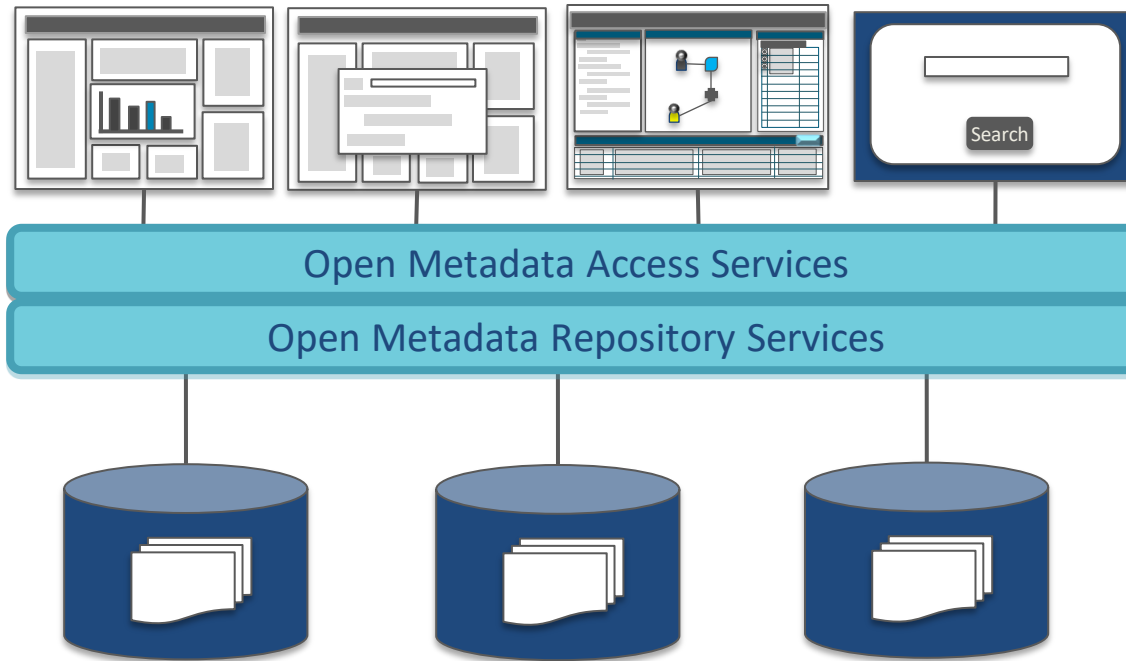
Use cases, Personas, Practitioners input

Data integration, availability and integrity best practices

Coco Pharmaceuticals persona

Jules Keeper, CDO

Tessa Tube,
Chief Researcher

Faith Broker
Chief Privacy Officr

Nancy Noah
Cloud Specialist

Erin Overview,
Information Architect

Bob Nitter,
Integration Developer

Callie Quartile,
Data Scientist

Gary Geeke
IT Infrastructure

https://egeria-project.org/practices/coco-pharmaceuticals/

Different personas need different services

Callie Quartile
Data Scientist

Find data
Understand data
Manage analytics models

Build data strategy
Define governance program
Monitor progress

Jules Keeper
Chief Data Officer

Different personas need different services

Tanya Tidie
Clinical Trials Administrator

Maintain accurate patient records
Catalog clinical trials data
Demonstrate good data management practices

Understand risks to organization
Set up protection
Monitor for suspicious activity

Ivor Padlock
Chief Security Officer

Using personas

- Open Metadata Types

- Access Service Identification

- Samples and API design
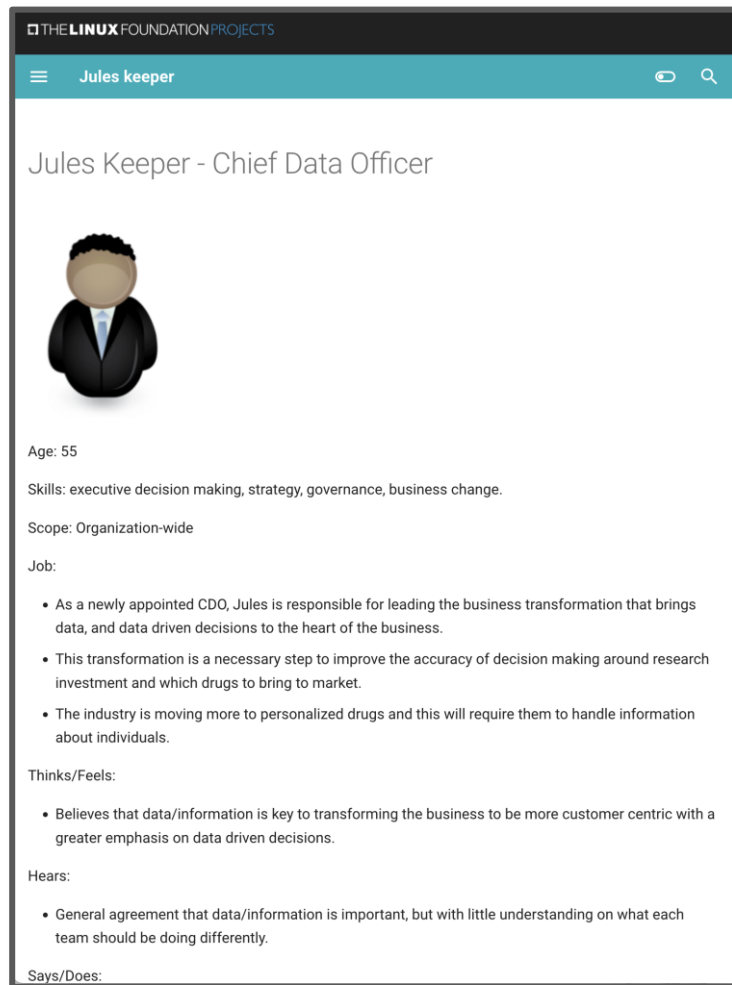
- Best Practices

Jules keeper

Jules Keeper - Chief Data Officer

Age: 55

Skills: executive decision making, strategy, governance, business change.

Scope: Organization-wide

Job:

- As a newly appointed CDO, Jules is responsible for leading the business transformation that brings data, and data driven decisions to the heart of the business.
- This transformation is a necessary step to improve the accuracy of decision making around research investment and which drugs to bring to market.
- The industry is moving more to personalized drugs and this will require them to handle information about individuals.

Thinks/Feels:

- Believes that data/information is key to transforming the business to be more customer centric with a greater emphasis on data driven decisions.

Hears:

- General agreement that data/information is important, but with little understanding on what each team should be doing differently.

Says/Does:

Use cases show the collaboration

Assign Owner

New Database

Shared Metadata

Classify Data

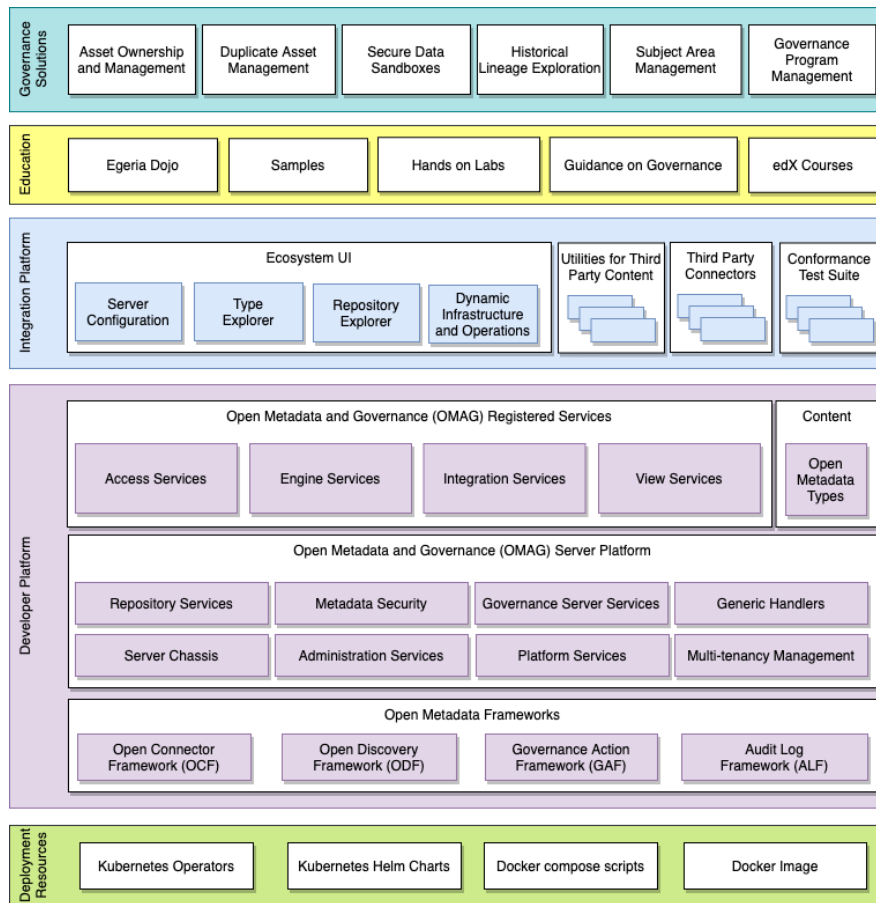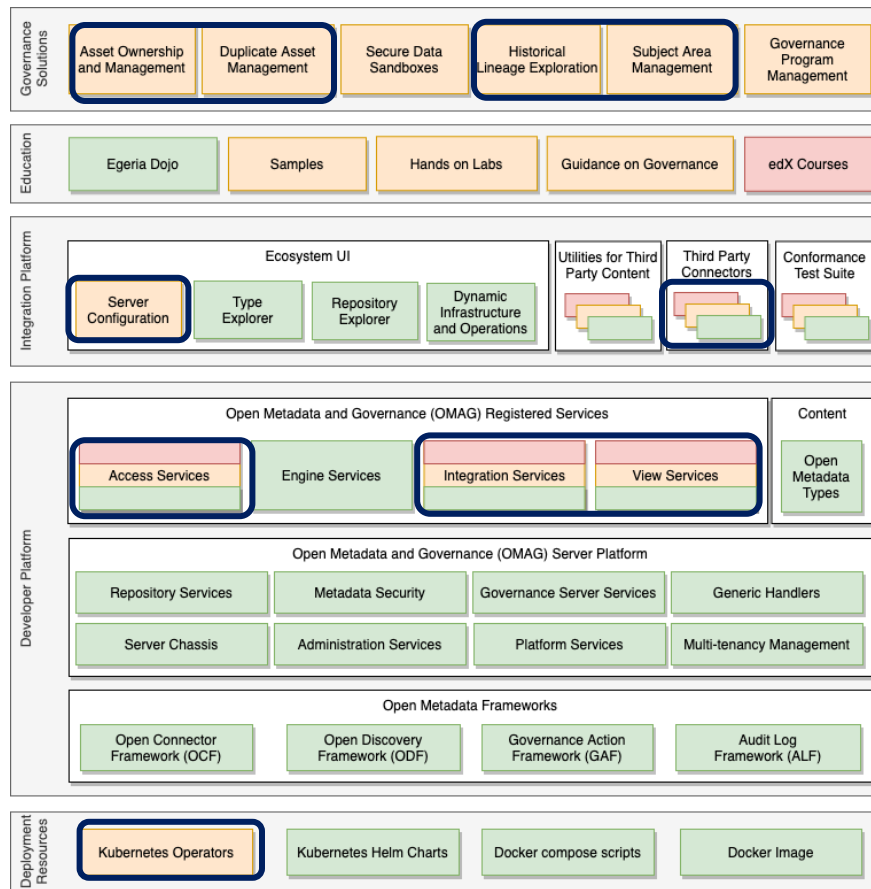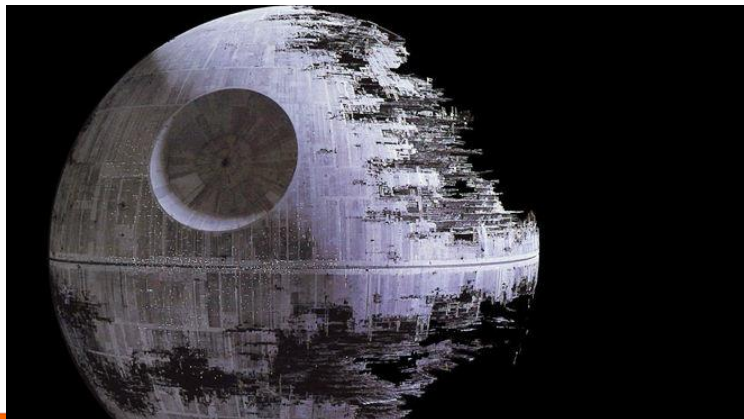Use Data
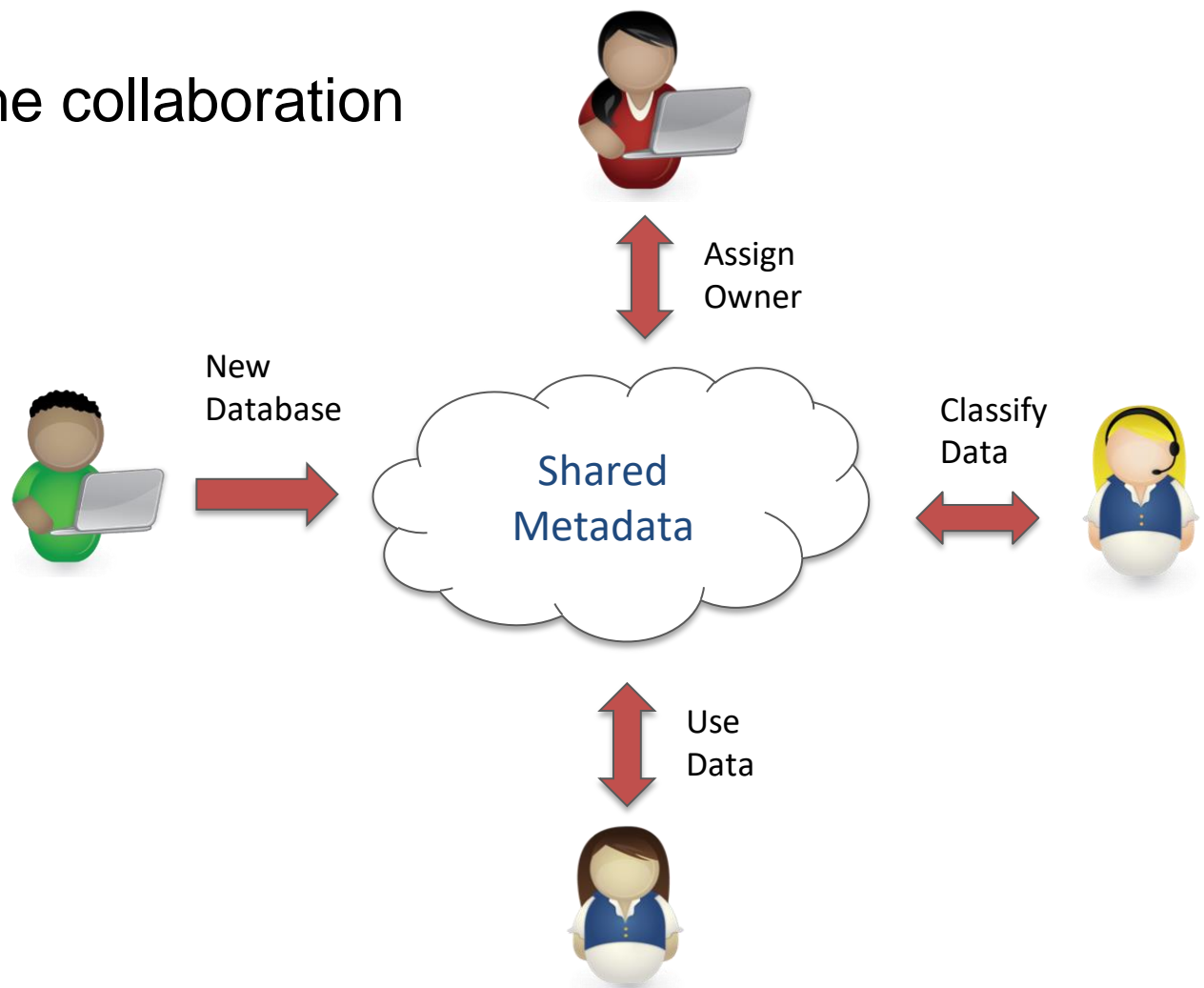
Egeria's full stack

- Egeria is more than code
  - Deployment
  - Education
  - Content packs
  - Solution cookbooks
  - Conformance test suite

Development Status

- Development in plain sight
- Monthly releases
  - What is ready, goes …
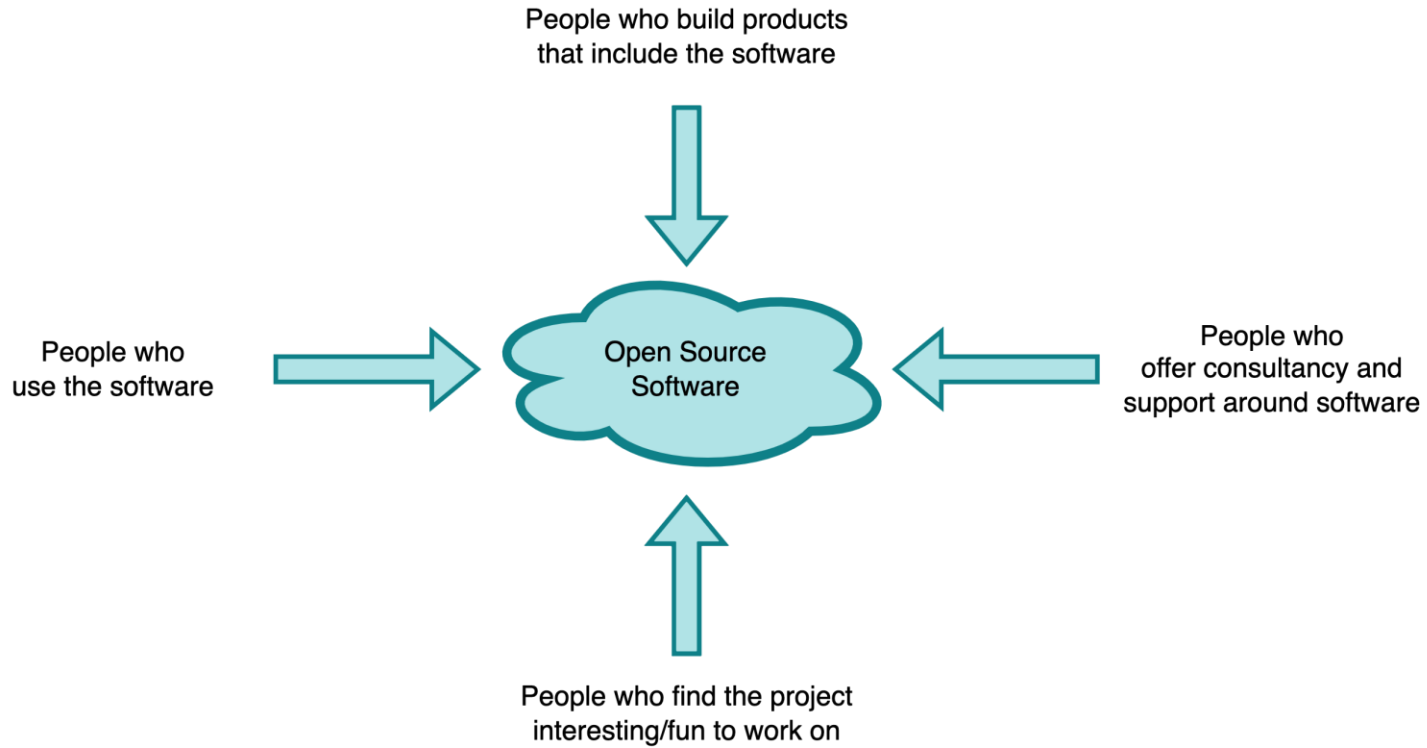
Why do vendors work with Egeria?

- Linking old platform with new platform

- Linking multiple versions of own product deployed across an enterprise

- Offload integration implementation costs

- Access to third party metadata:

  - open source technologies and other vendor's metadata

- Governing third party technologies

Open-Source Software (OSS)

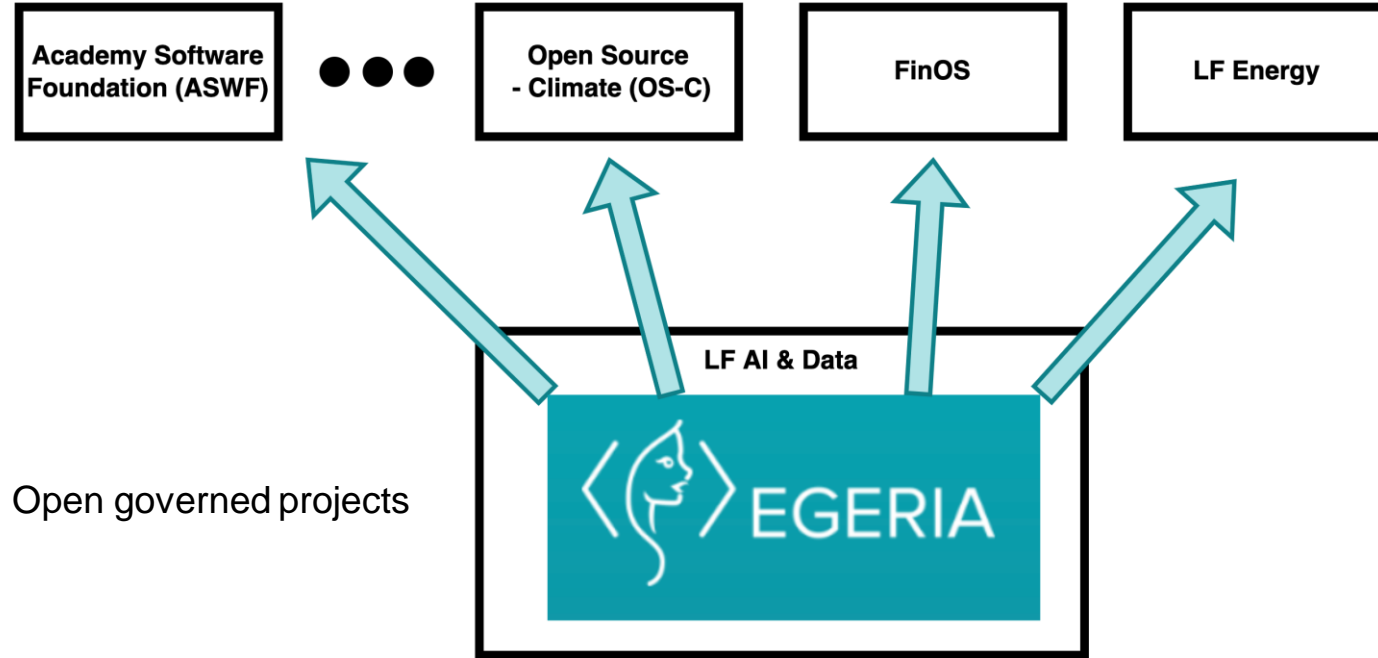# How does open-source work?

People who build products
that include the software

People who
use the software

Open Source
Software

People who
offer consultancy and
support around software

People who find the project
interesting/fun to work on

Add in the power of the open-source foundations

Open governed projects

In summary

- Software does not stand alone. It works best as part of an ecosystem.

- Open-Source Software (OSS)'s capability is rising.

- Open governance of OSS creates a more active, diverse community.

- And it is not just software …

- If you have software in your products, think of how it could be enhanced if it is part of an ecosystem.

- If Egeria is of interest, join the community …

    - https://egeria-project.org/guides/community/

Questions?

THANK YOU!

EGERIA