

Going Live with Enterprise Solutions



SHACL (Shapes Constraint Language) – An Introduction

Irene Polikoff, TopQuadrant
EDW, San Diego, April 2018



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

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



Irene Polikoff

- **SH**Apes **C**onstraint **L**anguage
 - W3C Recommendation to define the “shape” of data
 - Shapes are defined by constraints and validation rules
 - Shapes provide facilities 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:

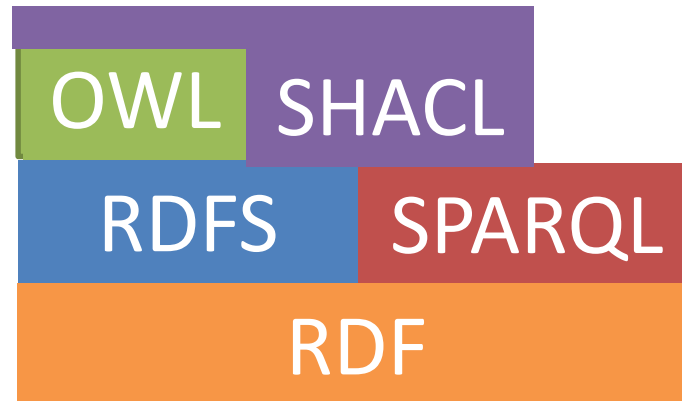
*Is that term used correctly?
You can't say that here!
What can you infer?*

Vocabulary:

Shared terms can we use

Statements:

Saying things



Query:

What did you say?

Why talk about SHACL at EDW?

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

SHACL ensures data quality and enhances the data

Motivation for SHACL at W3C

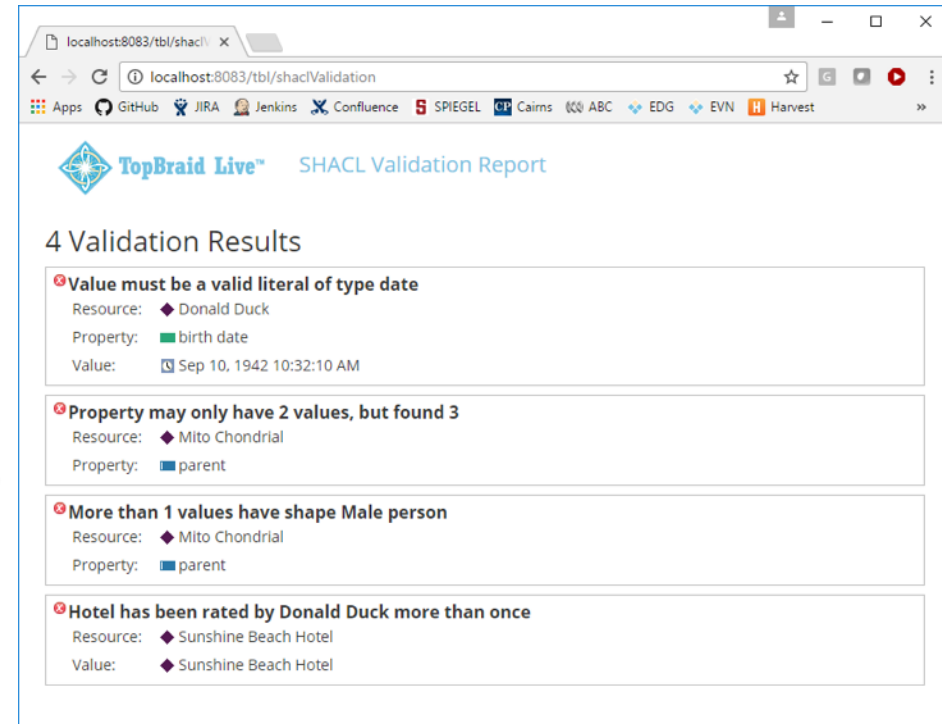
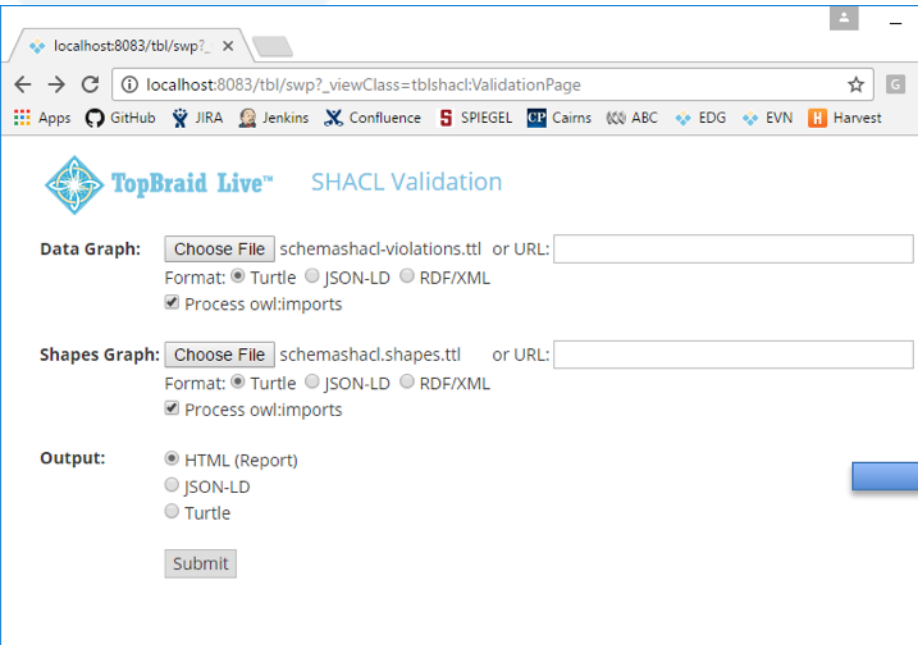
- SHACL is rooted in the world of the Semantic Web W3C standards
 - Until SHACL no official W3C recommendation for validating RDF
- RDF Schema / OWL are “schema languages”, however...
 - Descriptive, not prescriptive or restrictive
 - OWA, CWA, UNA ⇒ classification inferencing not validation
 - Even for inferencing/classification, OWA is often a problem - See using SHACL for classification rules:
<https://henrietteharmse.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

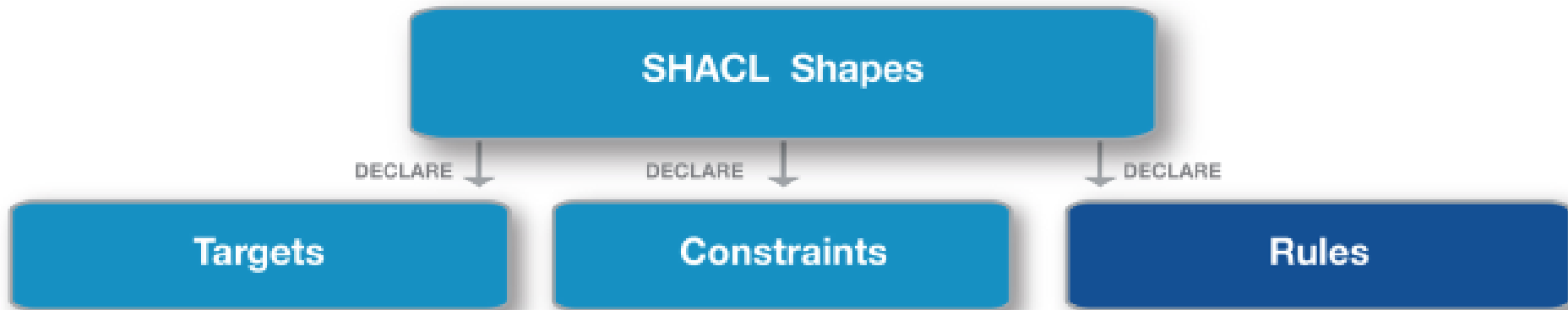
SHACL and Rules

- 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



Example Shapes

Explicit declaration is recommended, but SHACL processors will recognize shapes even without type statements

```
ex:PersonShape a sh:NodeShape ;
  sh:targetClass ex:Person ;
  sh:pattern "^http://mydomain.com/person/" ;
  sh:property [ # _:b1
    sh:path ex:firstName ;
    sh:minCount 1 ;
    sh:datatype xsd:string ; ] ;
  sh:property ex:Person-lastName .

ex:Person-lastName sh:path ex:lastName ;
  sh:minCount 1 ;
  sh:maxCount 1 ;
  sh:datatype xsd:string ;
  sh:maxLength 20 .
```

Node Shape

Property Shape
as a blank
node

Property Shape
as a URI

- Node shapes are used to:
 - Specify constraints on the target nodes
 - Group property shapes

Example Data

@prefix example: <http://example.org/> .
@prefix ex: <http://mydomain.com/person/> .

ex:Alice a ex:Person .

ex:Bob a ex:Student .

ex:Jack a ex:Person .

ex:Jill a ex:Teacher .

ex:Student rdfs:subClassOf ex:Person .

ex:Alice ex:firstName "Alice";

ex:lastName "Jones";

ex:knows example:Bob .

example:Bob ex:firstName "Bob";

ex:lastName "Brown".

ex:Jack ex:firstName "Jack";

ex:lastName "Smith" ;

ex:lastName "Jones".

ex:Jill ex:firstName 1 .

valid

not valid,
wrong URI

not valid, two
last names

not validated

Will be validated against
the PersonShape – based
on its target statement

Will NOT be validated
against the PersonShape –
not a target

Variation on the Shapes Example - 1

```
ex:PersonShape a sh:NodeShape ;
  sh:targetSubjectsOf ex:SSN ;
  sh:pattern "^http://mydomain.com/person/" ;
  sh:property [ # _:b1
    sh:path ex:firstName ;
    sh:minCount 1 ;
    sh:datatype xsd:string ; ] ;
  sh:property [ # _:b2
    sh:path ex:lastName ;
    sh:minCount 1 ;
    sh:maxCount 1 ;
    sh:datatype xsd:string ;
    sh:maxLength 20 ; ] .
```

Targets can be specified in terms of class membership, in terms of subjects or objects in statements with certain predicates and, more generally, using SPARQL

Variation on the Shapes Example - 2

```
ex:Person a sh:NodeShape ;
a rdfs:Class ;
sh:pattern "^http://mydomain.com/person/" ;
sh:closed true; sh:ignoredProperties ( rdf:type ) ;
sh:property [ # _:b1
  sh:path ex:firstName ;
  sh:minCount 1 ;
  sh:datatype xsd:string ; ] ;
sh:property [ # _:b2
  sh:path ex:lastName ;
  sh:minCount 1 ;
  sh:maxCount 1 ;
  sh:datatype xsd:string ;
  sh:maxLength 20 ; ] .
```

Shape can also be a class. In this case, target is implicit – all members of the class

Shape can be declared as “closed”. This means that the only valid properties are those defined in property shapes or put into an “ignore” list

Variation on the Shapes Example - 3

```
ex:PersonShape a sh:NodeShape ;
  sh:targetSubjectsOf ex:SSN ;
  sh:pattern “^http://mydomain.com/person/” ;
  sh:property [ # _:b1
    sh:path ex:firstName ;
    sh:minCount 1 ;
    sh:datatype xsd:string ; ] ;
  sh:property [ # _:b2
    sh:path ex:lastName ;
    sh:minCount 1 ;
    sh:maxCount 1 ;
    sh:datatype xsd:string ;
    sh:maxLength 20 ; ] ;
  sh:property
    [ sh:path rdf:type ;
      sh:or (
        [ sh:hasValue ex:Customer ]
        [ sh:hasValue ex:Person ]
      ) ] .
```

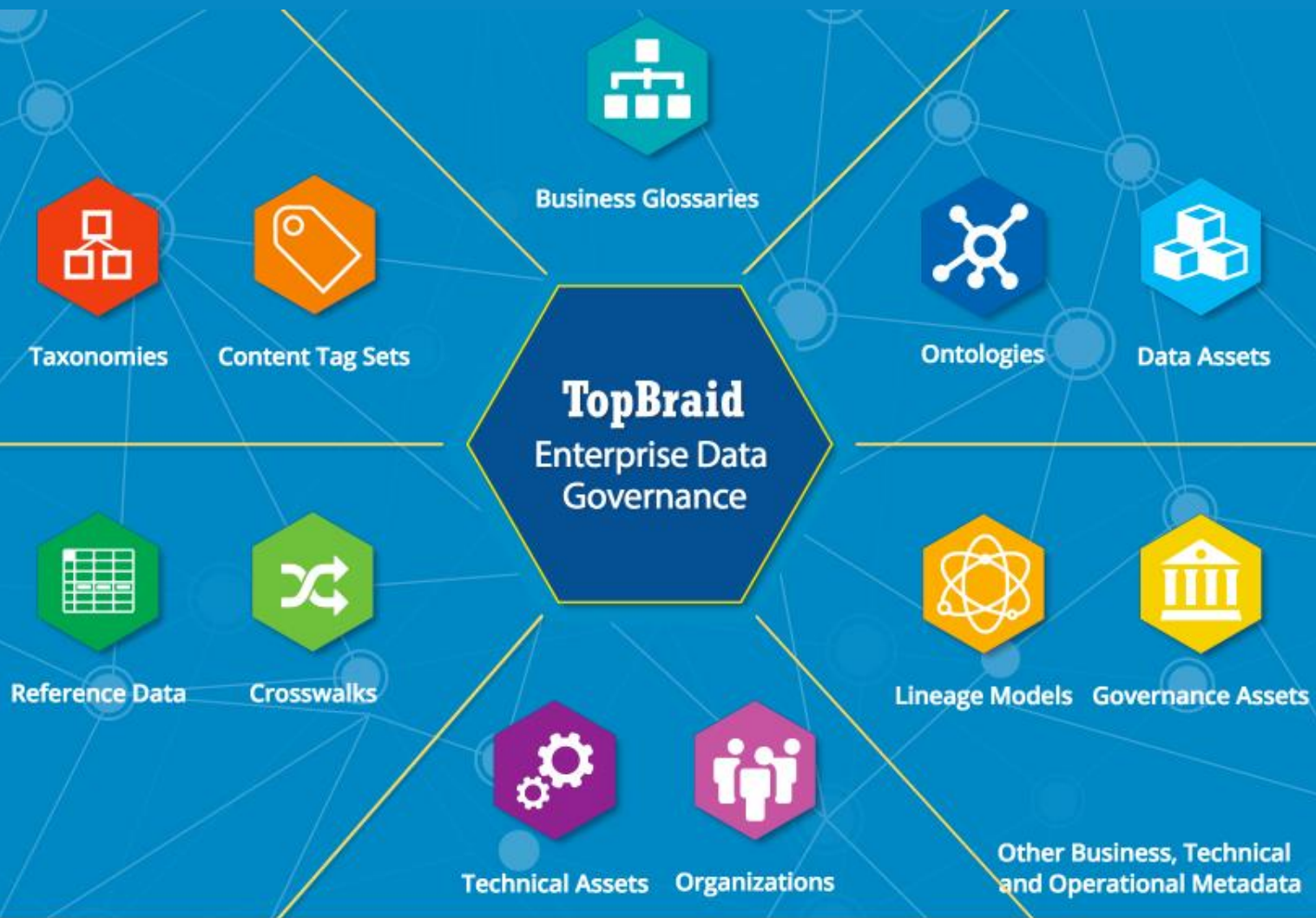
Shapes can be compounded using logical expressions with OR, AND and NOT

SHACL Terminology (recap)

- Targets (of a shape)
 - determine what nodes/resources are to be validated against a shape
 - during the validation, targets are referred to as “focus nodes”
- Node Shapes
 - specify conditions a target node itself must comply with
 - used to group property shapes
- Property Shapes
 - specify conditions that related nodes (property values) must comply with
- Constraint Components
 - predefined CCs in SHACL Core form “SHACL Core vocabulary”
 - e.g., sh:minCount, sh:datatype, sh:pattern, etc.
 - users can create new CCs – domain specific data validation languages
- Shapes Graph, Data Graph
 - These are “roles” – any graph can be declared to be a shapes graph or a data graph
- Validation Report
 - RDF graph with validation results
 - SHACL includes a vocabulary for describing results

SHACL and Enterprise Data Management

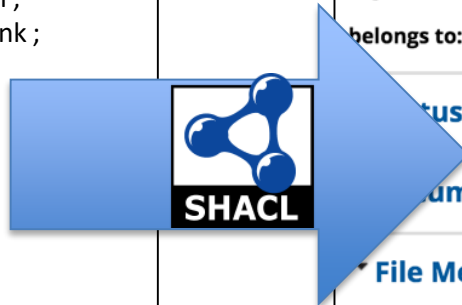
SHACL for Data Governance



TopQuadrant™ 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 ;
.
```



Identifiers Metadata

type: ▾ ✕

label: Lang ▾ ✕

alternative label: Lang ▾ ✕

name: ✕

acronym: Lang ▾ ✕

identifier: Lang ▾ ✕

tag: Lang ▾ ✕

belongs to: ▾ ✕

File Metadata

name: ✕

format: ▾ ✕

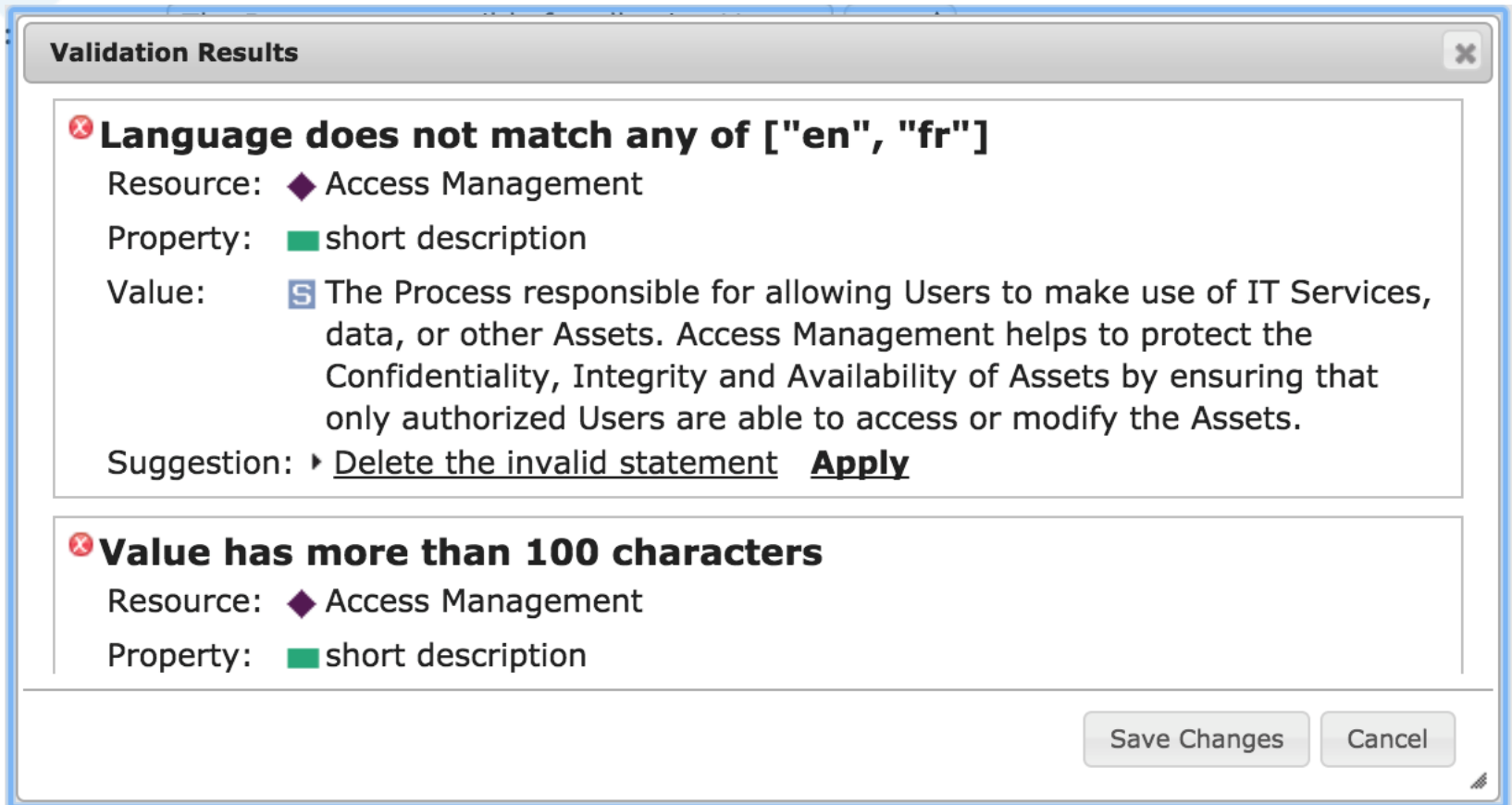
source link: ▾ ✕

File System Metadata




resides in: ▾ ✕



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.



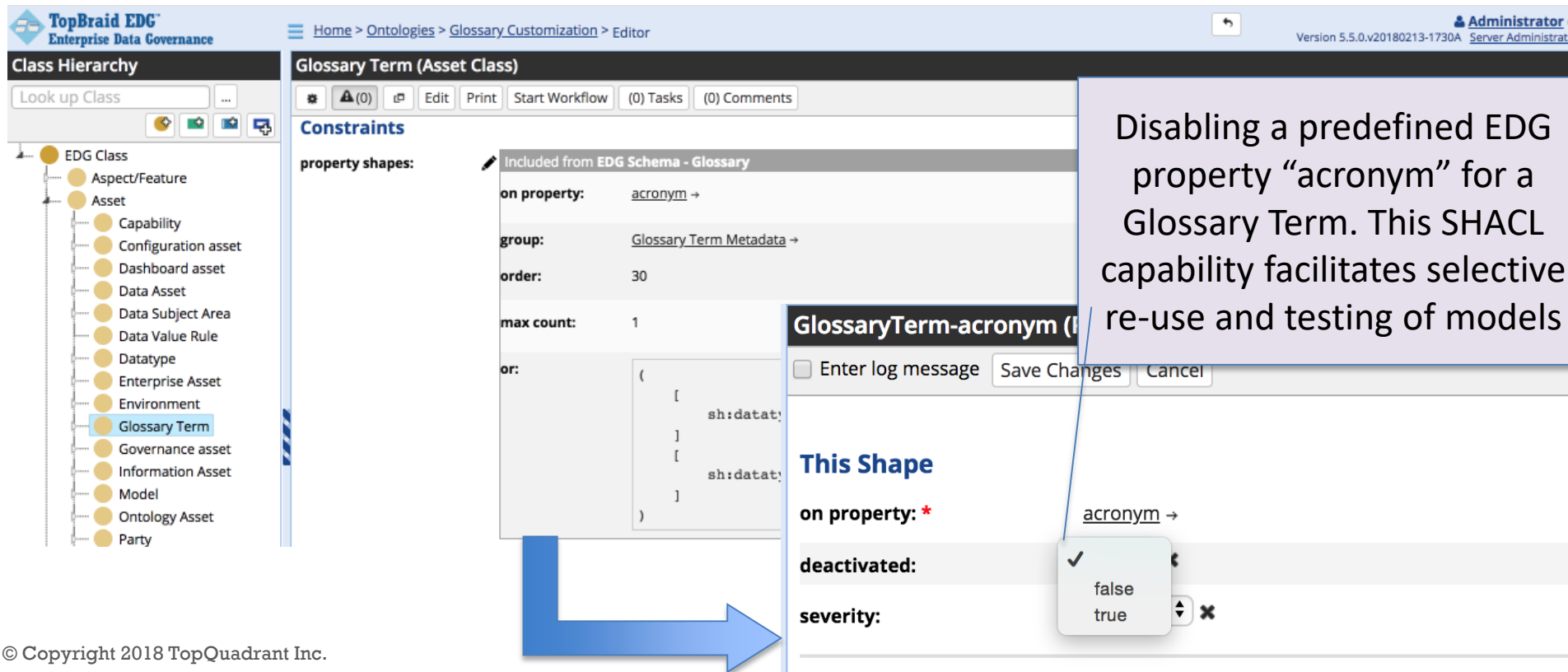
Validation Results

✘ Language does not match any of ["en", "fr"]
Resource:  Access Management
Property:  short description
Value:  The Process responsible for allowing Users to make use of IT Services, data, or other Assets. Access Management helps to protect the Confidentiality, Integrity and Availability of Assets by ensuring that only authorized Users are able to access or modify the Assets.
Suggestion: ▶ Delete the invalid statement **Apply**

✘ Value has more than 100 characters
Resource:  Access Management
Property:  short description

Save Changes Cancel

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.



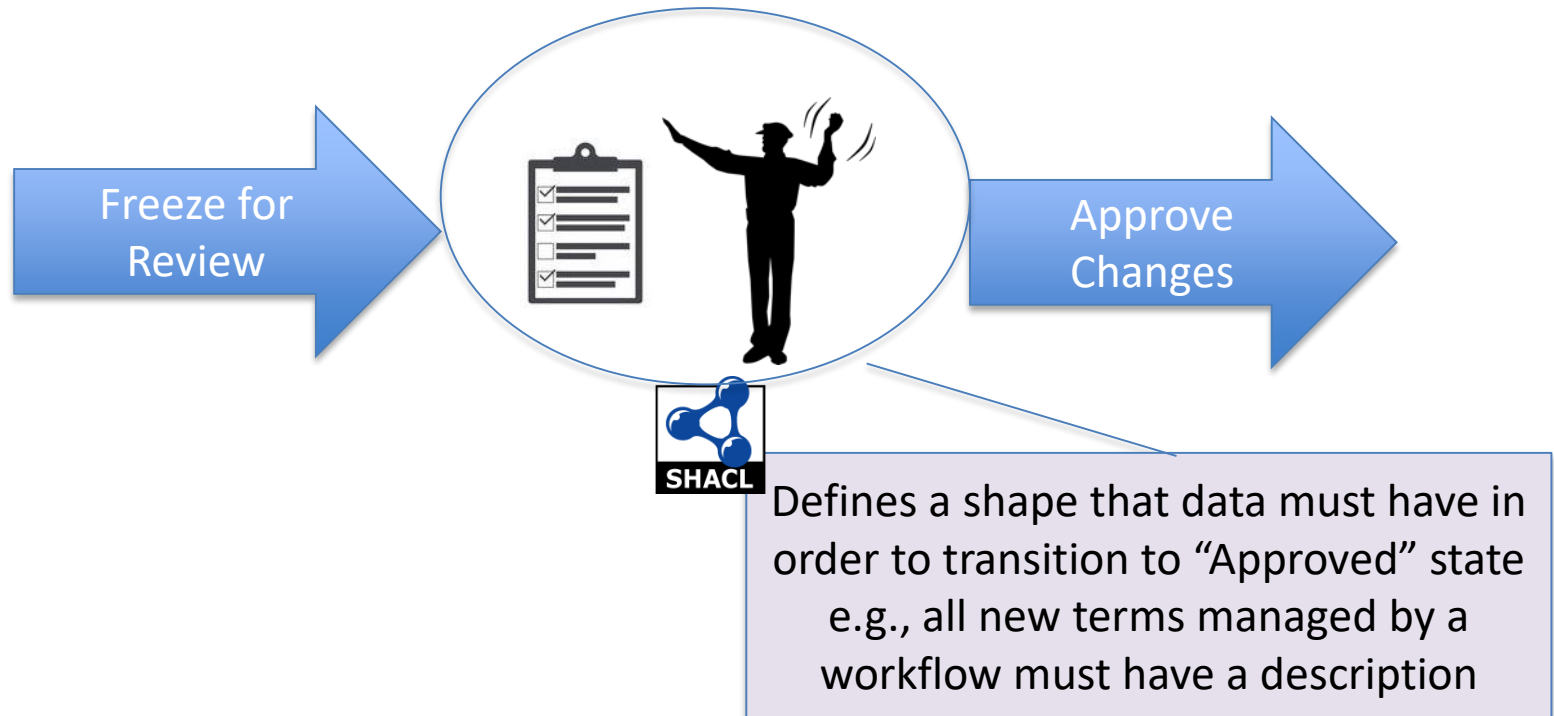
The screenshot shows the TopBraid EDG Enterprise Data Governance interface. On the left is a 'Class Hierarchy' tree with 'Glossary Term' selected. The main area is titled 'Glossary Term (Asset Class)' and shows a 'Constraints' section. A 'property shapes' table is visible, with one row for 'Included from EDG Schema - Glossary' showing constraints for the 'acronym' property. A modal dialog titled 'GlossaryTerm-acronym (I' is open, showing a 'This Shape' section with 'on property: *' set to 'acronym'. The 'deactivated:' checkbox is checked, and a dropdown menu is open showing 'false' and 'true' options. A blue arrow points from the 'deactivated:' checkbox area to the text box on the right.

Disabling a predefined EDG property "acronym" for a Glossary Term. This SHACL capability facilitates selective re-use and testing of models

TopQuadrant™ Ensure Governance Workflows

TopBraid EDG supports configurable workflows.

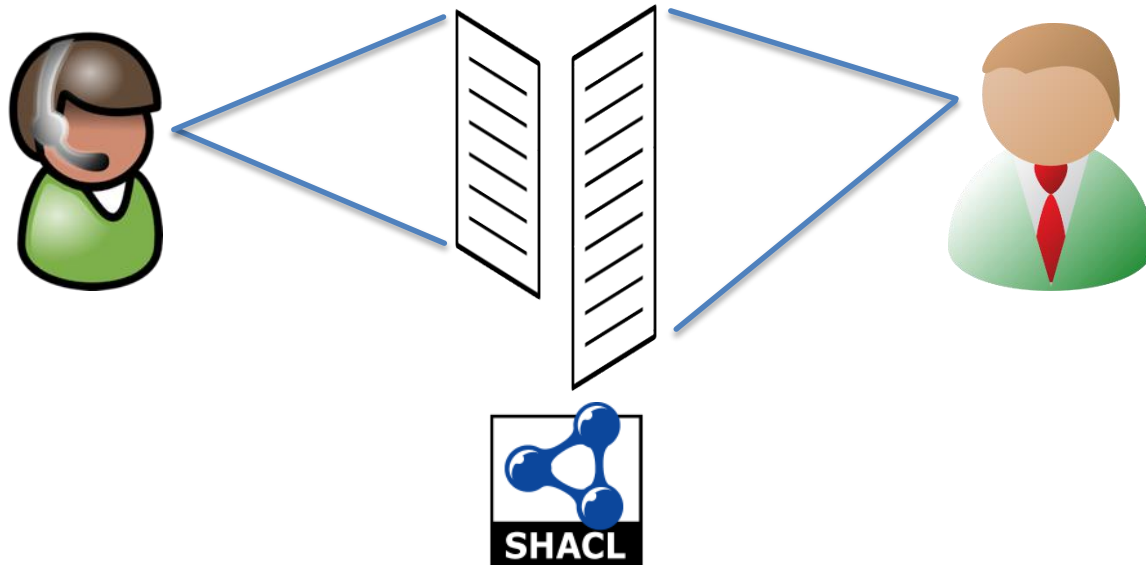
Each transition of a workflow from one state to another can include SHACL statements specifying conditions that data must satisfy in order to pass on to the next state in a workflow.



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.

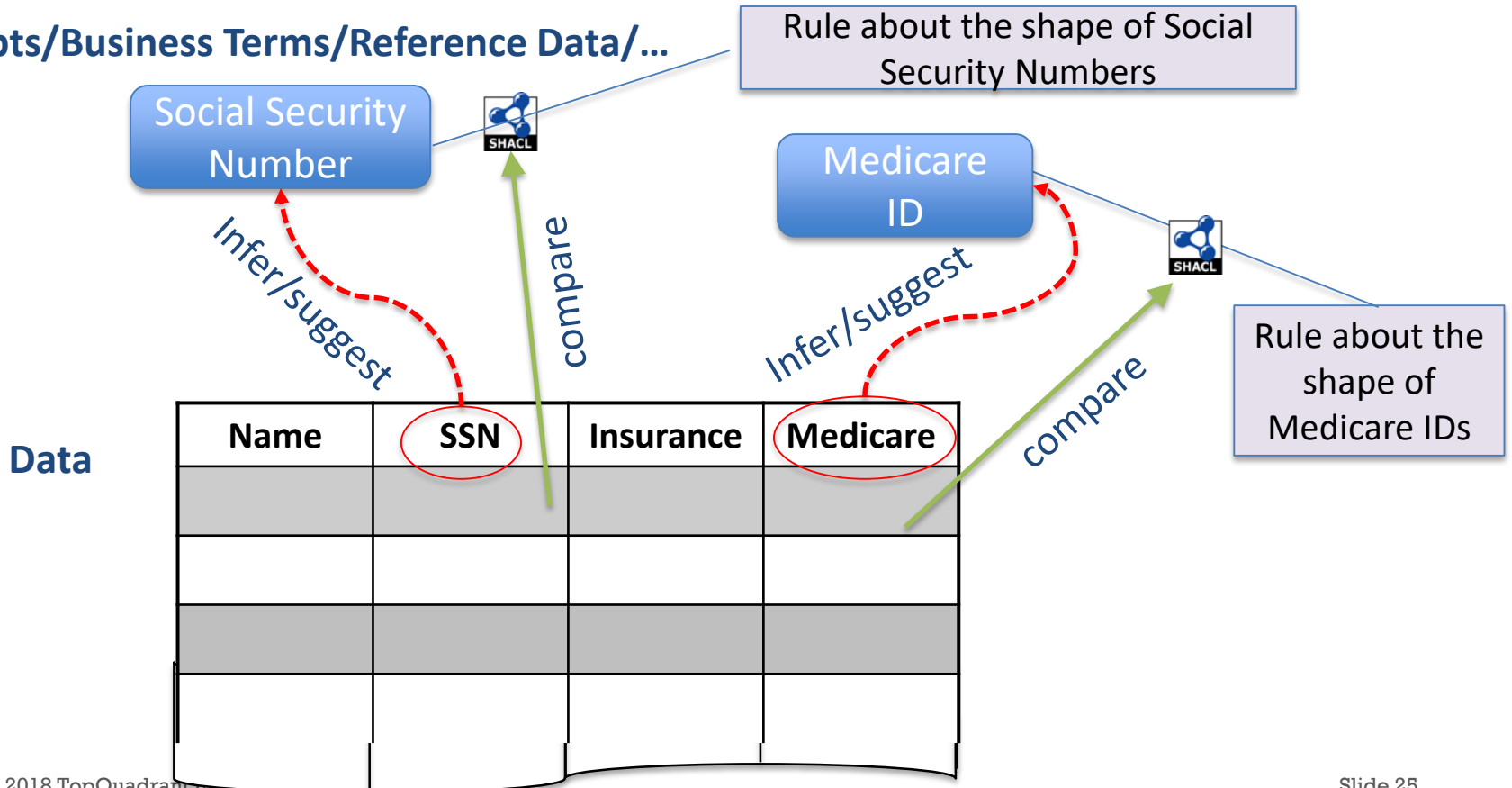


Data Discovery and Mapping Automation

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.

Concepts/Business Terms/Reference Data/...




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**

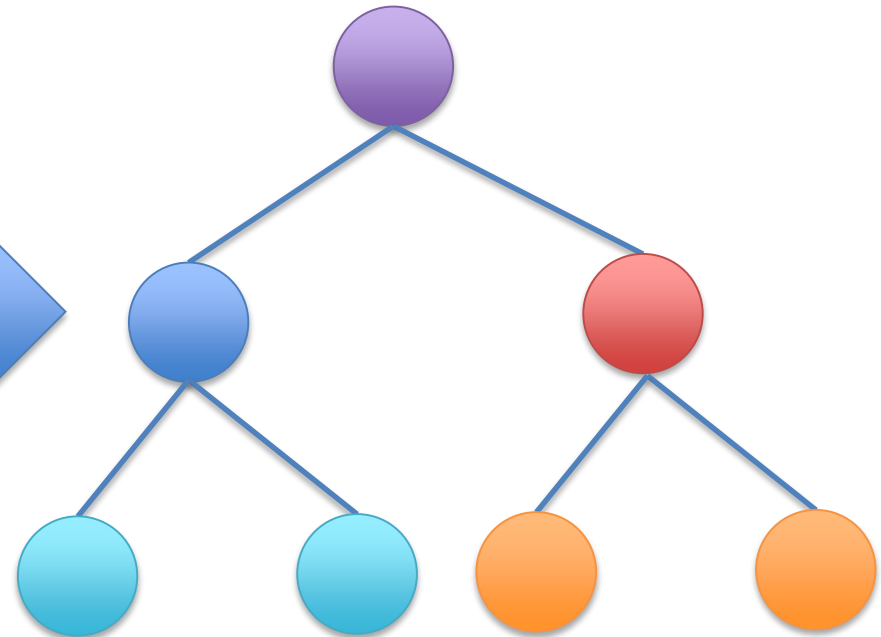
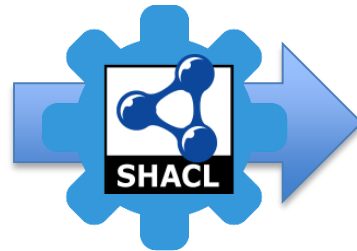
[Apply all top suggestions](#)

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

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.

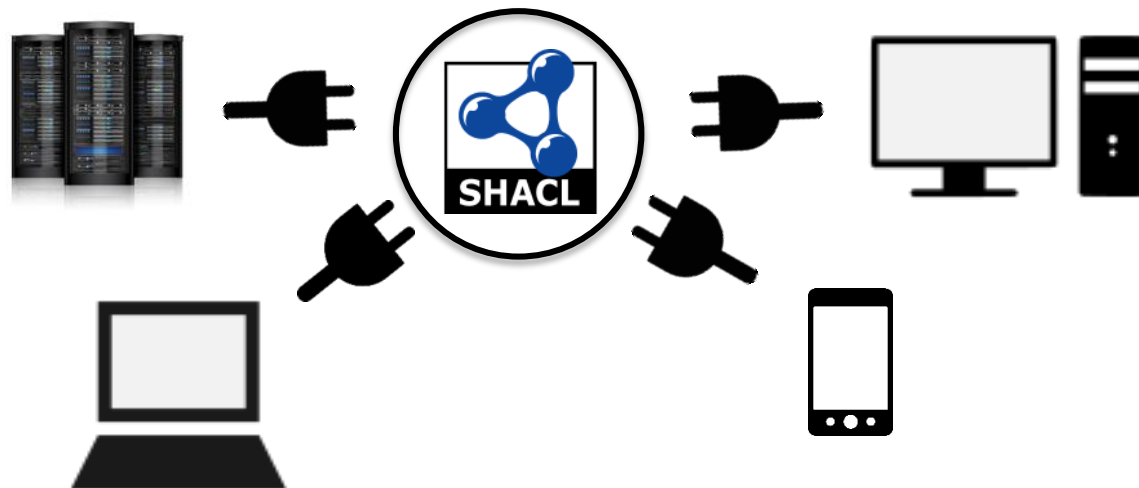
N	N	N	N

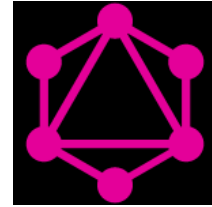


TopQuadrant™ Data Interface Configurability

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.





- 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
}
```



⚙️ ⚠️(0) Edit Print Start Workflow (0) Tasks (0) Comments Show History

<http://example.org/ontologies/new#User>
Last edited by Administrator on Mar 16, 2018 2:36:13 PM

Labels and Description

label: User
comment: A user account

Constraints

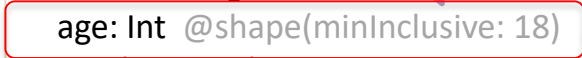
property shapes:

on property:	age
order:	1
datatype:	integer
min inclusive:	18

on property:	gender
order:	2
in:	1. MALE 2. FEMALE

on property:	name
order:	0
datatype:	string

on property:	purchases
order:	3
class:	Purchases



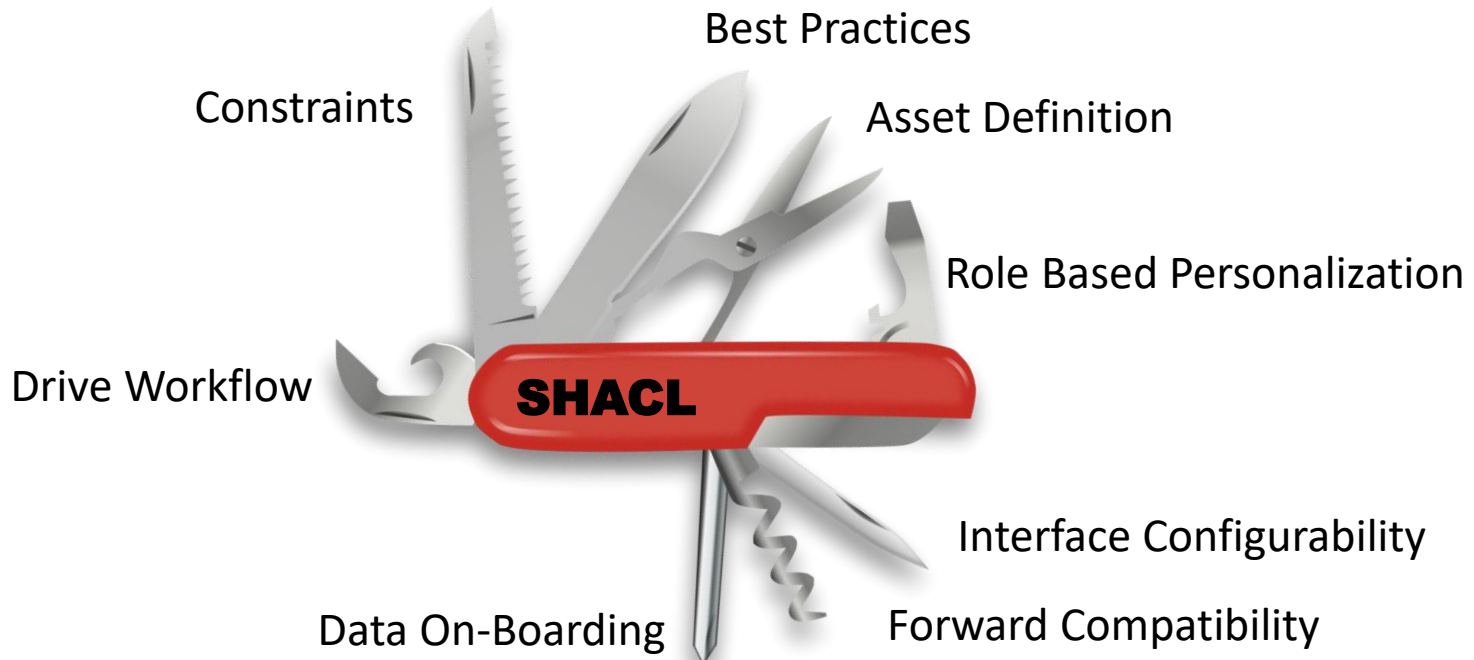
Opens Up Many Possibilities

- 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

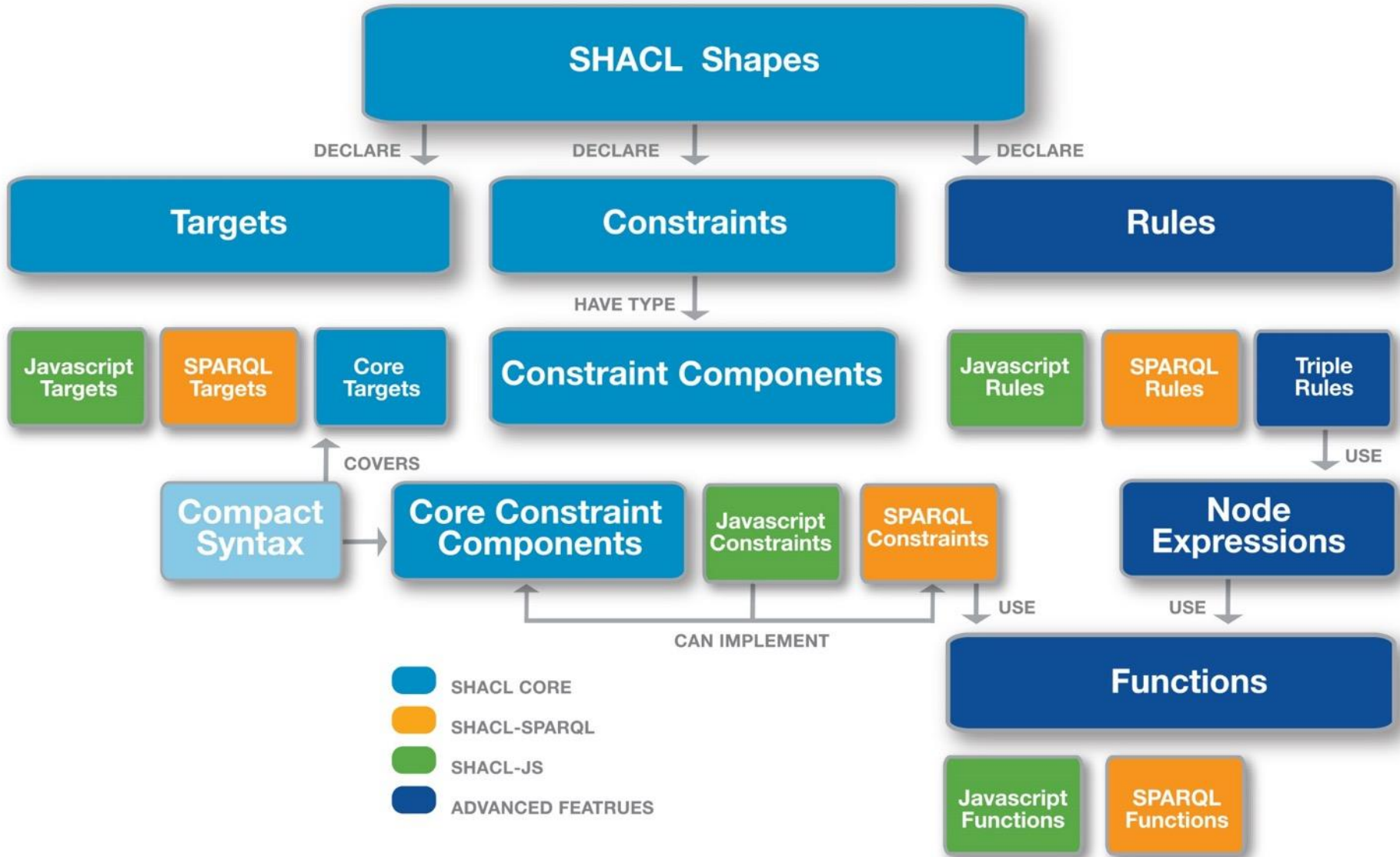
Learn SHACL Once, Use for Many Needs

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





... 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>

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