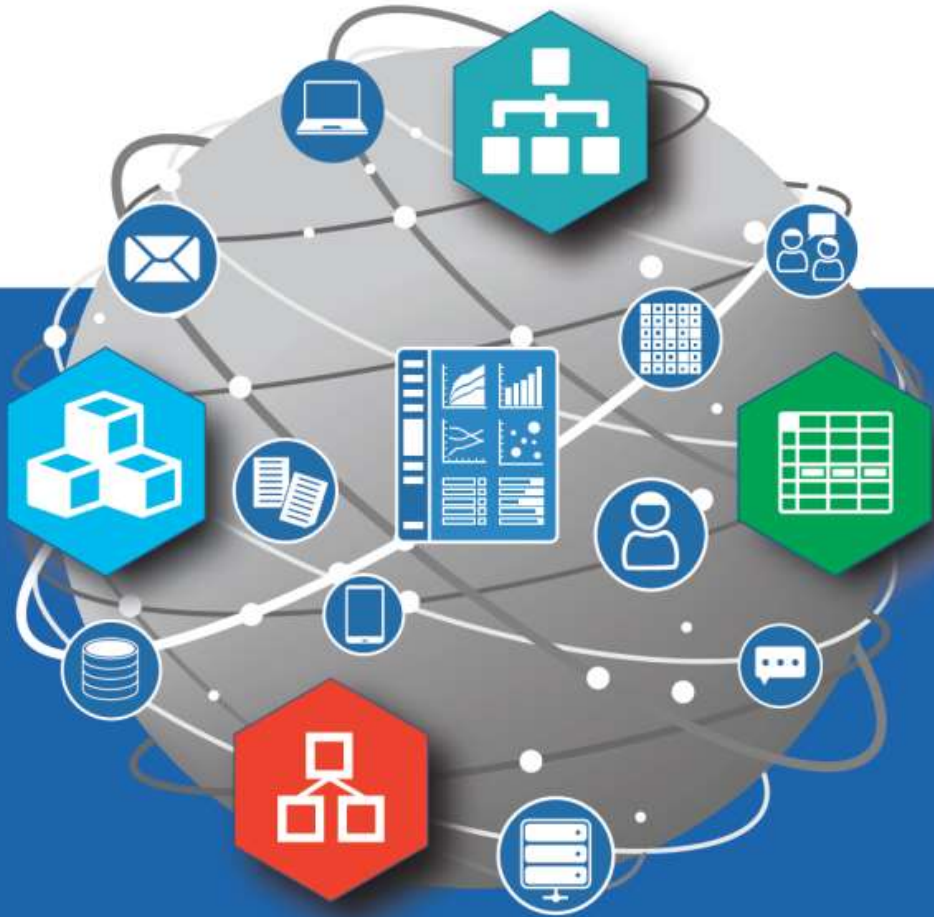




Getting Started with Data Governance – Some Practical Guidance

March 12, 2020



FOUNDATION

- TopQuadrant was founded in 2001
- Strong commitment to standards-based approaches to data semantics

MISSION

- Empower people and drive results — by making enterprise information meaningful



FOCUS

- Provide comprehensive data governance solutions using knowledge graph technologies

Today's Program



Getting Started with Data Governance – Some Practical Guidance

- Some quick observations about Data Governance
- Common Data Challenges
- How to pick your Area of Focus
- Aligning Solution Capabilities and Focus Area
- Iterative and Incremental Approach to Data Governance
- Q & A



Irene Polikoff

What is Data Governance?

There are Many Definitions / Descriptions of Data Governance

- “Data Governance is a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods.”
... DGI Framework, David Plotkin, Wikipedia
- "Delivering trusted information for smarter business decisions across your entire information supply chain" ... IBM
- “Data governance is the universal approach to managing and securing data in organizations”
... DGPO webinar

What is Data Governance?

“Data governance is a term without an authoritative definition.”

... Daniel Howard, Bloor Research, January 2020

“Data governance encompasses the people, processes, and information technology required to create a consistent and proper handling of an organization's data across the business enterprise.”

Source: Data Governance - Wikipedia

Having Difficulty Getting Started?

- Relax
- You already did!
- You are managing data today
- Can you do it better?
- What causes you to ask the question?



Common Data Challenge I

I. We don't know where all the data resides, or how it is used

- We may be maintaining duplicate data
- We may have critical data that stakeholders don't know about and, thus, can't use



Common Data Challenge II

II. It takes us too long to get data when we need it

- We don't have the right tools, processes or people to reliably locate, request or provide data



Common Data Challenge III

III. Our data isn't of high enough quality

- We know of specific problems with data quality
- We can't trust the data that comes out of our systems in reports – adding or aggregating data that is not compatible

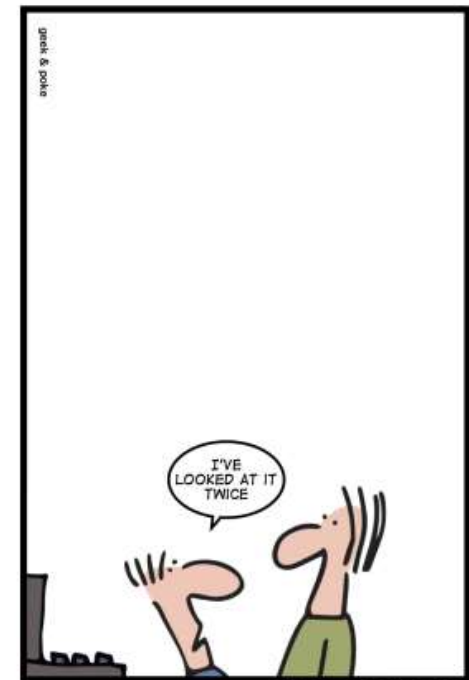


Common Data Challenge IV

IV. We have low data competency

- Business users don't know enough about data to make informed use of it
- Technical users don't know enough about business to advise business users
- Lack a common business terminology for data negatively impacts reporting and analysis

SIMPLY EXPLAINED



4 EYES PRINCIPLE

Common Data Challenge V

V. Data is not readily available to consumers

- Access is blocked
- Formats are not appropriate for intended use



Common Business Drivers

- New Regulations
- Competitive Pressures
- Merger or Acquisition
- New Product Offering
- Growing Complexity of Data Ecosystem
 - e.g., New Data Sources

Why the Need to Address these Data Challenges Now?

Important: Pick a Focus Area

- You will not be able to address all problems at once
- Pick one to start with
- Working towards too many goals dilutes your efforts
- Pick a business/organization area to work with
 - Where the need is strong
- Expand out after success

Solution Areas and Capabilities

Reference Data Management

Reference Datasets

Enumerations

Crosswalks

Metadata Management

Data Systems Inventory

Business Activities

Business Glossary

Data Lineage
(at Business and Logical Levels)

Data Dictionary / Catalog

Data Quality

Vocabulary Management

Taxonomies

Content Corpora

Content Tagging

Supported by an integrated information management / collaboration environment with search, roles, access control, etc.

Auto-Crosswalking

Automation

Auto-Tagging

Automapping Data
Elements to Terms

Term Extraction

Rule Extraction

Governance Model/Framework

Policies

Workflows

Metrics

Problem → Solution

I. Don't know where all the data resides, or how it is used

Data Systems Inventory

Data Lineage

II. It takes too long to get data when we need it

Data Systems Inventory

Data Request Workflow

III. Data isn't of high enough quality

Reference Data Management

Data Quality

IV. Low data competency

V. Data is not readily available to consumers

Problem → Solution

Data Systems Inventory

Business Glossary

Data Dictionary / Catalog

Data Request Workflow

Data Dictionary / Catalog

Data Security and Privacy Policies

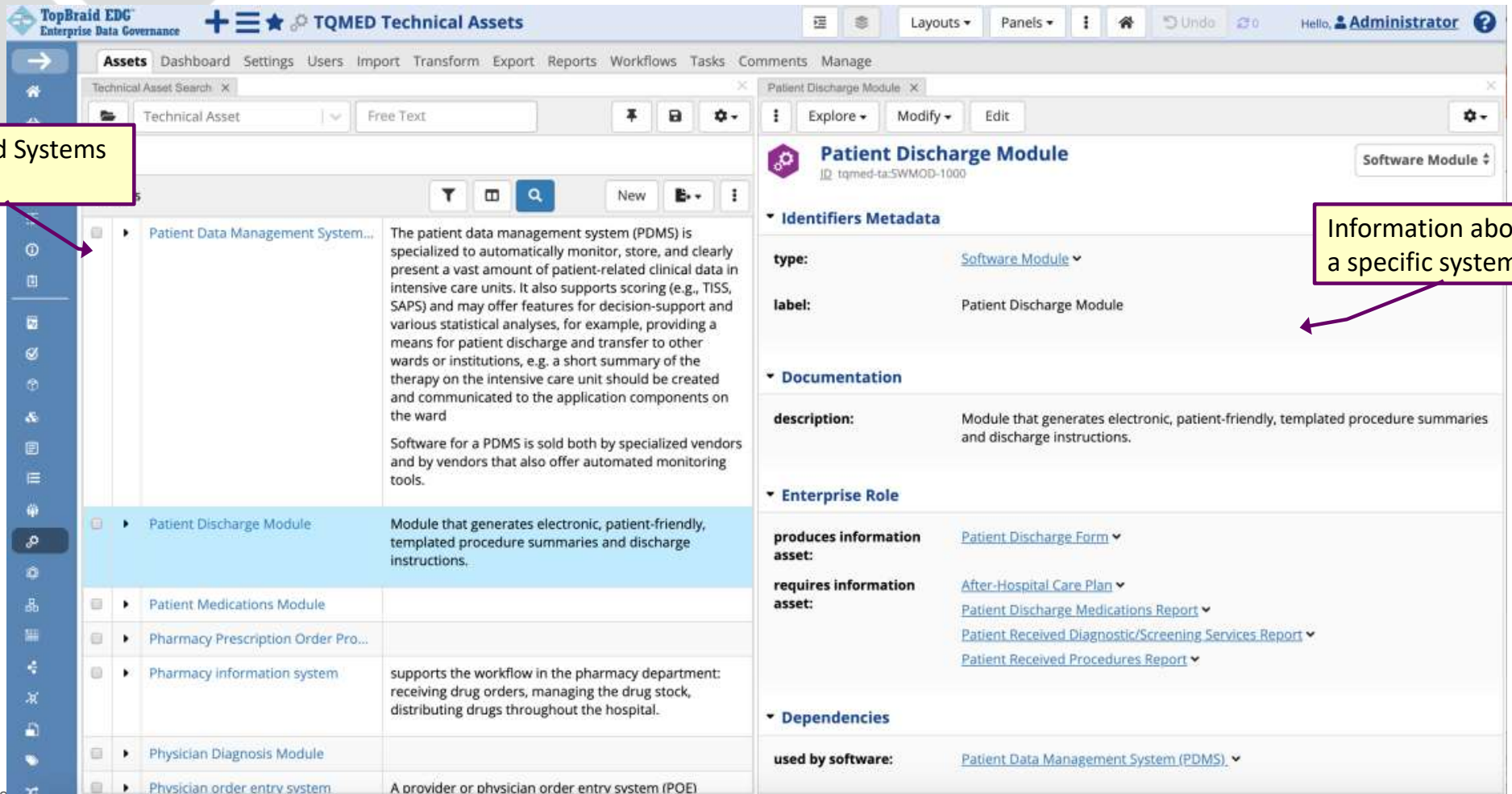


An Example Starting Point

- I. Don't know where all the data resides, or how it is used

Data Systems Inventory

Example Data Systems Inventory

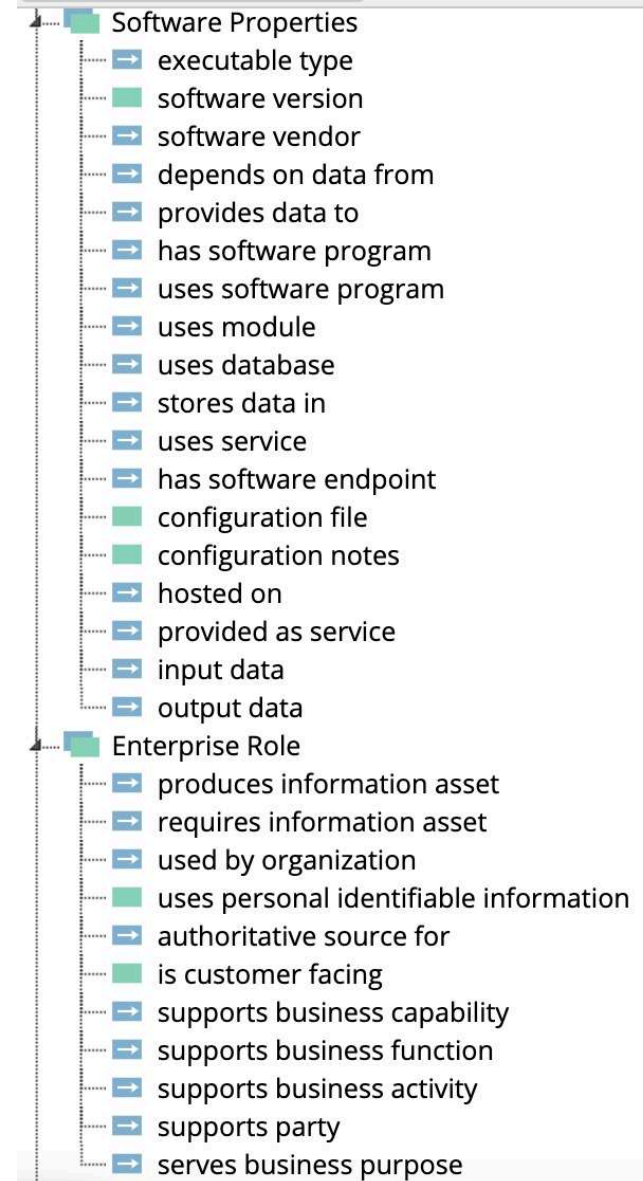
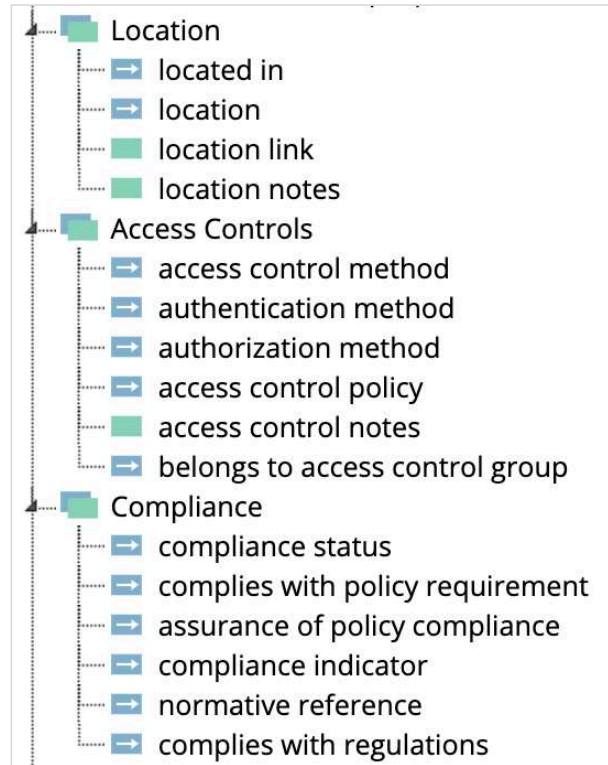
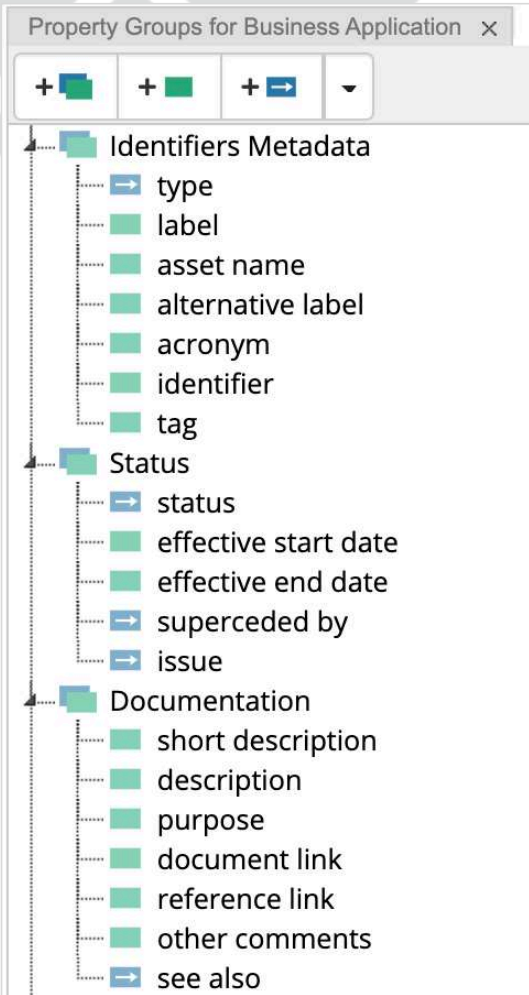


The screenshot displays the TopBraid EDG Enterprise Data Governance interface. The main window shows a list of technical assets under the heading 'TQMED Technical Assets'. A yellow callout box labeled 'Cataloged Systems' points to the left sidebar navigation menu. The 'Patient Discharge Module' is selected and highlighted in blue. A detailed view of this module is shown on the right, with a yellow callout box labeled 'Information about a specific system' pointing to the 'type' field.

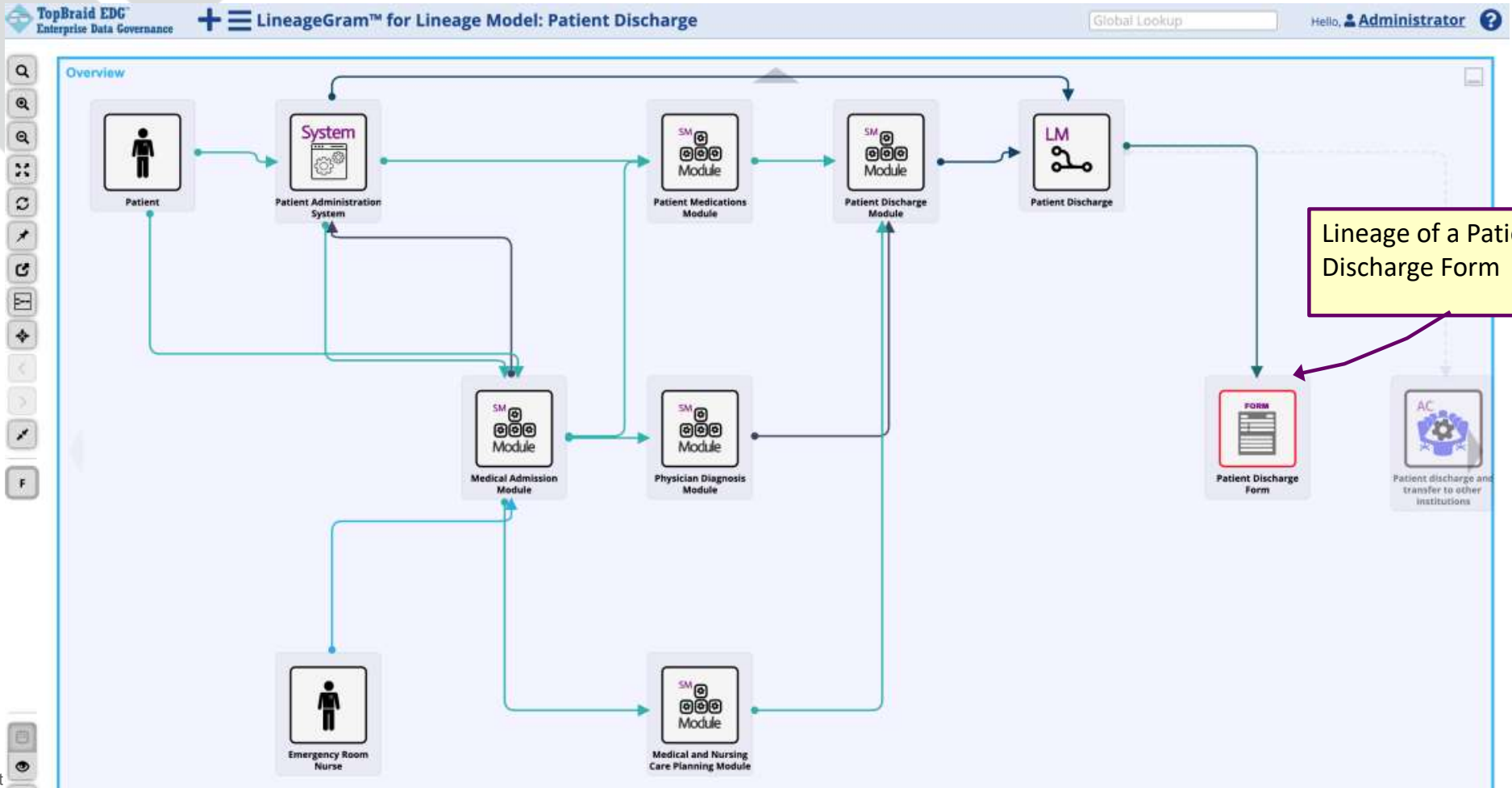
Asset Name	Description
Patient Data Management System...	The patient data management system (PDMS) is specialized to automatically monitor, store, and clearly present a vast amount of patient-related clinical data in intensive care units. It also supports scoring (e.g., TISS, SAPS) and may offer features for decision-support and various statistical analyses, for example, providing a means for patient discharge and transfer to other wards or institutions, e.g. a short summary of the therapy on the intensive care unit should be created and communicated to the application components on the ward. Software for a PDMS is sold both by specialized vendors and by vendors that also offer automated monitoring tools.
Patient Discharge Module	Module that generates electronic, patient-friendly, templated procedure summaries and discharge instructions.
Patient Medications Module	
Pharmacy Prescription Order Pro...	
Pharmacy information system	supports the workflow in the pharmacy department: receiving drug orders, managing the drug stock, distributing drugs throughout the hospital.
Physician Diagnosis Module	
Physician order entry system	A provider or physician order entry system (POE)

Patient Discharge Module	
Identifiers Metadata	
type:	Software Module
label:	Patient Discharge Module
Documentation	
description:	Module that generates electronic, patient-friendly, templated procedure summaries and discharge instructions.
Enterprise Role	
produces information asset:	Patient Discharge Form
requires information asset:	After-Hospital Care Plan Patient Discharge Medications Report Patient Received Diagnostic/Screening Services Report Patient Received Procedures Report
Dependencies	
used by software:	Patient Data Management System (PDMS)

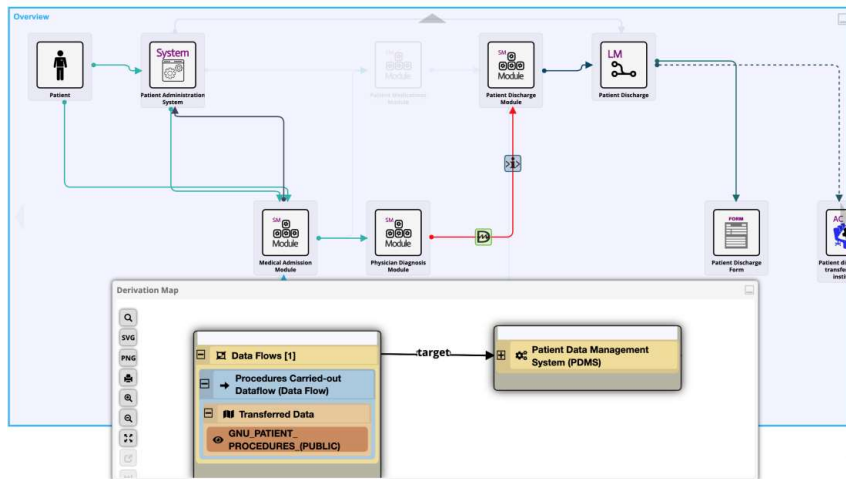
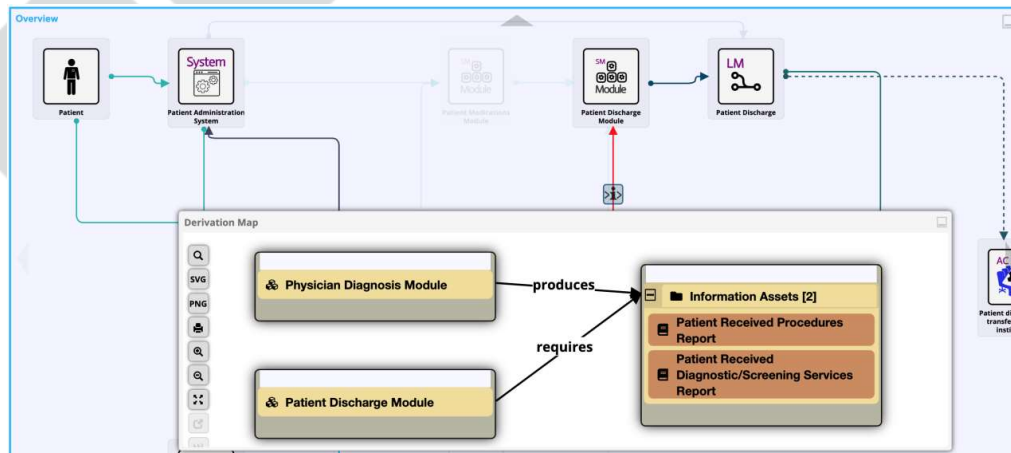
What Type of Information Should be Captured?



What Can You Get As a Result?

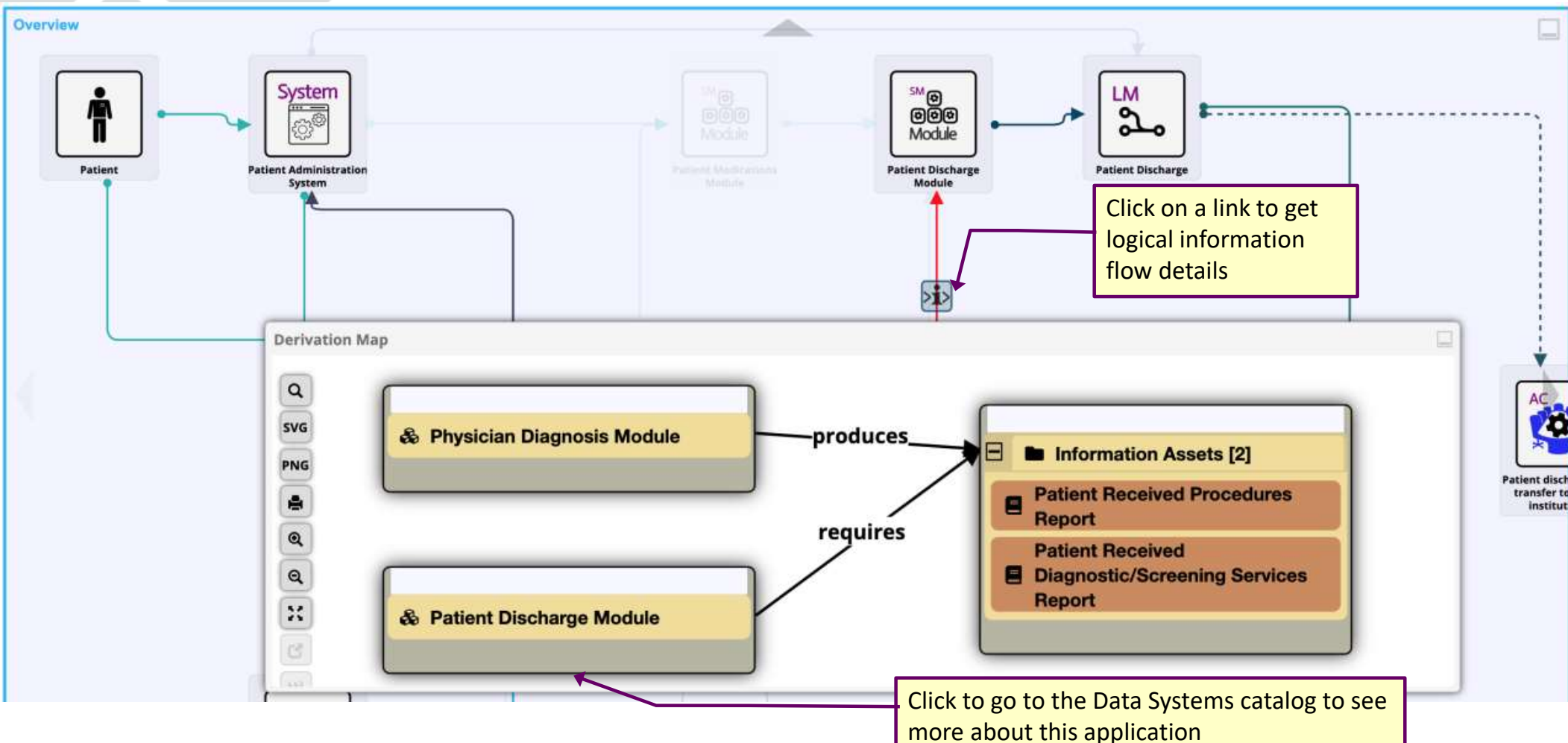


Different Levels of Lineage: Logical vs Physical

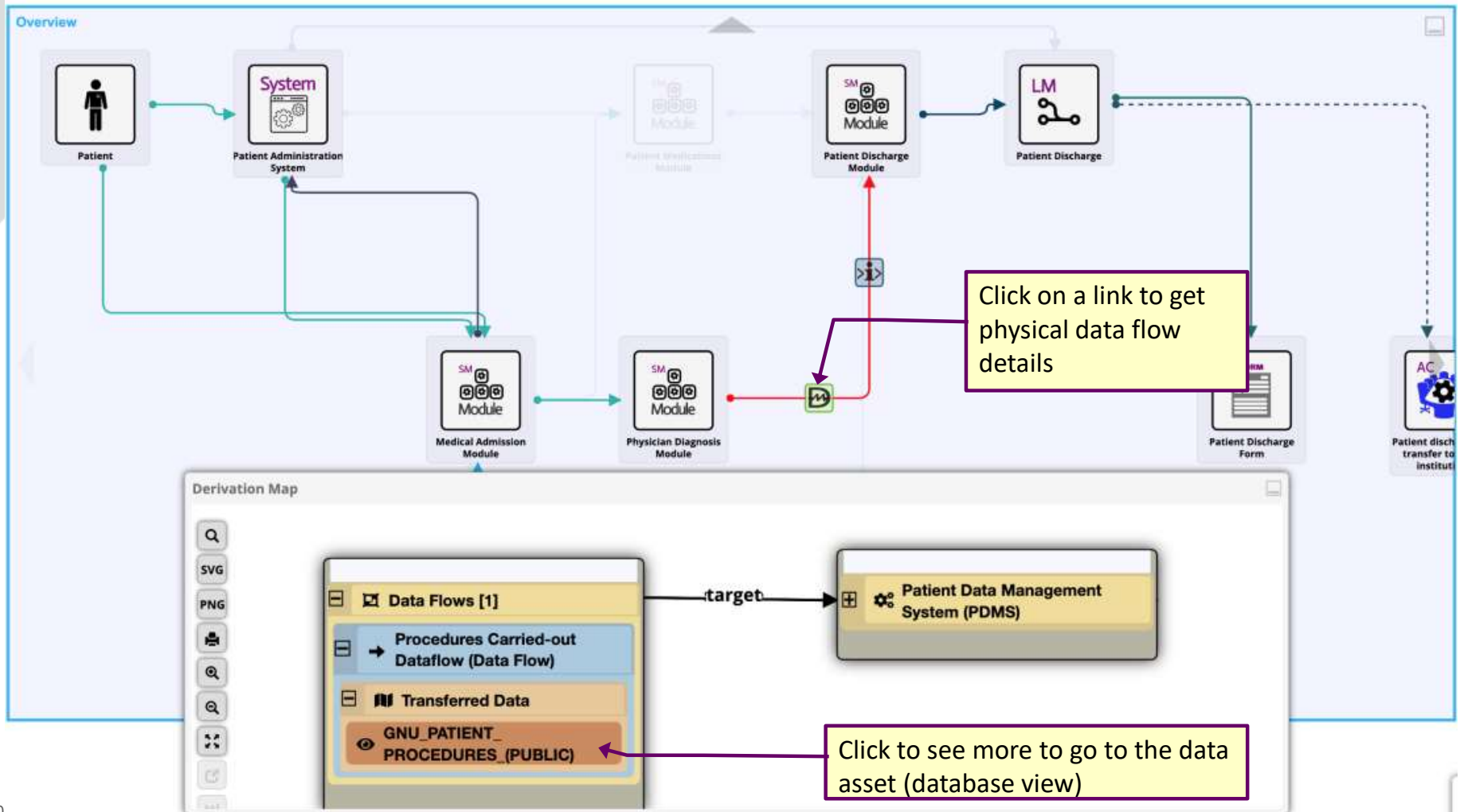


- Links are clickable and provide different level of details
- Logical – higher level, more business user friendly
- Physical – developer or data analyst level, requires data cataloging

Logical: Information Flow Example



Physical: Data Flow Example



Example of the Information Required for the Physical Lineage

GNU_PATIENT_PROCEDURES_(PUBLIC)
URI http://example.org/technical-assets-collections/TQMED_Technical_Assets#DBVIEW-1000

Identifiers Metadata

type: Database View

label: GNU_PATIENT_PROCEDURES_(PUBLIC)

asset name: GNU_HEALTH_PATIENT_PROCEDURES

Table/View Properties

table kind: Operational Data

record count: 3256

view of: PUBLIC

view output column: CODE (PUBLIC.GNU_PATIENT_PROCEDURES), DATE (PUBLIC.GNU_PATIENT_PROCEDURES), DESCRIPTION (PUBLIC.GNU_PATIENT_PROCEDURES), PATIENT (PUBLIC.GNU_PATIENT_PROCEDURES)

view input column: CODE (PUBLIC.GNUHEALTH_LAB_TEST_TYPE), CODE (PUBLIC.GNUHEALTH_SURGERY), DESCRIPTION (PUBLIC.GNUHEALTH_SURGERY)

Data Sample

CODE	DATE	DESCRIPTION	PATIENT
K35.80	2014-07-06T23:55:57.619Z	APPENDECTOMY	1
R77	2014-07-06T08:45:57.518Z	COMPLETE BLOOD COUNT	1
R77	2014-07-05T10:05:20.432Z	COMPLETE BLOOD COUNT	2
K80.42	2014-07-05T13:30:10.72Z	CHOLECYSTECTOMY	2
R82	2014-07-06T08:36:20.22Z	URINE ANALYSIS	1

1 to 5 of 5 entries

[Previous](#) [Next](#)

Data Quality Records

Data Quality Records

0 rows

Data Quality Record	Created	Data Quality Status	Quality Issues	Verified

Patient Procedures Database View used in the Dataflow

Database Sample

Data Statistics

Database Columns

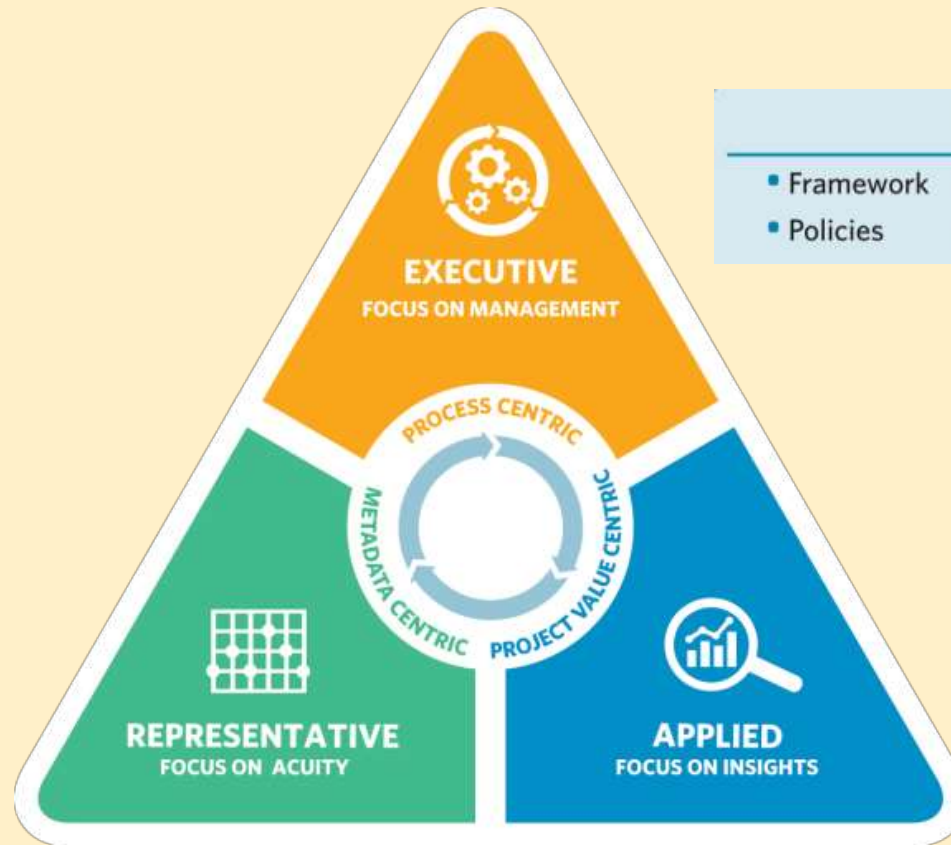
Data Governance Triangle

Common Primary Modes of DG Operation

- Process Centric
- Project Value Centric
- Metadata Centric

Executive Governance:

- Framework
- Methodology
- Stewardship
- Process
- Policies
- Metrics
- Maturity Assessment



Applied Governance:

- Lineage
- Impact Analysis
- Metrics
- Data Quality
- Compliance Reporting

Representative Governance:

- Glossaries
- Taxonomies
- Ontologies
- Data
- Technical and Enterprise Assets
- Reference Data

Installed Workflow Templates

Instances of Workflow template (for Working Copies)

Refine: Excel CSV Print

Workflow template (for Working Copies)

- [+ Add or Edit Asset](#)
- [+ Assign Business Term](#)
- [+ Basic Review Workflow](#)
- [+ Basic workflow](#)
- [+ Data Quality Review](#)
- [+ Delete Asset](#)
- [+ Issue Management](#)
- [+ Issue Resolution](#)
- [+ No Review Workflow](#)
- [+ Propose Business Term](#)
- [+ Request Access](#)
- [+ Review With Escalation Workflow](#)
- [+ Stakeholder Meeting Review](#)
- [+ Version Increment Workflow](#)
- [+ Voting Workflow](#)

1 to 15 of 15 entries 1 row selected Previous Next

Basic workflow Print

Basic workflow

The workflow template consists of the following stages.

▶ **Uncommitted** (editable)
This is the initial status.

It means that changes are being made or are expected to be made. Any change created so far is "work in progress" and has not been written to the production copy.

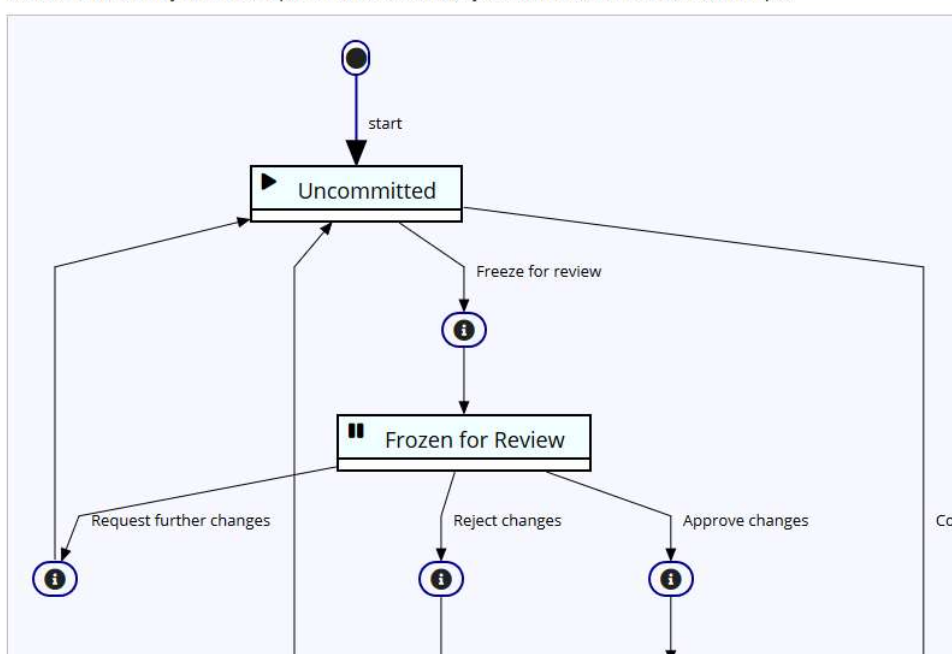
Possible Transitions:

- To Committed, The following conditions have to be satisfied:
 - Required permission for the production copy: editor [teamwork:editor]
- **State Entry Action**
Commit the workflow to production
- To Frozen for Review, The following conditions have to be satisfied:
 - Working copy requires changes: true
 - Required permission for the working copy: manager

▶ **Approved** (not editable)
It means that the changes have been approved.

Possible Transitions:

Workflow transitions are controlled using the diagram below or the text next to it. Rectangular shapes in the diagram depict workflow states. State transitions are depicted as clickable oval shapes, positioned over arrows connecting states. Clicking on a transition shape in the diagram or on an icon in the text next to the diagram, will move the workflow into the next state. The current state and possible transitions from it are highlighted in green. If you have permissions to execute a transition, you will see a person icon connected by a dotted arrow to the transition shape.



```

graph TD
    Start((start)) --> Uncommitted[▶ Uncommitted]
    Uncommitted --> Frozen[Frozen for Review]
    Frozen --> Uncommitted
    Frozen --> Request((Request further changes))
    Frozen --> Reject((Reject changes))
    Frozen --> Approve((Approve changes))
    Request --> Uncommitted
    Approve --> Approved[▶ Approved]
    
```

Governance Roles

Patient Discharge Module
ID: tqmed-ta:SWMOD-1000

Stewardship

▼ Governance Roles

- data steward:
- business steward:
- technical steward:
- application steward: Irene.Polikoff (User)
- data governor:
- data producer:
- data consumer:
- subject matter expert:
- responsible:
- accountable:
- consulted:
- informed: Jesse.Lambert (User)
Robert.Coyne (User)

Stewardship View
of an asset

A useful role list to get
started, but should be
adjusted based on the
decided practices

Governance Subject Areas and Roles

TopBraid EDG Enterprise Data Governance | Governance Model | Global Lookup | Hello, Administrator

Navigation: Governance Areas | Organizational Structure | Metrics | Dashboards | Issues | Policies | Other Governance Assets | Dashboard | Settings | Users | Import | Transform | Export | More...

Governance Areas – business areas, data domains

Manage Your Governance Areas and Roles

Organize your asset collections into **business areas** and **data subject areas**. Enable **governance roles** for these areas to assign responsible parties.

Buttons: Add Data Subject Sub-Area | Add governance role...

Governance Role Assignment

Governance capabilities: Metrics, Issues, Policies, etc.

- Business & Data Subject Areas
 - Hospital Management
 - Facilities and Equipment Data Domain
 - Financial Data Domain
 - HR Data Domain
 - Pharmacy Data Domain
 - Patient Care
 - Ambulatory Care Data
 - Clinical Trials Data Domain
 - Electronic Health Record**
 - Business Steward (2)
 - Technical Steward (2)
 - Reference Data Management

Electronic Health Record (Data Subject Area)

Information about patient collected across healthcare information systems.

Data includes demographics, medical history, medication and allergies, immunization status, laboratory test results, radiology images, vital signs, personal statistics like age and weight, and billing information.

Buttons: Details | Delete

Include all assets included in this Subject Area in the index for Search the EDG

Asset Collections

Create new Choose type... in this data subject area

Add Existing Choose type... in this data subject area

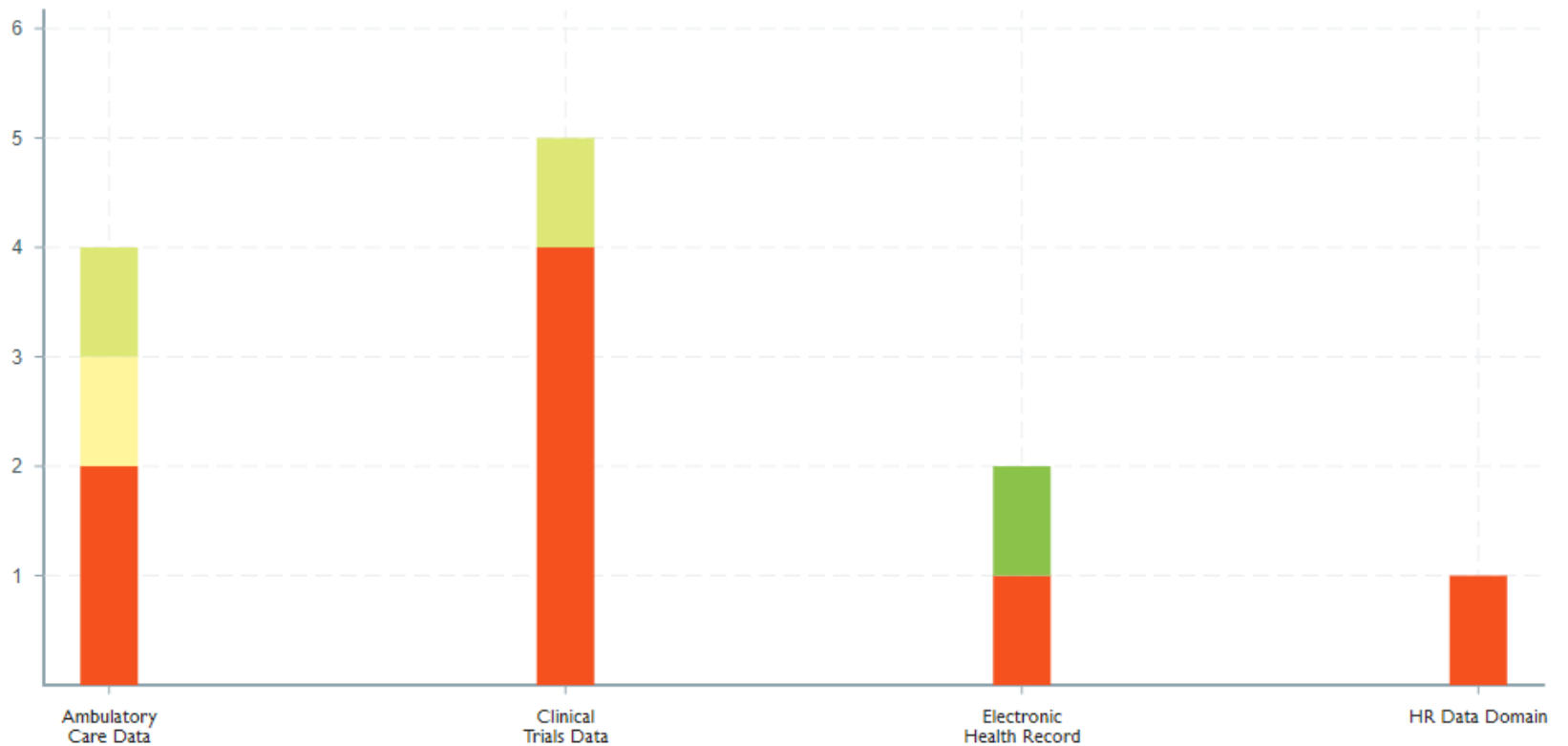
Name	Workflows	+Roles	Remove
OpenMRS Data Catalog	0	None	Remove
GNU Health Database Catalog	0	None	Remove
Patient Information Glossary	0	None	Remove
NUCC Healthcare Provider Taxonomy	0	None	Remove

Asset collections in an area

Governance Role Assignments

Governance Role Assignments by Subject Area

Role Assignments across Subject Area

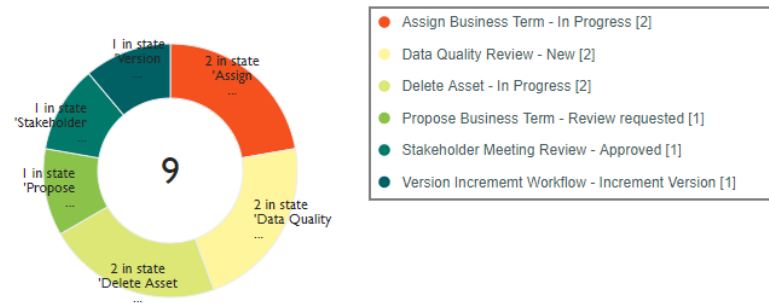


Metrics on Running Workflows

Workflows per Subject Area

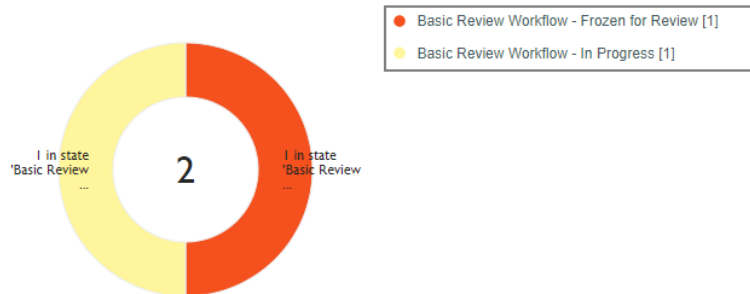
Ambulatory Care Data

My Asset Collections: [GNU Health Data Catalog](#), [NUCC Healthcare Provider Taxonomy](#), [OpenMRS Data Catalog](#), [Patient Information Glossary](#)
 My roles: Business Steward



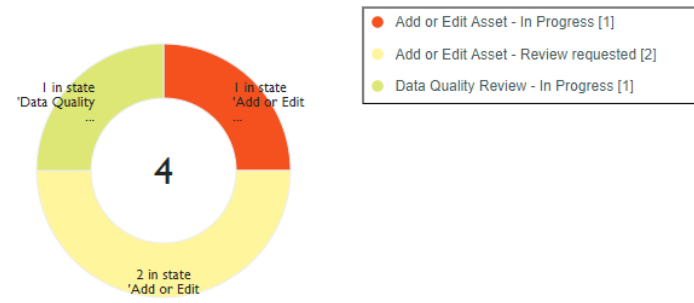
Electronic Health Record

My Asset Collections: [Country Codes](#), [Patient Release Glossary](#)
 My roles: Business Steward



Clinical Trials Data Domain

My Asset Collections: [Clinical Trials Datasets](#), [ICD-10](#), [ICD-10 to NDC crosswalk](#), [National Drug Code \(NDC\)](#)
 My roles: Business Steward



Search Access to All Assets

Faceted filters

Link to details view

Links to LineageGram, NeighborGram traceability maps

Link to Comments

Graph	>
Type	▼
Concept	150
Glossary Term	74
Business Term	24
Database Column	4
Business Activity	2
System	2
String Data Element	2
Report	1
Form	1
Numeric Data Element	1
Show entries 11 to 13 of 13	
Has Broader	>
Column Of	>
Physical Datatype	>
Sub-Activity Of	>
Produces Information Asset	>
Requires Information Asset	>
Has Exact Match	>
Has Related	>

< Previous | 1 - 10 of 202 | Next >

10 per page

DIAGNOSIS (PUBLIC.GNUHEALTH_LAB)

An [Database Column], [String Data Element] from GNU Health Datasource

(0) Comments



INFO DIAGNOSIS (PUBLIC.GNUHEALTH PATIENT EVALUATION)

An [Database Column], [String Data Element] from GNU Health Datasource

(0) Comments



Diagnosis Related Groups (DRGs)

An inpatient or hospital classification system used to pay a hospital or other provider for their services and to categorize illness by diagnosis and treatment. A classification scheme used by Medicare that clusters patients into 468 categories on the basis of patients' illnesses, diseases and medical problems. Groupings of diagnostic categories drawn from the International Classification of Diseases and modified by the presence of a surgical procedure, patient age, presence or absence of significant comorbidities or complications, and other relevant criteria. System involving classification of medical cases and payment to hospitals on the basis of diagnosis. Used under Medicare's prospective payment system to reimburse inpatient hospitals, regardless of the cost to the hospital to provide services.

An [Glossary Term], [Concept] from Healthcare Taxonomy

(0) Comments



Principal Diagnosis

The medical condition that is ultimately determined to have caused a patient's admission to the hospital. The principal diagnosis is used to assign every patient to a diagnosis related group. This diagnosis may differ from the admitting and major diagnoses.

An [Glossary Term], [Concept] from Healthcare Taxonomy

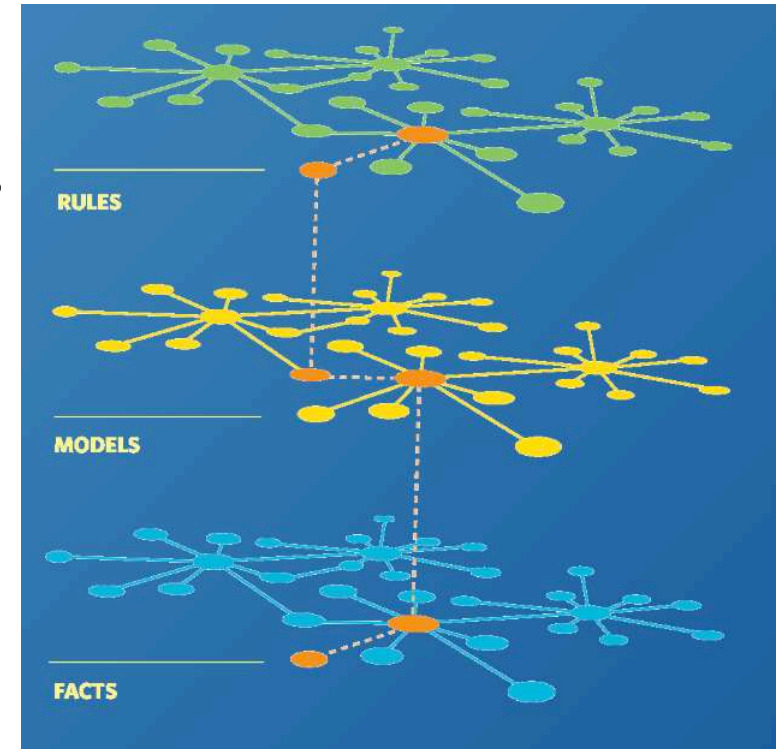
(0) Comments



Benefits of a Knowledge Graph based Platform for Data Governance 2.0

TopBraid Enterprise Data Governance (EDG):

- Is flexible and extensible, based on standards
- Integrates reasoning and machine learning
- Enables people (UI) and software (APIs/web services) to view, follow and query
- Bridges data and metadata “silos” for a seamless data governance
- Delivers Knowledge-driven data governance



As an enterprise knowledge graph infrastructure, TopBraid EDG supports Data Governance 2.0 and applications of AI / ML

Thank You !



... Questions?



To Learn More about TopBraid EDG and Knowledge Graphs:

EDG Product Info:

- <https://www.topquadrant.com/products/topbraid-enterprise-data-governance/>

Contact us: at info@topquadrant.com to:

- Discuss vocabulary management solutions (glossaries, taxonomies, ontologies)
- Request a more targeted demo of TopBraid EDG – Vocabulary Management
- Ask for a free EDG evaluation account



More Resources ...

More Webinar Recordings, Slides, Q&A:

- <https://www.topquadrant.com/knowledge-assets/topquadrant-webinars/>

Short Videos:

- TopBraid EDG “Quick Grok” Videos
<https://www.topquadrant.com/knowledge-assets/videos/>
- TopBraid EDG Animated Video
https://www.topquadrant.com/project/edg_agile_modular/

Blog:

- <https://www.topquadrant.com/the-semantic-ecosystems-journal/>

Data Governance White Papers

- <https://www.topquadrant.com/knowledge-assets/whitepapers/>