Going Live with Enterprise Solutions

TopQuadrant™

FIBO in SHACL: The Next Step in Data Validation and Interoperability

Irene Polikoff, TopQuadrant

EDW, San Diego, April 2018

TopQuadrant™

- CEO and co-founder at TopQuadrant
- W3C SHACL Working Group co-chair

Irene Polikoff

- SHACL Working Group
  - Was charted in October 2015
  - Successfully completed its charter in July 2017
  - Currently, in the “maintenance mode”

What is SHACL?

- SHApes Constraint Language
  - W3C Recommendation to define the “shape” of data
  - Shapes are defined by constraints and validation rules
  - Shapes provide fascilities for input checking and reporting
  - By design, a schema language for RDF & LinkedData (the W3C graph based, or network-of-nodes, data model)

- SHACL Shapes:
  - Shapes for node types reside in a “Shape Library”
  - Shape “inheritance” down the node tree
  - Shapes can be extended, reused, and shared

SHACL's Role in the Sematic Web Stack

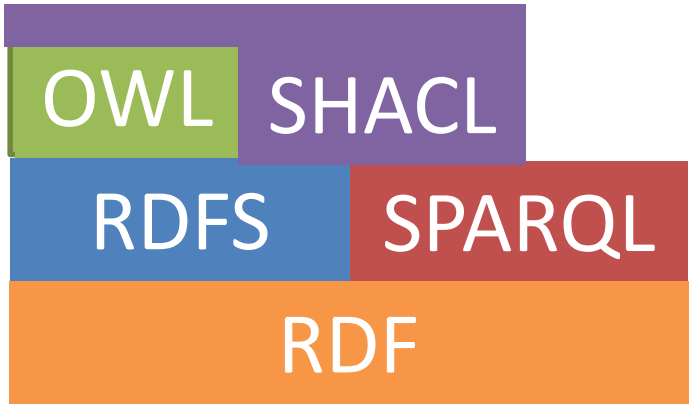
*W3C = World Wide Web Consortium led by Tim Berners-Lee

Classification:

What is the type of this thing?

Rules:

Is that term used correctly?
You can't say that here!
What can you infer?



Vocabulary:

Shared terms can we use

Query:

What did you say?

Statements:

Saying things

Why talk about SHACL at EDW?

(Knowledge) graphs are powerful for enterprise data integration and reuse

SHACL ensures data quality and enhances the data

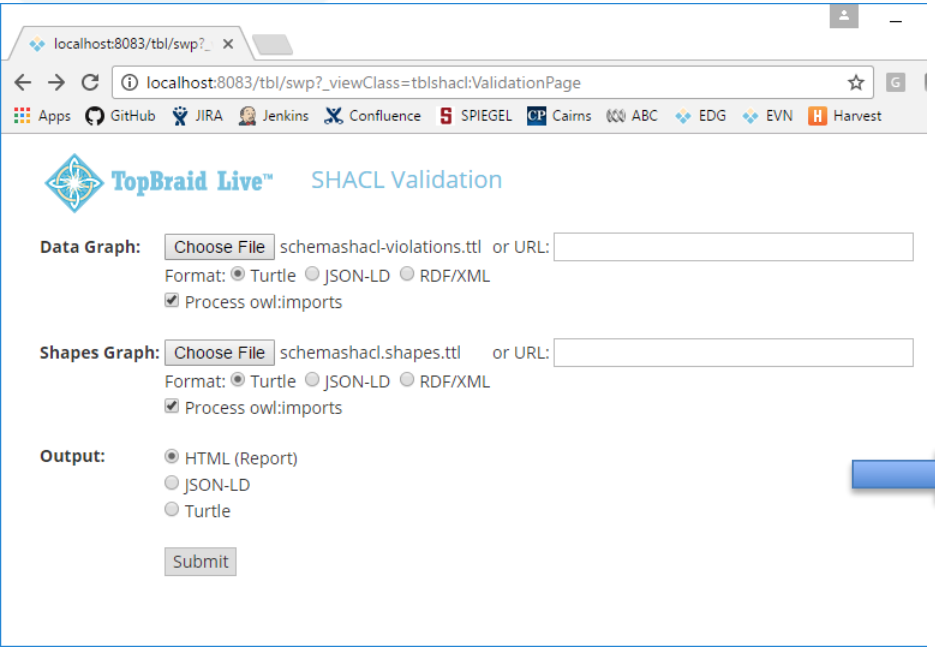
Strong adoption of SHACL is happening – including in Financial Services

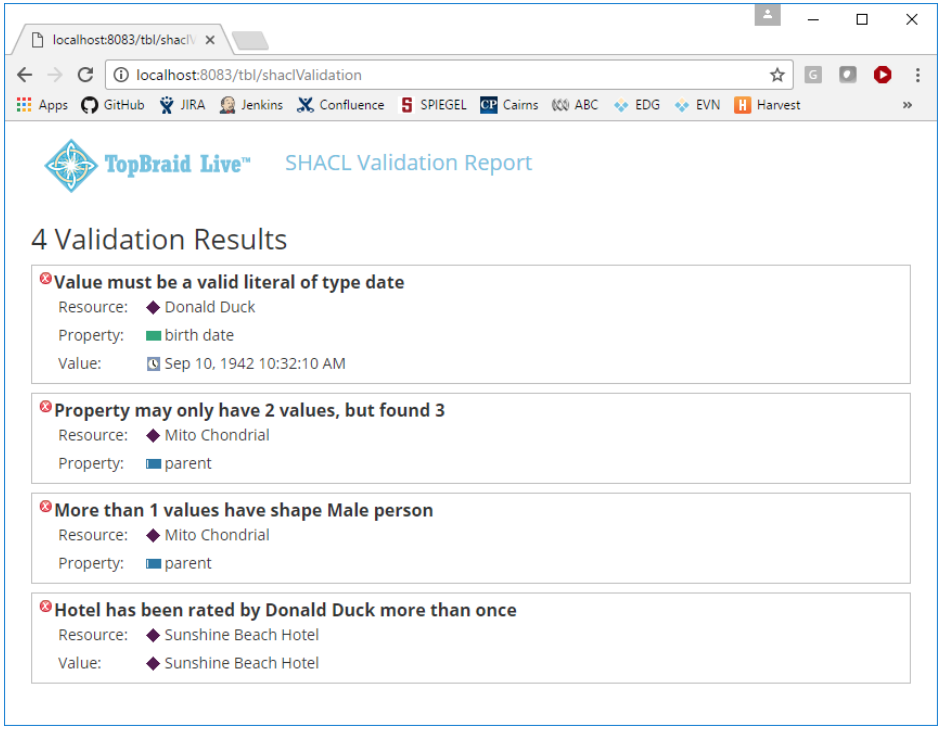
Motivation for SHACL at W3C

- SHACL is rooted in the world of the Semantic Web W3C standards
  - Until SHACL no official W3C recommendation for validating RDF
- RDF Schema / OWL are “schema languages”, however...
  - Descriptive, not prescriptive or restrictive
  - OWA, CWA, UNA ⇒ classification inferencing not validation
  - Even for inferencing/classification, OWA is often a problem - See using SHACL for classification rules: https://henriettеharmse.com/2018/03/15/classification-with-shacl-rules/
- W3C member submissions (semi-official)
  - SPIN, Resource Shapes, ShEx
  - SHACL unifies the best features of each to cover most use cases
- Other Standalone / ad-hoc solutions have also evolved
  - Most on top of SPARQL & RDFS/OWL
- SHACL gives us a vendor neutral solution and a governing body
  - A standard on par with XML, HTML, RDF, etc.

OWA: Open World Assumption
CWA: Closed World Assumption
UNA: Unique Name Assumption

SHACL Ensures conformance of RDF data to a defined schema





- Not conforming data is reported in the "Validation Report" – report structure is defined in the SHACL standard

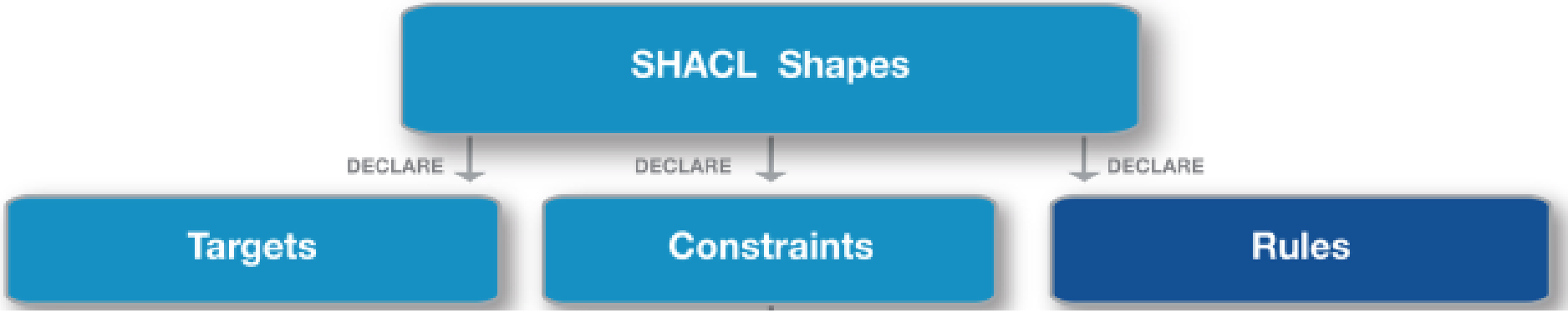
SHACL and Rules

- SHACL let's you define a "shape" data should adhere to
  - This is effectively a "data quality rule" – if condition is not met, SHACL engine reports an exception
- SHACL also let's you 'infer' new data if condition is met
  - This is a more general data or business rule – if condition is met, SHACL engine can add new fact(s)

What else SHACL can be used for

- Extended, but not limited to, goals:
  - interface building
  - data structure communication
  - code generation
  - data integration
  - rule-based inferencing
  - …

High Level View of SHACL

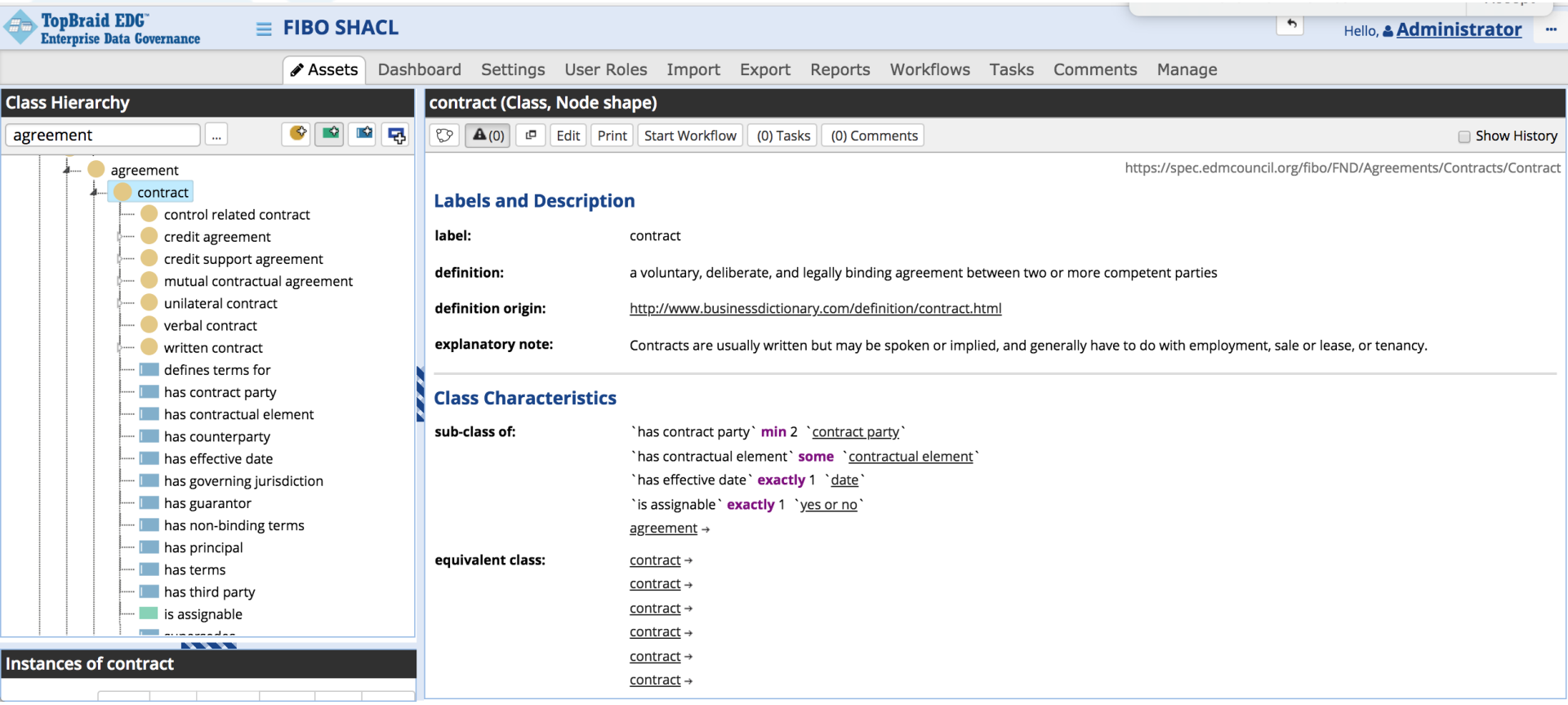


FIBO is maintained in OWL published to various formats

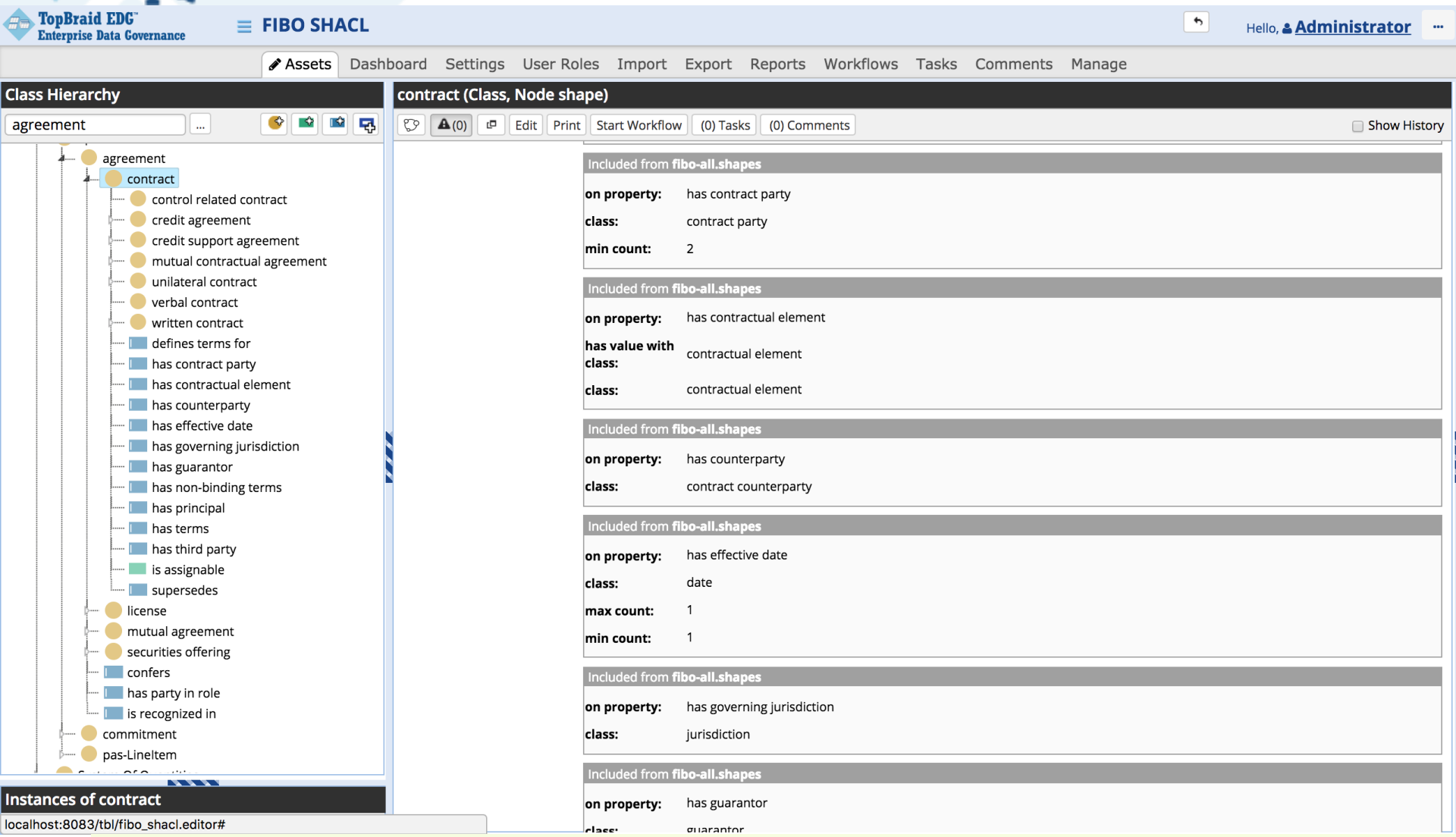
- FIBO SKOS
- Spreadsheet data dictionary
- …

- FIBO can be published in SHACL
- Caveat: SHACL is more expressive than OWL

Automatically converted FIBO SHACL

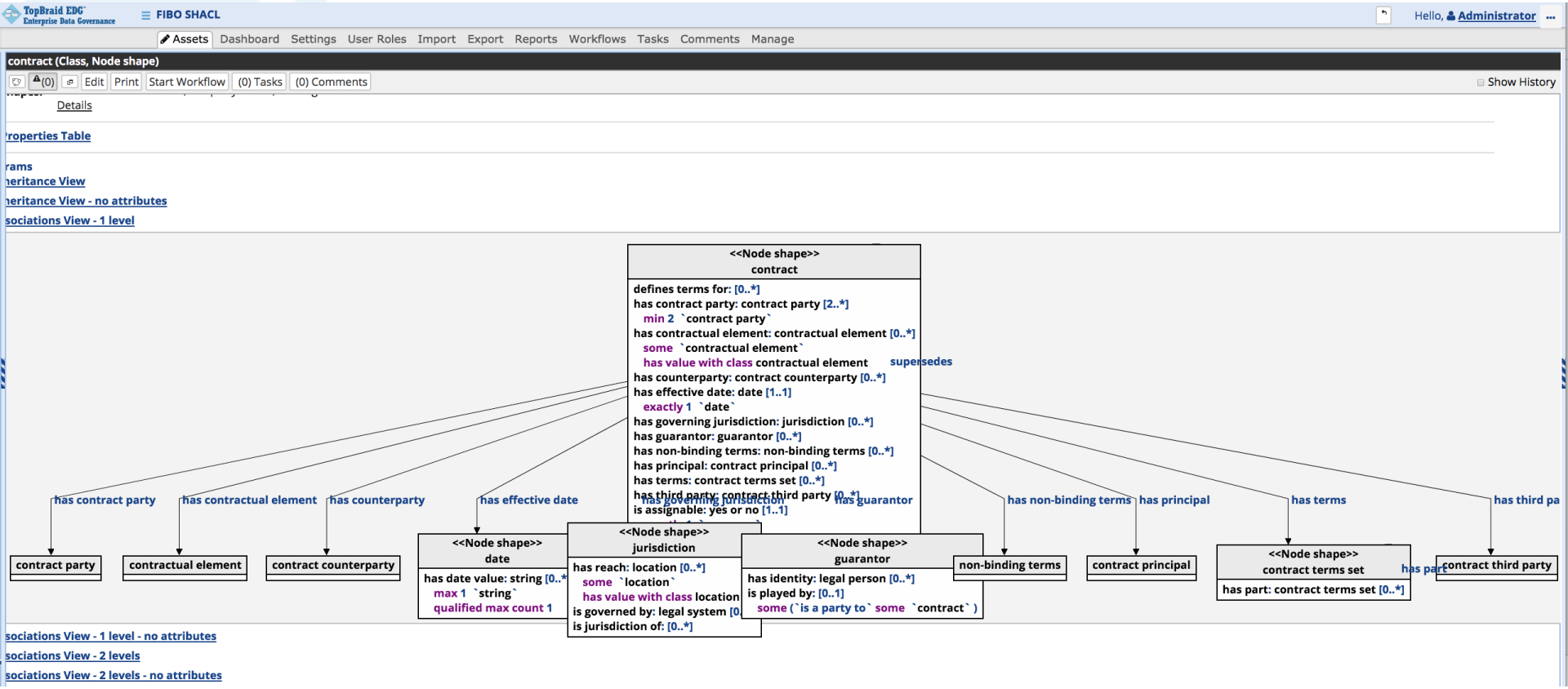


Automatically converted FIBO SHACL

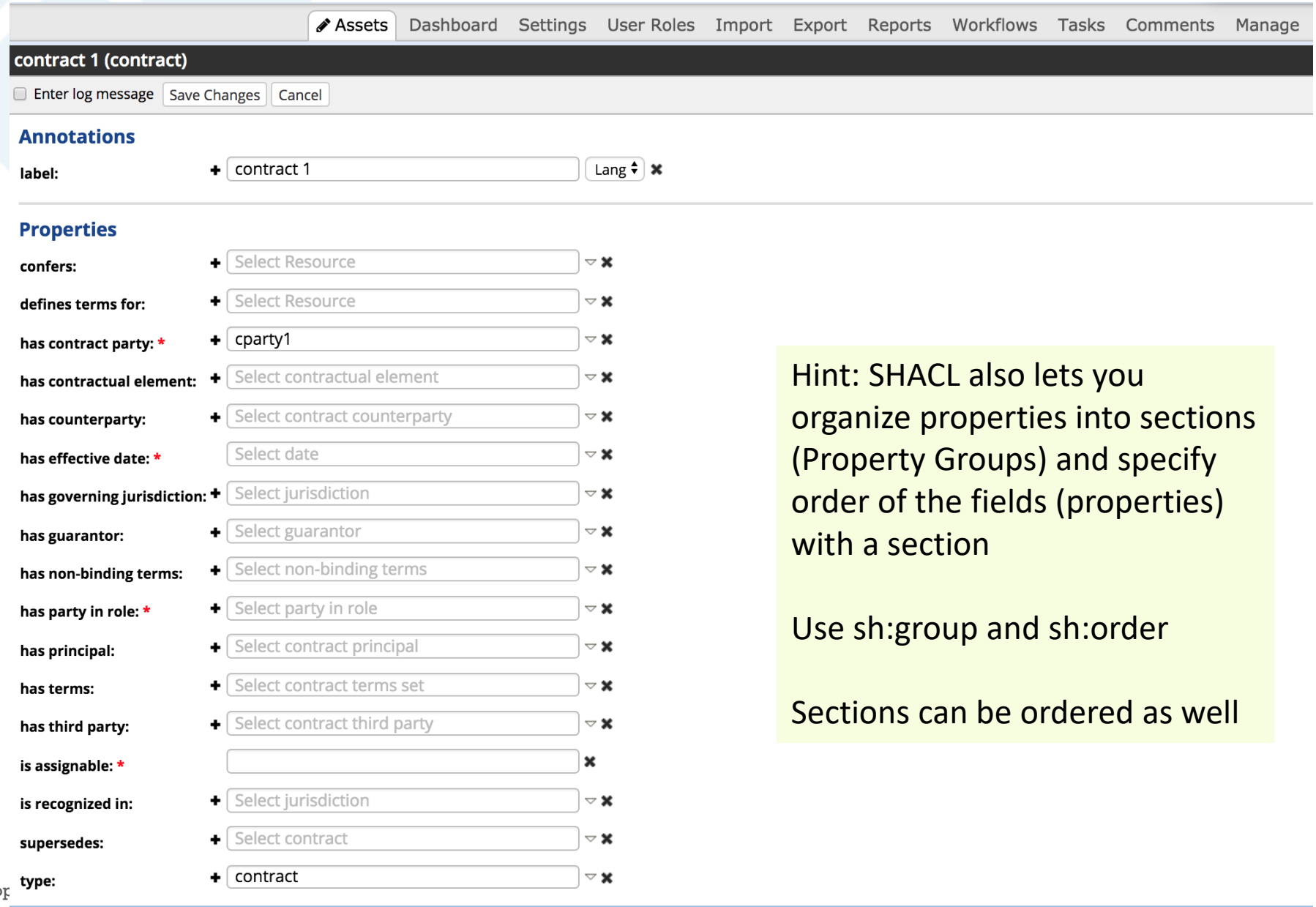


Fixes up unnecessary qualified cardinality restrictions: e.g. range of the property ‘has contract party’ is ‘contract party’ – qualified cardinality is not necessary and misleading

Automatically converted FIBO SHACL



Automated Form Generation for a Contract

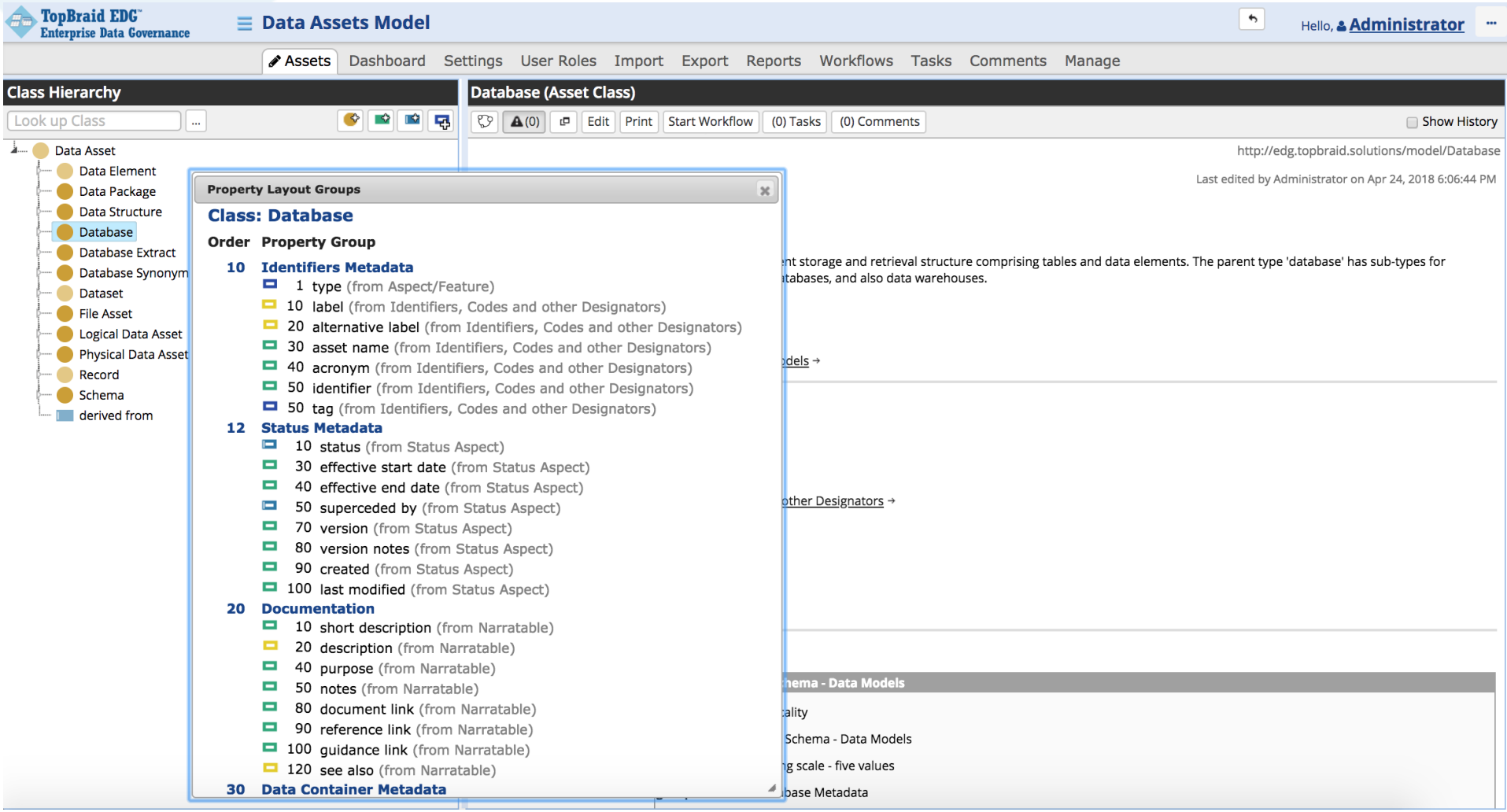


Hint: SHACL also lets you organize properties into sections (Property Groups) and specify order of the fields (properties) with a section

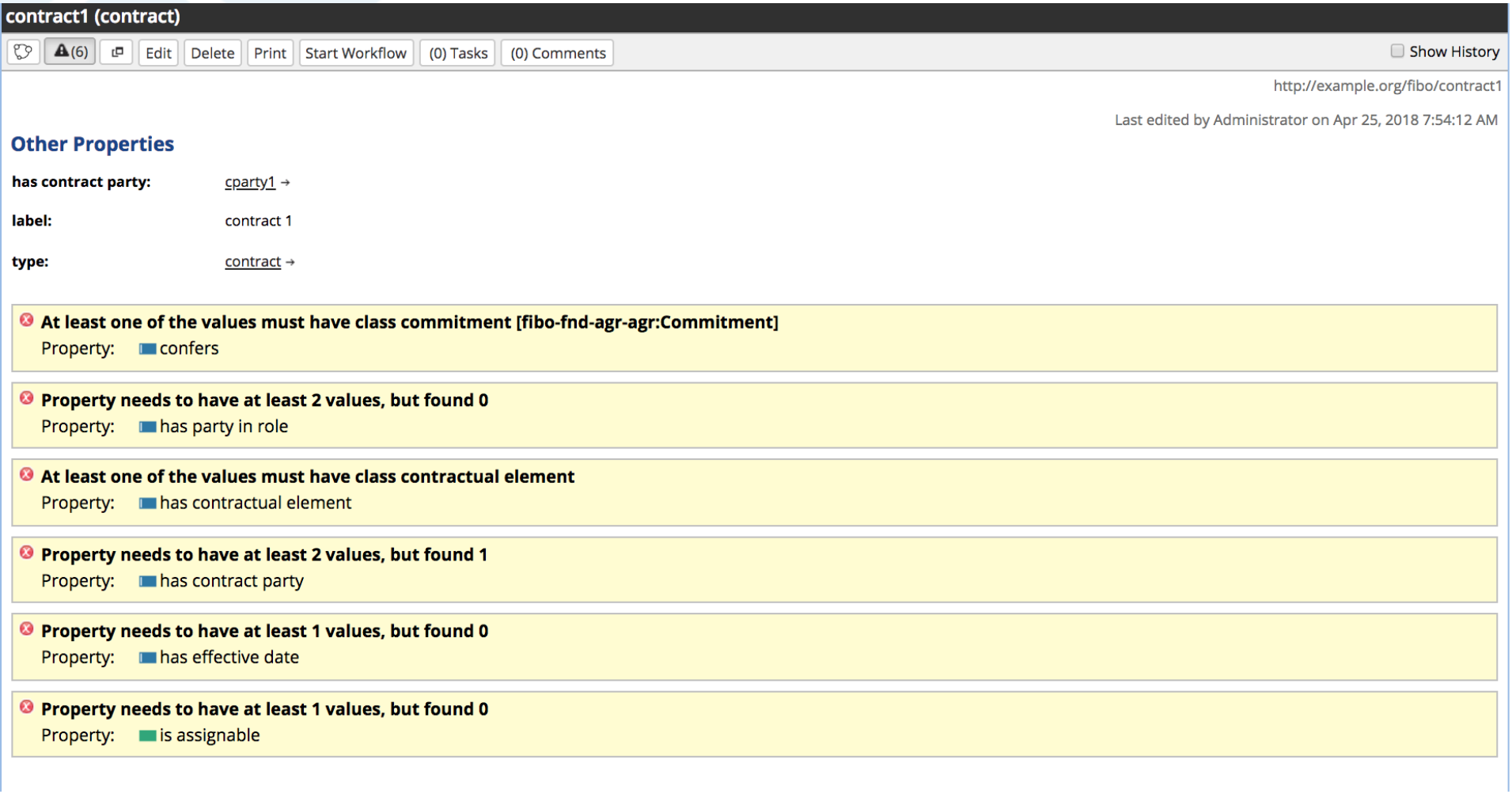
Use sh:group and sh:order

Sections can be ordered as well

Example of ordered Sections using one of pre-built classes (Database) from EDG ontologies



Data Validation for a Contract



Other features and benefits e.g., automated APIs/GraphQL queries with SHACL

TopQuadrant™

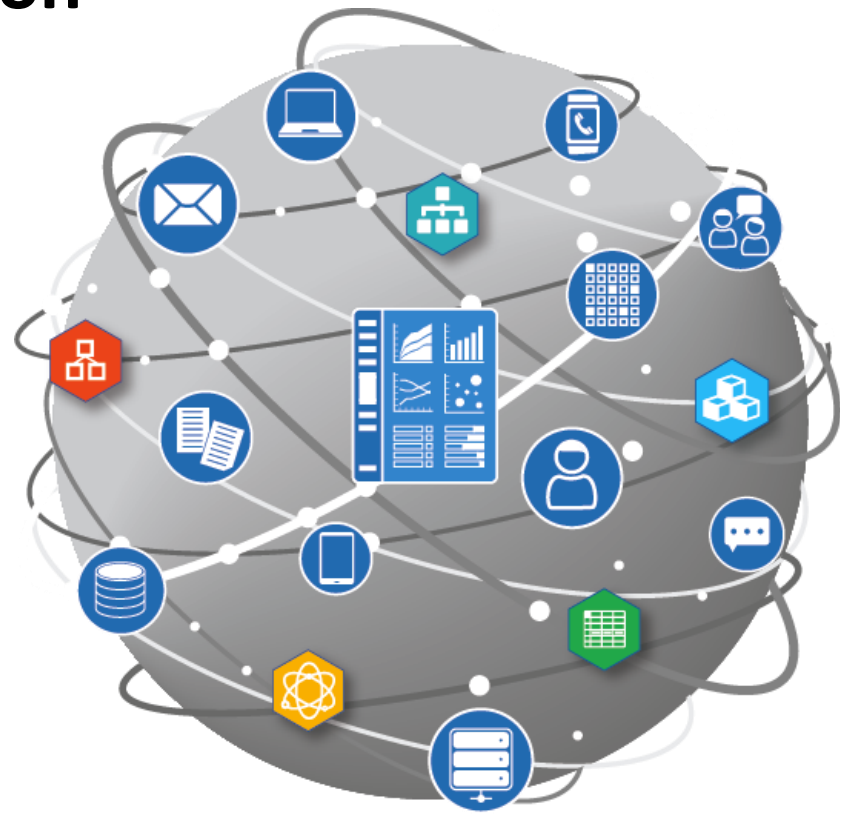
SHACL and Enterprise Data Management

SHACL for Data Governance

Asset Definition

SHACL is used in TopBraid EDG to identify the information about an asset that is subject to governance (e.g. provenance, security, essential metadata).

EDG comes with over 100 asset types including taxonomies and ontologies –***pre-built using SHACL***.



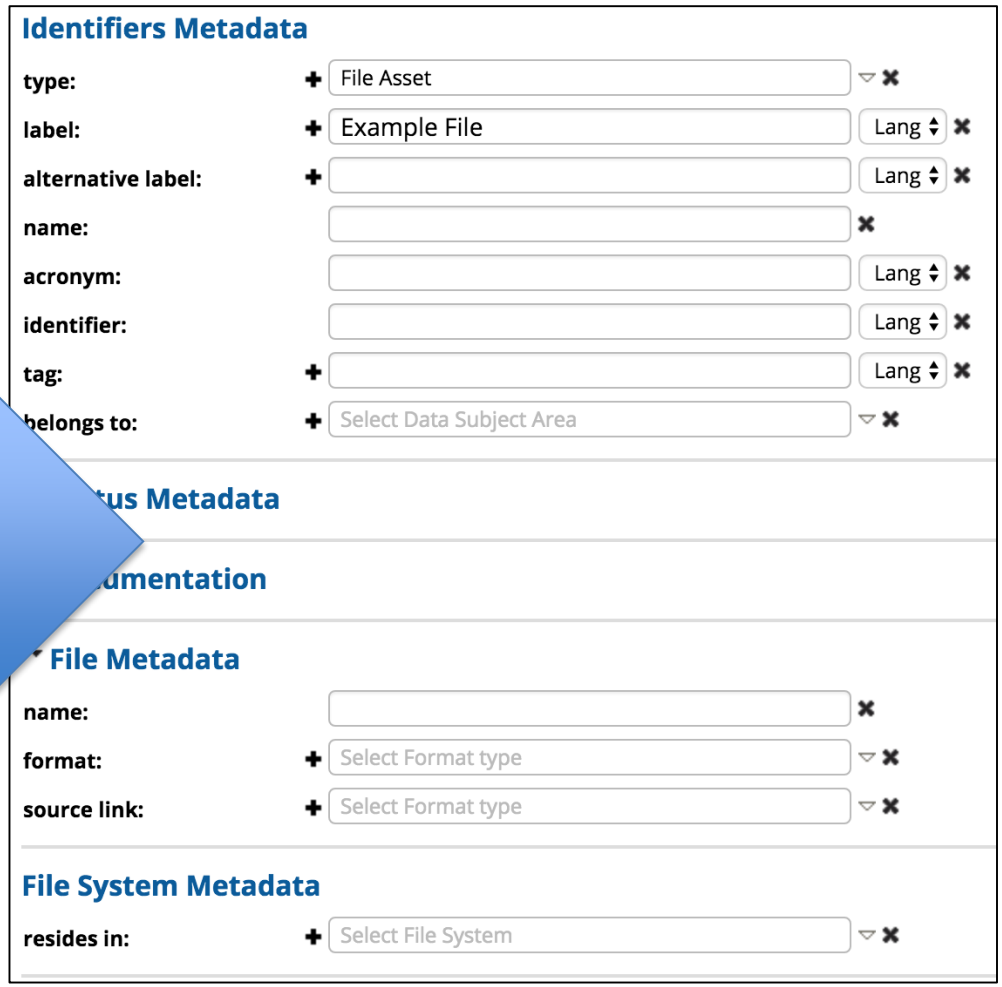
TopBraid EDG™
Enterprise Data Governance

Asset Definition

Models drive the User Experience (UX) i.e., what fields will appear on an (edit, search or browse) form, order of the fields, their division into sections and other relevant information.

```
edg:FileAsset
  rdf:type edg:AssetClass ;
  dash:abstract "false"^^xsd:boolean ;
  rdfs:label "File Asset" ;
  rdfs:subClassOf edg:AccessControllable ;
  rdfs:subClassOf edg:ComplianceAspect ;
  rdfs:subClassOf edg:Identifiable ;
       ⋮                ⋮
  sh:property edg:File-formatType ;
  sh:property edg:File-name ;
  sh:property edg:File-residesIn ;
  sh:property edg:File-sourceLink ;
.
edg:File-formatType
  rdf:type sh:PropertyShape ;
  sh:path edg:format ;
  sh:class edg:FormatType ;
.
edg:File-name
  rdf:type sh:PropertyShape ;
  sh:path edg:name ;
  sh:datatype xsd:string ;
.
edg:File-residesIn
  rdf:type sh:PropertyShape ;
  sh:path edg:residesIn ;
  sh:class edg:FileSystem ;
.
```

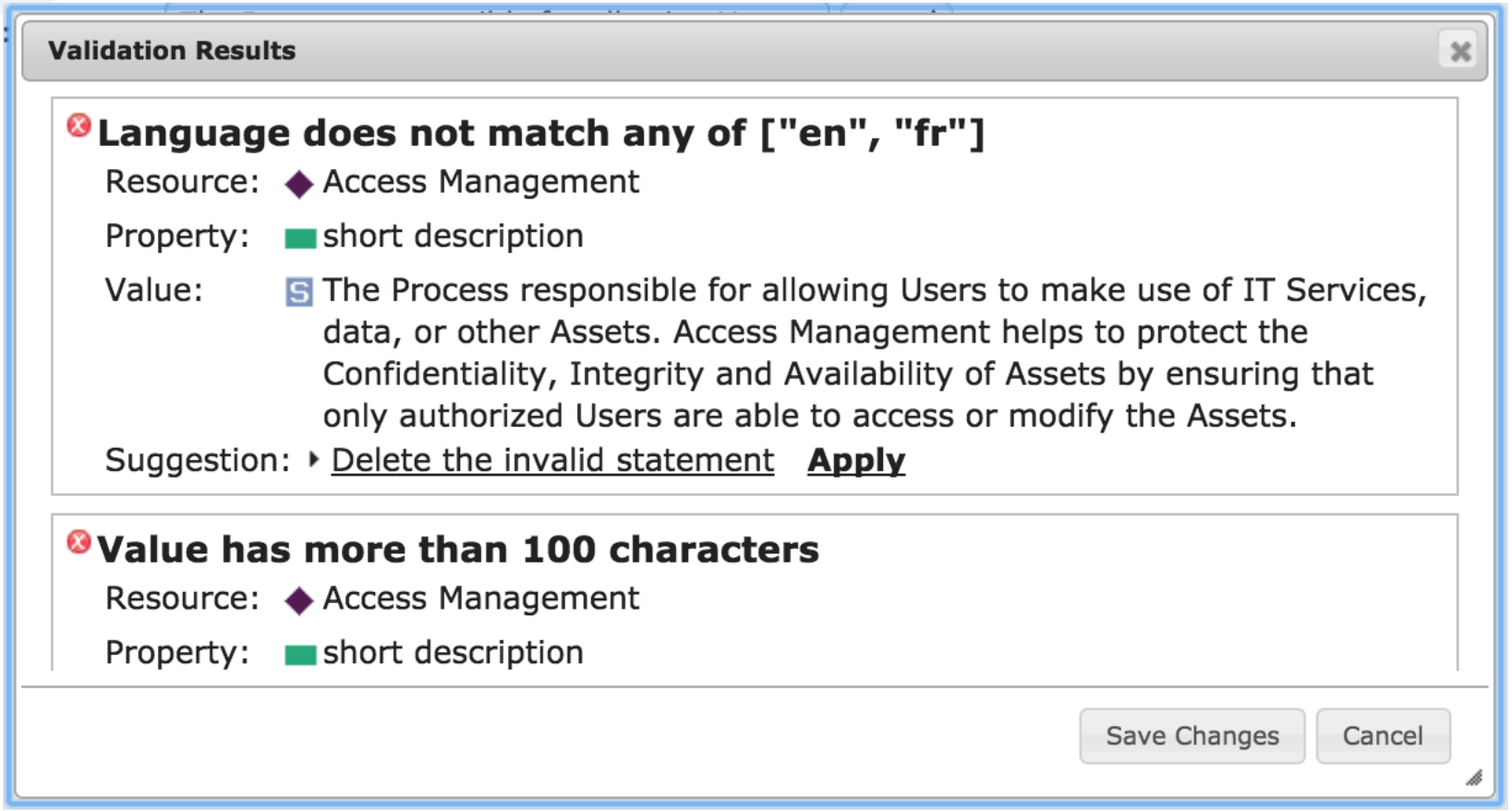
SHACL



Ability to Define Data Quality Rules

SHACL can be used to enforce organization's best practices. E.g.:

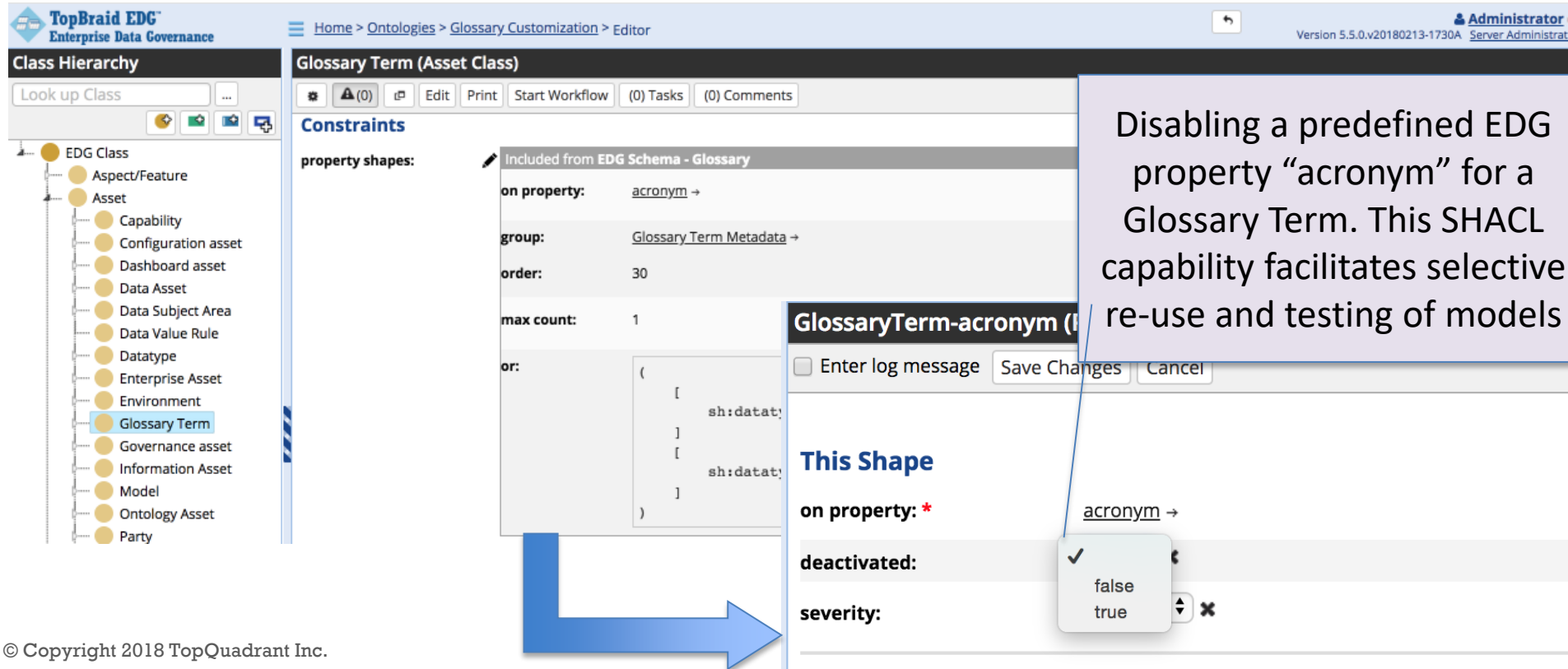
A business term must have a short description in both (and only) English and French with length under 100 characters.



TopQuadrant™ Ease of Configuration & Forward Compatibility

Organizations often need to configure the pre-defined models.

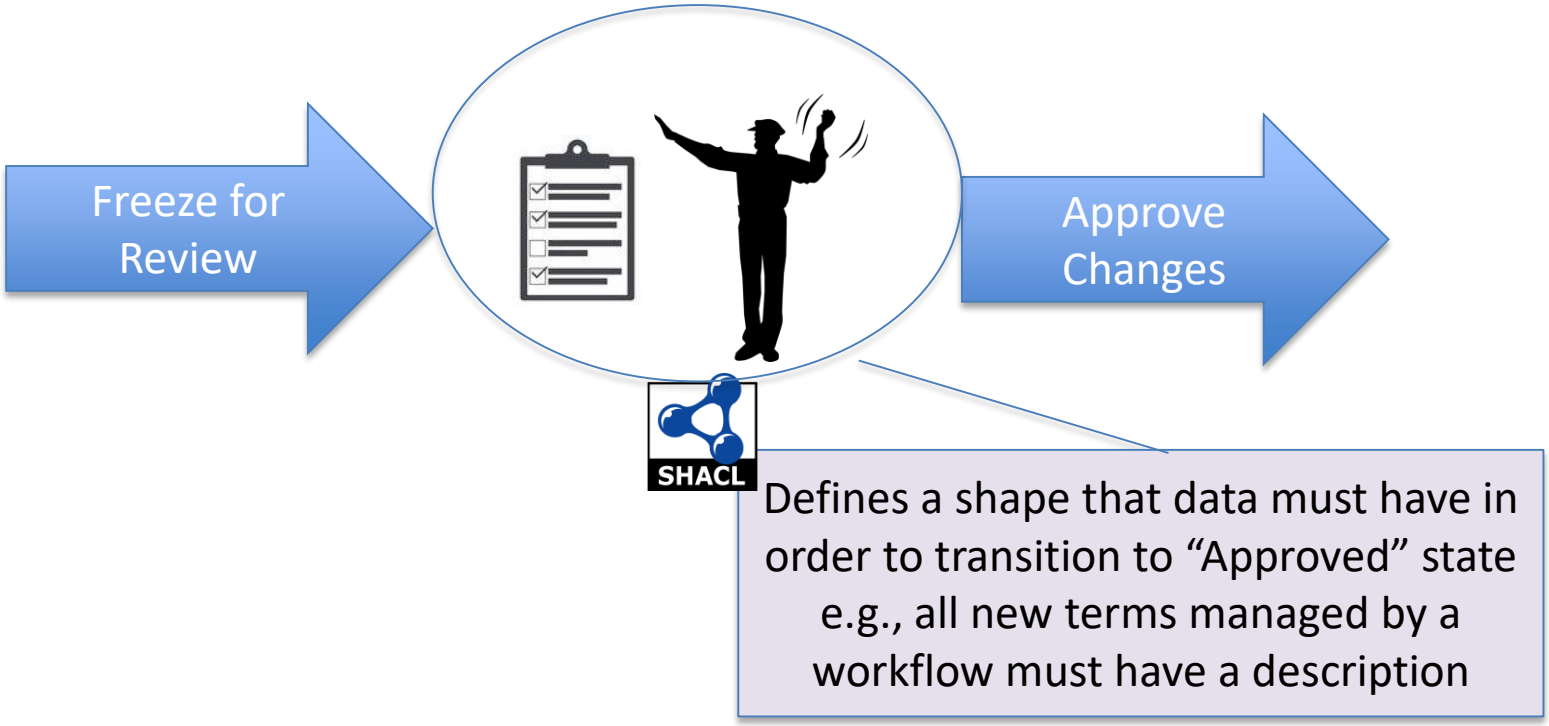
SHACL not only offers flexibility to define new attributes and relationships for an asset type, but also lets users 'disable' predefined attributes and relationships if they decide they don't need to capture certain information.



Ensure Governance Workflows

TopBraid EDG supports configurable workflows.

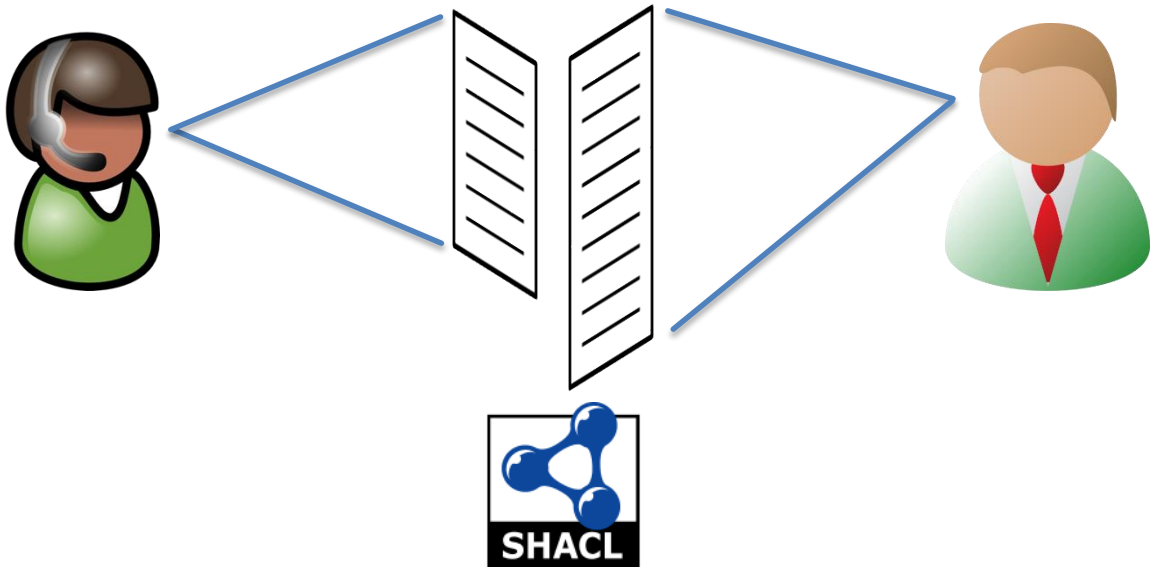
Each transition of a workflow from one state to another can include SHACL statements specifying conditions that data must satisfy in order to pass on to the next state in a workflow.



Role Based Personalization

SHACL can be used to provide different views or perspectives for different users

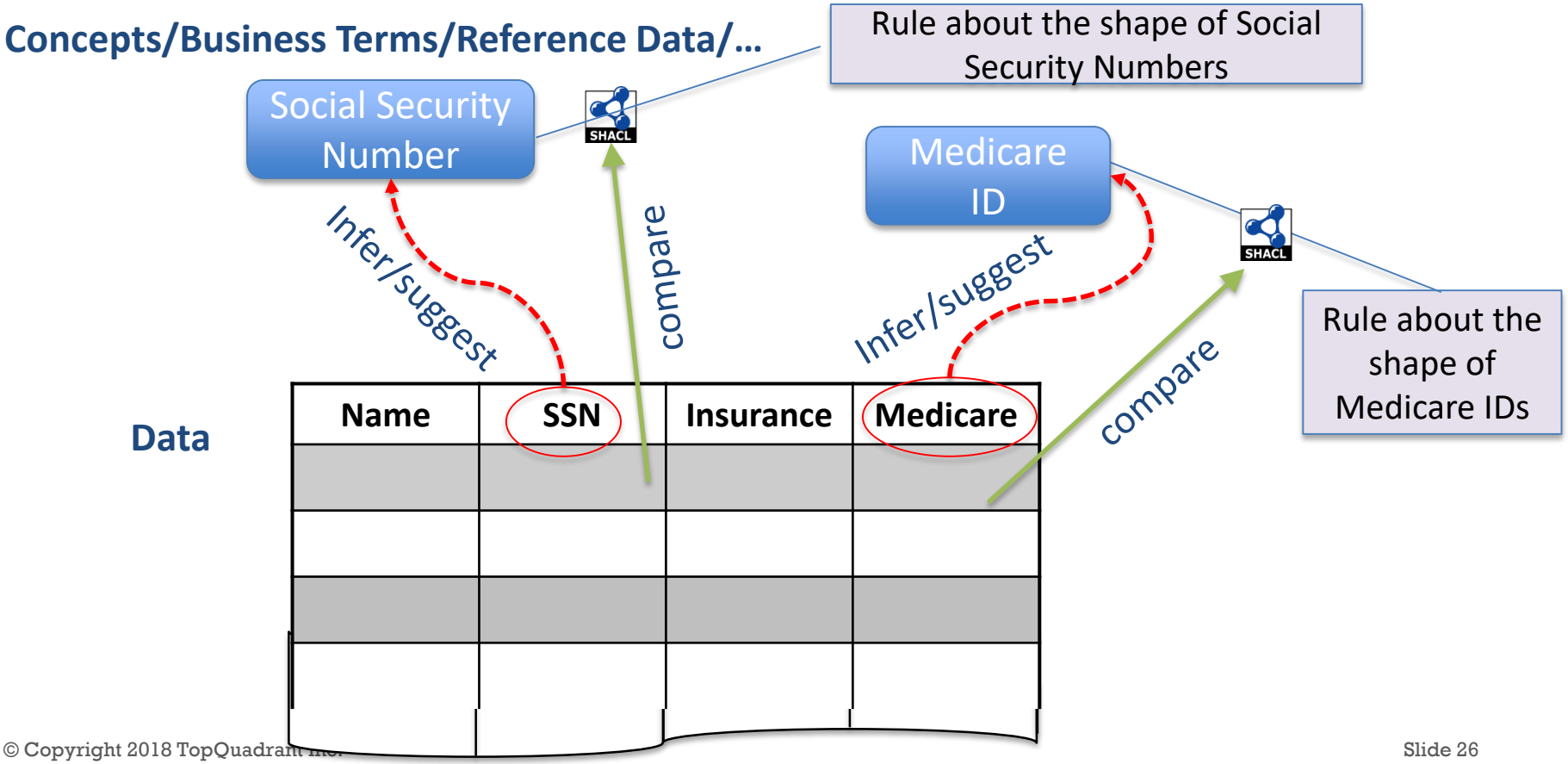
SHACL let's us define multiple shapes for a given asset type. These shapes can be associated with different user roles. As a result, different users will get different views of the information. To support unique needs of each role, some fields can be specified as view-only. Other fields can be calculated based on available values.



Data Discovery and Mapping Automation

SHACL rules can be used to create new connections across collected information.

Rules defined in SHACL are directly executable and can be used to infer new information – for example, from existing technical metadata and results of data profiling. This SHACL capability is used extensively by EDG to cross-link, create new facts, add smartness and automation.



Auto-discovery with Confidence Calculation

TopQuadrant™

TopBraid EDG
Enterprise Data Governance

Home > Data Assets Collections > Northwind > Problems and Suggestions

Problems and Suggestions (3)

Download as TSV Download as JSON

Minimum suggestion confidence: 0%

Suggested mapping to term
Resource: CATEGORYNAME (NORTHWIND.DBO.ALPHABETICAL LIST OF PRODUCTS)
Property: maps to term
Suggestion: Map column to Product Category, based on 5 out of 5 samples (Confidence: 100%) **Apply**

Suggested mapping to term
Resource: CATEGORYNAME (NORTHWIND.DBO.PRODUCTS BY CATEGORY)
Property: maps to term
Suggestion: Map column to Product Category, based on 5 out of 5 samples (Confidence: 100%) **Apply**

Suggested mapping to term
Resource: CATEGORYNAME (NORTHWIND.DBO.SALES BY CATEGORY)
Property: maps to term
Suggestion: Map column to Product Category, based on 5 out of 5 samples (Confidence: 100%) **Apply**

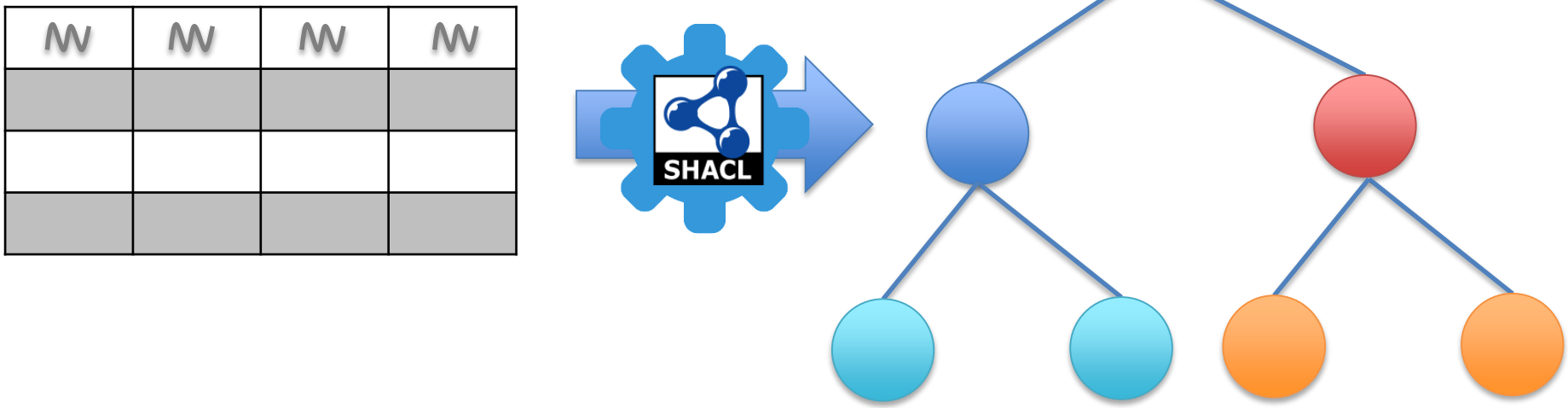
Apply all top suggestions

New facts auto-created by SHACL can have confidence attached to them

Data “on boarding” Ease and Configurability

SHACL can be used to define data transformations.

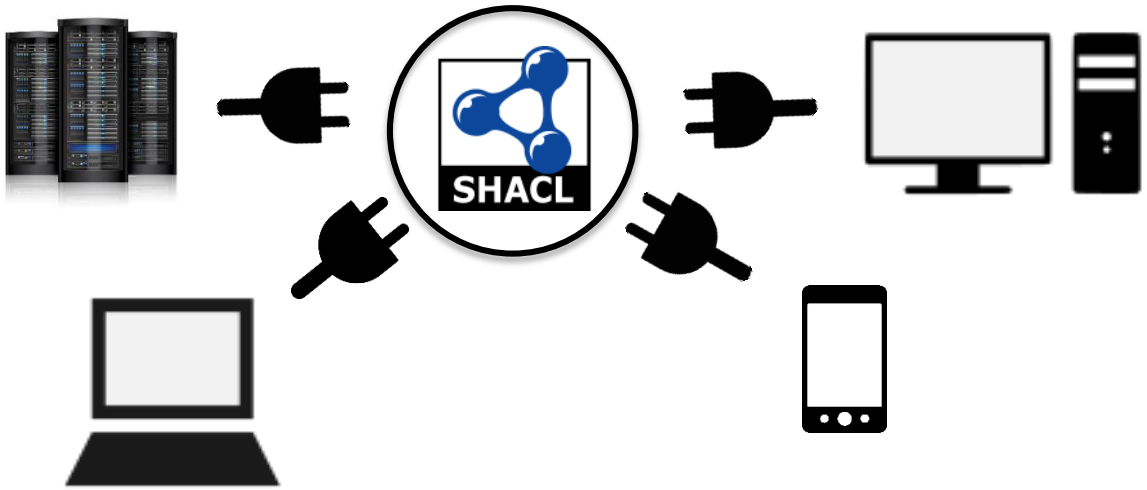
When importing information from different sources (e.g., spreadsheets, XML, RDBMS), information needs to be mapped to the asset definitions in TopBraid EDG. SHACL let's us do this.



Data Interface Configurability

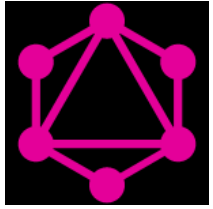
SHACL can define data interfaces.

Just like SHACL can be used to specify different views of the information for different roles, it can also define different views or data structures to be delivered to systems that must interact with TopBraid EDG to access the information it stores. The definitions can include computations and transformations needed to satisfy the requirements of consuming systems.



TopQuadrant™

GraphQL



graphql.org/

- World's most popular graph "query language"
- Not really a query language, but rather a way to define data interfaces – JSON structures
- Very similar to SHACL – can serve as a compact syntax for SHACL

```
# A user account
type User {
   name: String!
   age: Int
   gender: Gender
   purchases: [Purchase]
}

type Purchase {
   # The internal ID of the product
   productId: String!
   date: String}

enum Gender {
   FEMALE
   MALE
}
```

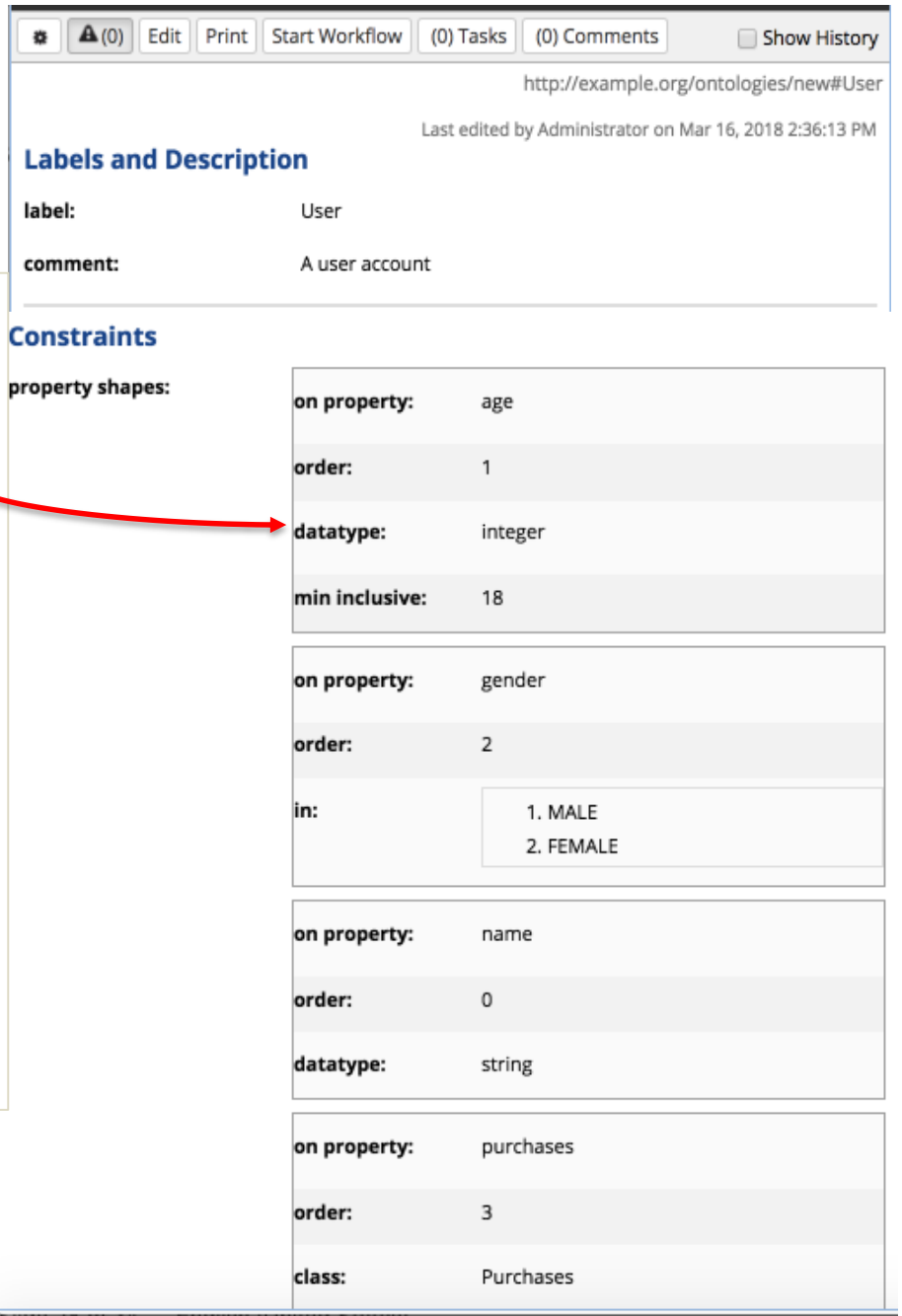
GraphQL as a simple RDF Modelling Language

Converted to SHACL on the fly.
GraphQL “directives” to express constraints, etc.

```
# A user account
type User {
   name: String!
   age: Int  @shape(minInclusive: 18)
   gender: Gender
   purchases: [Purchase]
}

type Purchase {
   # The internal ID of the product
   productId: String! @shape(minLength: 8, pattern: "[0-9]+")
   date: String @shape(datatype: "xsd:date")
}

enum Gender {
   FEMALE
   MALE
}
```



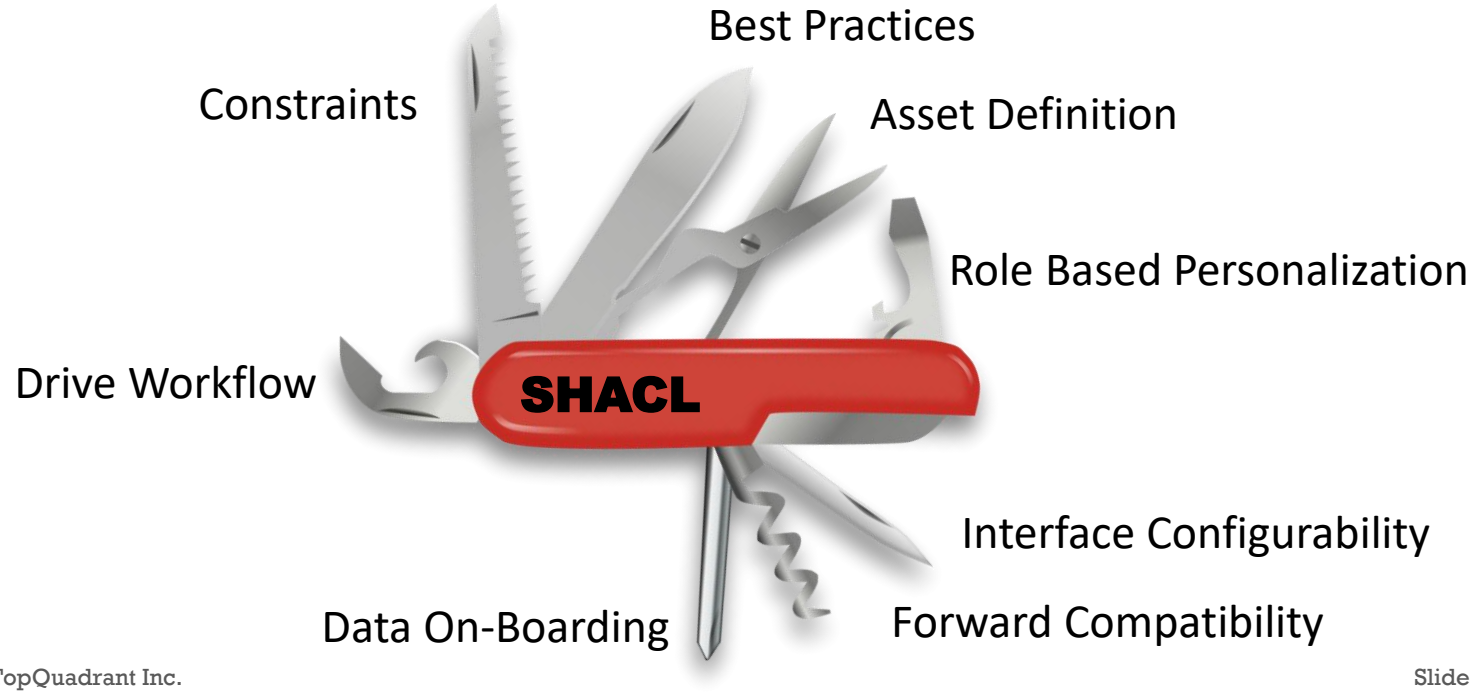
TopQuadrant™ Opens Up Many Possibilities

- Every GraphQL service treated as RDF
  - GraphQL schemas converted to SHACL shapes
  - JSON instances converted to RDF nodes
- Every RDF source treated as GraphQL
  - SHACL shapes converted to GraphQL schemas
  - RDF graphs exposed as GraphQL service
- GraphQL libraries and tools add to the RDF infrastructure
- RDF+SHACL brings IDs (URIs) to GraphQL objects, richer query, inferencing

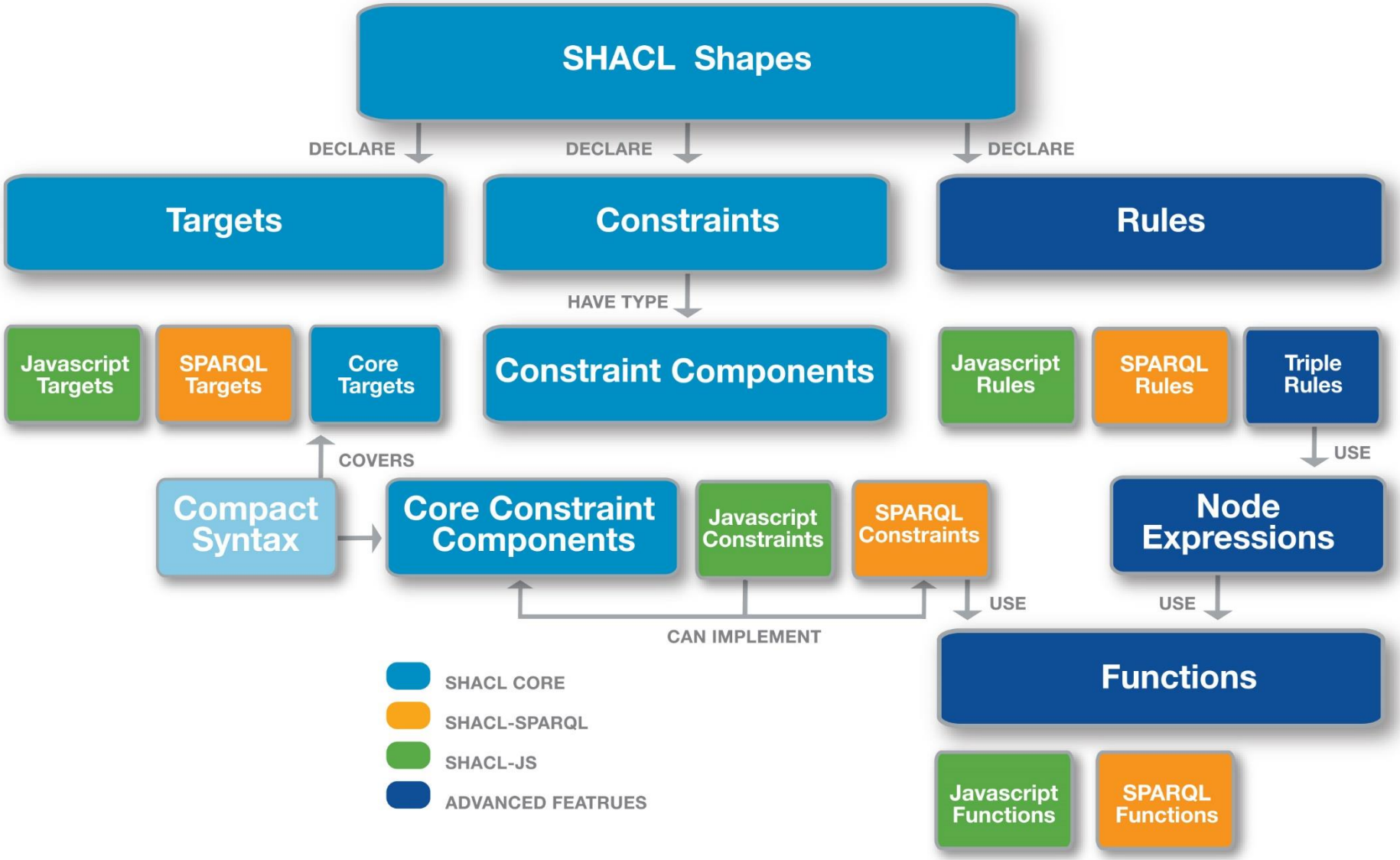
Learn SHACL Once, Use for Many Needs

A single language for delivering a range of capabilities

All of the value propositions and features listed so far are important for a data governance tool. Each could be supported using its own technical approach requiring a separate investment in understanding and learning it. Supporting all of them with SHACL brings consistency. Users do not need to learn different approaches to take advantage of these capabilities.



Detailed View of SHACL



In Conclusion

- I have given a brief overview of SHACL
- Demoed FIBO in SHACL
- Discussed SHACL's features and use case – for data governance and beyond

- Come to the booth for more !

… Questions?

Resources for learning SHACL

- SHACL Community Group https://www.w3.org/community/shacl/
- SHACL W3C Wiki https://www.w3.org/2014/data-shapes/wiki/Main_Page
  - Links to implementations, WG deliverables, Meeting minutes, …
  - Some historic info – not so useful anymore
- TQ's SHACL page - http://www.topquadrant.com/technology/shacl/
  - Tutorials, articles, presentations – for example:
  - AN OVERVIEW OF SHACL FEATURES AND SPECIFICATIONS
  - USING SHACL DATA CONSTRAINTS IN THE TOPBRAID WEB PRODUCTS EVN AND EDG
  - HOW TO DEFINE CONSTRAINTS ON RDF:LISTS USING SHACL
  - HOW TO USE TOPBRAID AS A DATA VALIDATION SERVER
- SHACL OVERVIEW WEBINARS:
  - https://www.topquadrant.com/knowledge-assets/topquadrant-webinars/#TQ-SHACL-overview-webinar

TopQuadrant™

Thank You !

To contact me, use: irene@topquadrant.com

To Learn More about TopBraid EDG, see:

EDG Product Info:

- http://www.topquadrant.com/products/topbraid-edg/
- http://www.topquadrant.com/products/topbraid-edg-gov-packs/

EDG demos/webinar recordings:

- https://www.topquadrant.com/knowledge-assets/topquadrant-webinars/#TQ-EDG-appliedDG-part2-webinar
  Webinar: Applied Data Governance – Part 2: A Day in the life of a Technical Data Steward
- https://www.topquadrant.com/knowledge-assets/topquadrant-webinars/#TQ-EDG-realworldDG-part1-webinar
  Webinar: Applied Data Governance – Part 1: A Day in the life of a Business Steward
- http://www.topquadrant.com/knowledgeassets/videos/#edgoverviewdemo
  Webinar: Data Governance for the Connected Enterprise: TopBraid EDG in Action
- http://www.topquadrant.com/knowledge-assets/topquadrant-webinars/#TQ-EDG-metadata-mgt-webinar
  Webinar: Metadata Management is Key to Data Governance Initiatives