Going Live with Enterprise Solutions

TopQuadrant™

# Semantic Data Governance for GDPR Compliance

Natasa Varytimou and Jack Spivak, TopQuadrant Inc.

October 12, 2017

TopBraid EDG™

Enterprise Data Governance

# Today's Program

## *Semantic Data Governance for GDPR Compliance*

- What is GDPR? Who? What? Where? How?
- What's so hard about data protection?
- How do semantic standards help solve the hard problems in GDPR compliance?
- Demo: How TopBraid EDG, an open-standards solution, can simplify, speed and ensure compliance.
- Some Quick Polls Throughout.
- Questions and Answers

**Jack Spivak**

**Natasa Varytimou**

TopBraid EDG™
Enterprise Data Governance

© Copyright 2017 TopQuadrant Inc.

# TOPQUADRANT COMPANY

## 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

© Copyright 2017 TopQuadrant Inc.

TopQuadrant™

# ENTERPRISE CUSTOMERS

© Copyright 2017 TopQu

# General Data Protection Regulations (GDPR)

## *Mandate: Protect Personally Identifiable Information (PII)*

- ✓ 7 guiding principles and 83 pages of regulations govern the protection of personal data.
- ✓ Generally applies to all personal data of EU residents or handled by EU companies.
- ✓ Protection ”by design” requires systems for compliance, verification, audit, and notification
- ✓ Full compliance required by May 25, 2018

© Copyright 2017 TopQuadrant Inc.

TopQuadrant™

# GDPR Facts

## *7 Guiding Principles – Standard of Care*

- Lawful, Fair and Transparent Processing ..... Article 5.1a
- Specified, Fair and Legitimate Purposes ..... Article 5.1b
- Data Minimization – Adequate , Relevant, Limited to Necessary ..... Article 5.1c
- Accurate and current ..... Article 5.1d
- Minimize duration of storage ..... Article 5.1e
- Secure Processing ..... Article 5.1f
- Accountability ..... Article 5.2

© Copyright 2017 TopQuadrant Inc.

# GDPR Facts

*Violations have significant consequences*

- 20MM Euro or 4% of Global Turnover
- Prohibited from processing of critical data
- Reputation Exposure and/or Damage
- Interruption of critical data supply chain
- Business model at risk

© Copyright 2017 TopQuadrant Inc.

# GDPR Complexity

## YOUR PII CHART™

Take time to inventory the identity relationships you have with the companies, organizations, and individuals you entrust with your personally identifiable information or PII. See how your identity is a PII Chart™, a picture of relationships you've created. Once you visualize the slices of your PII, managing your identity assets becomes easier.

Personally Identifiable Information

**LEGEND**

- **SOCIAL SECURITY NUMBER**
- **CONTACT INFORMATION** (email address, physical address, telephone and mobile numbers)
- **GOVERNMENT-ISSUED IDENTIFICATION** (driver's license, passport, birth certificate, library card)
- **BIRTH DATE, BIRTH PLACE**
- **ONLINE INFORMATION** (Facebook, social media, passwords, PINs)
- **GEOLOCATION** (smartphone, GPS, camera)
- **VERIFICATION DATA** (mother's maiden name, pets' and kids' names, high school, passwords)
- **MEDICAL RECORDS INFORMATION** (prescriptions, medical records, exams, images)
- **ACCOUNT NUMBERS** (bank, insurance, investments, credit cards)

Attribution: PII Chart Trademark, Identity Theft 911 LLC, Scottsdale, AZ

© Copyright 2017 TopQuadrant Inc.

# GDPR Complexity

*GDPR compliance must answer complex questions …*

- Who are my data partners?
- What data do I share with them?
- What countries are they in?
- Who encrypts what? And how?
- How do changes in structure or architecture affect compliance?
- Do I have compliance assets in my system for those countries?
- What 3rd country jurisdictions have regulatory authority for what data and/or what data processing?

Dozens of user personas
hundreds of user stories
Thousands of design tasks

© Copyright 2017 TopQuadrant Inc.

# Which Brings Us to the Point…

*Complexity of GDPR requirements for identification, tracking and protection … not just about the data.*

It's about:

- The data
- The processing, software, software systems
- The jurisdictions where data, software and processing are hosted
- The flows of the data through systems, jurisdictions and partner relationships
- And how, requirements that need to be met change situationally

© Copyright 2017 TopQuadrant Inc.

# Need simple answer, but this isn't it.

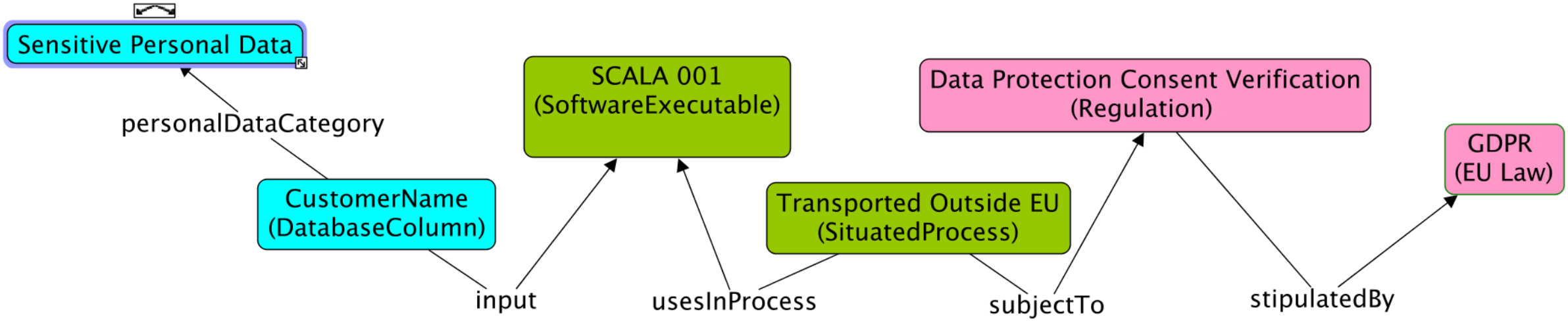
+

=

© Copyright 2017 TopQuadrant Inc.

# GDPR - What do we need to talk about?

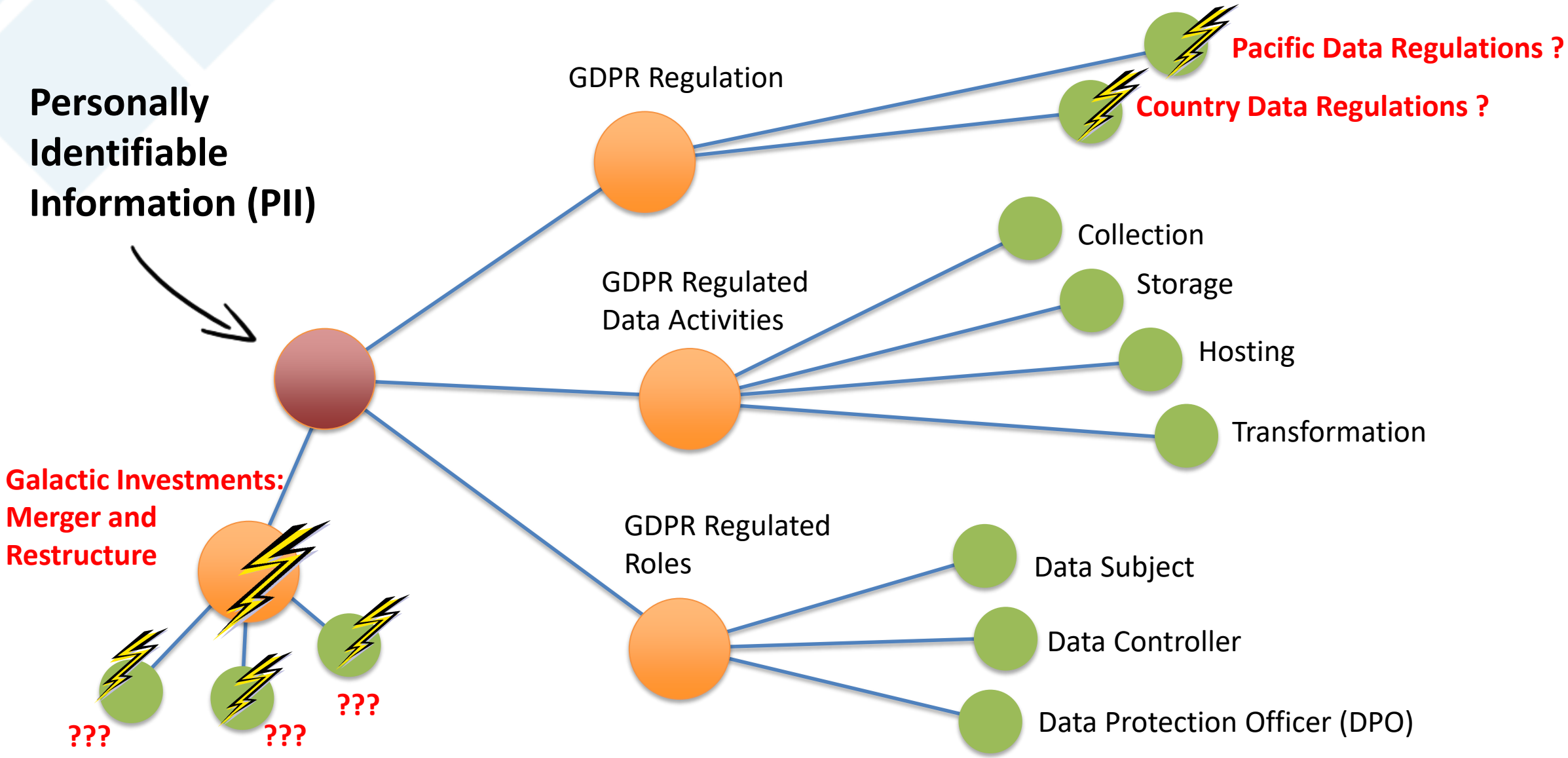
*Discovering the path between personal data …*




*… and specific GDPR obligations*

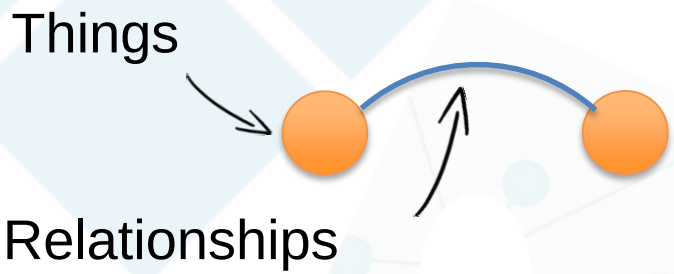
© Copyright 2017 TopQuadrant Inc.

# GDPR - What do we need to talk about?




© Copyright 2017 TopQuadrant Inc.

# Graph Representation: Value of Relationship



*Relationships …*

- provide a common language of meaning
- reveal causes and effects
- bridge domains of discourse for insight
- define “line-of-sights” for decision support
- place GDPR into a structured framework

*Relationships tame complexity and “pave the road” to compliance*

© Copyright 2017 TopQuadrant Inc.

# GDPR - What do we need to talk about?

## Regulated Data Actions

Transport Outside EU

Disclosure By Transmission

Processing

Archiving

Storage

Collection

Alteration

Adaptation

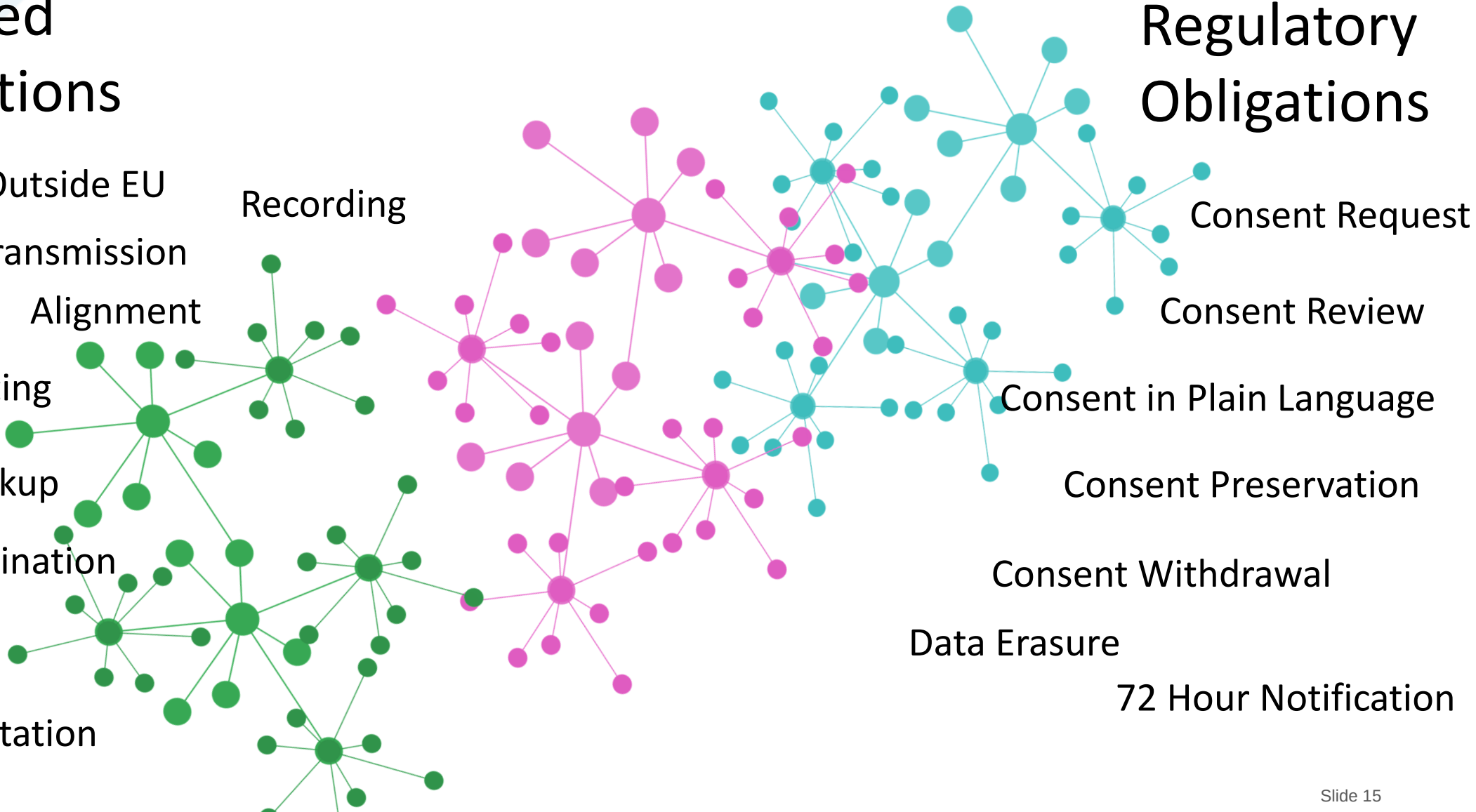
Recording

Alignment

Hosting

Backup

Combination



## Regulatory Obligations

Consent Request

Consent Review

Consent in Plain Language

Consent Preservation

Consent Withdrawal

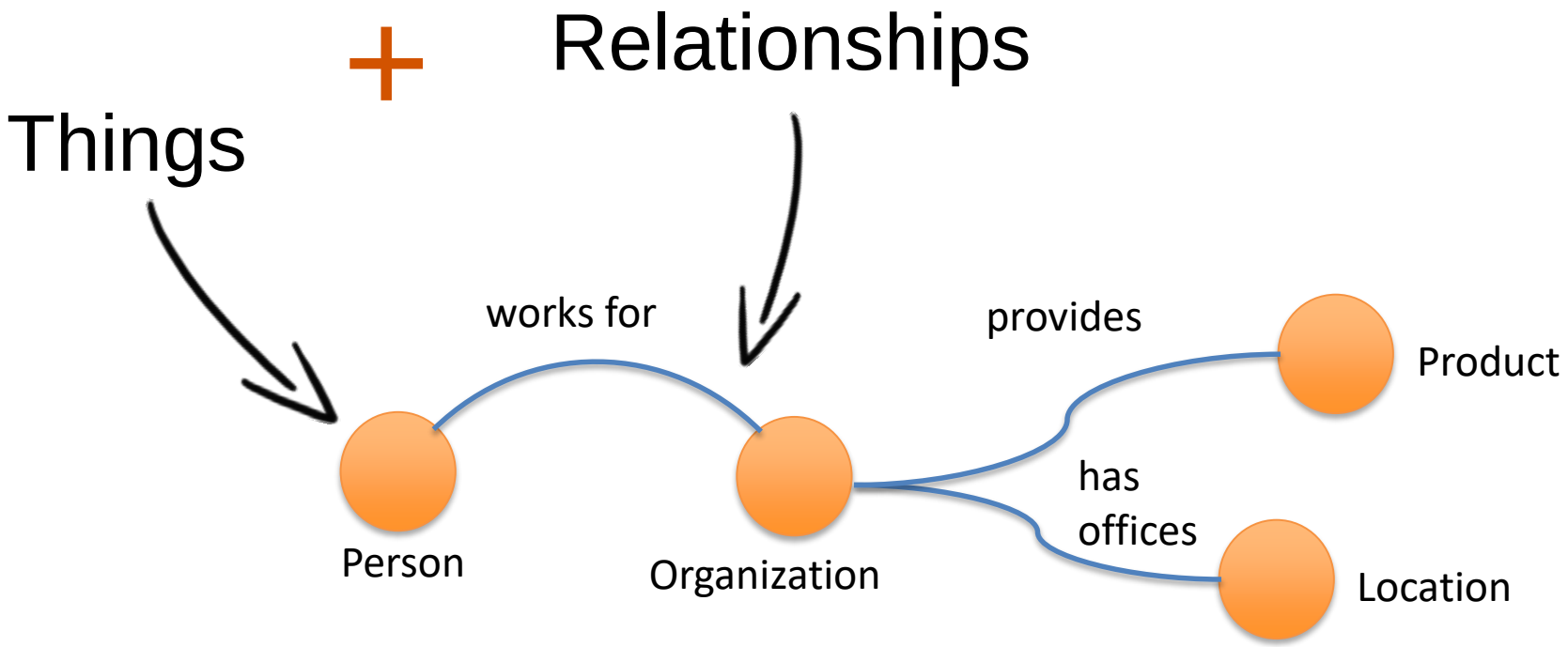
Data Erasure

72 Hour Notification

© Copyright 2017 TopQuadrant Inc.

# GDPR – Saying What We Mean

## Semantic graph =



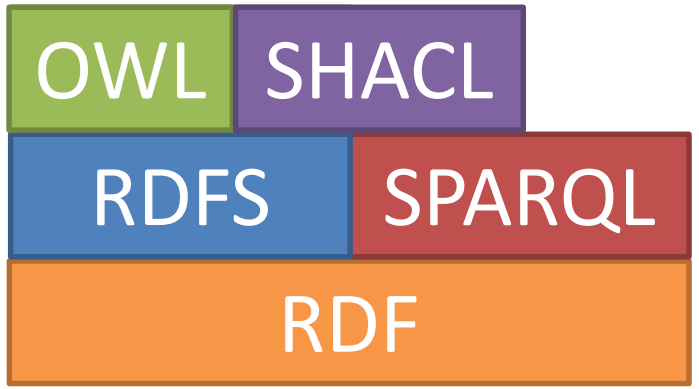

© Copyright 2017 TopQuadrant Inc.

# Standards for Meaningful Expression

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

### Classification:
*What is this thing?*

### Rules:
*Is that term used correctly?*
*What do you need to know?*
*You can't say that here!*




### Vocabulary:
*Shared terms can we use*

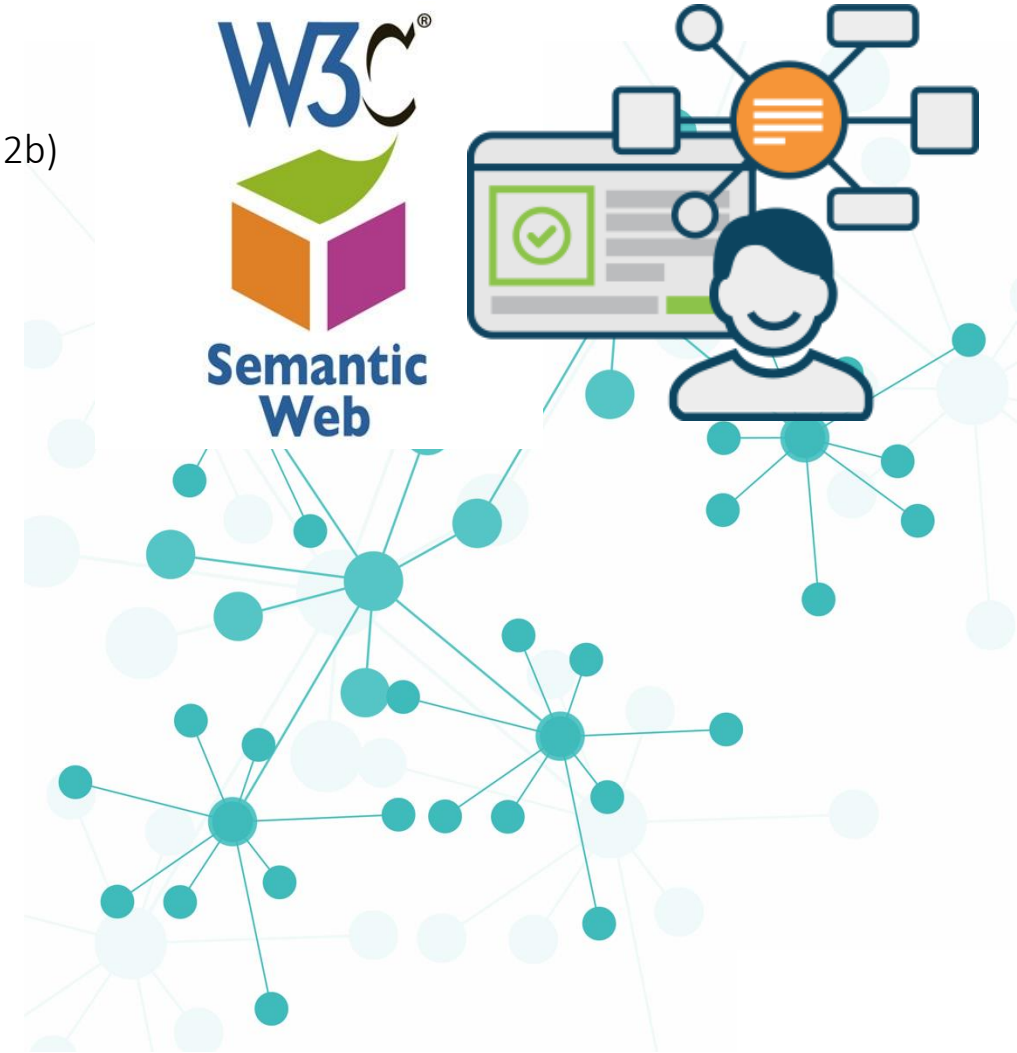
### Query:
*What did you say?*

### Statements:
*Saying things*

© Copyright 2017 TopQuadrant Inc.

# OpenSource Standards Help Compliance

- GDPR requires data ”portability” (Article 13 Para 2b)
- Interoperability with enterprise, vendor and partner systems
- Common standardized language for description and collaboration.
- Well thought out.
  - No need to re-invent the wheel.
- Extensive adoption

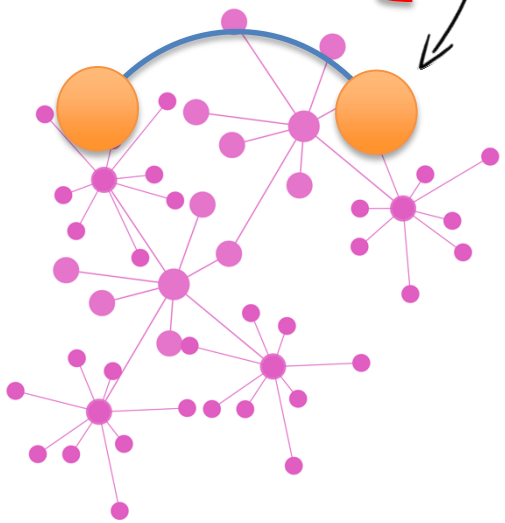



© Copyright 2017 TopQuadrant Inc.

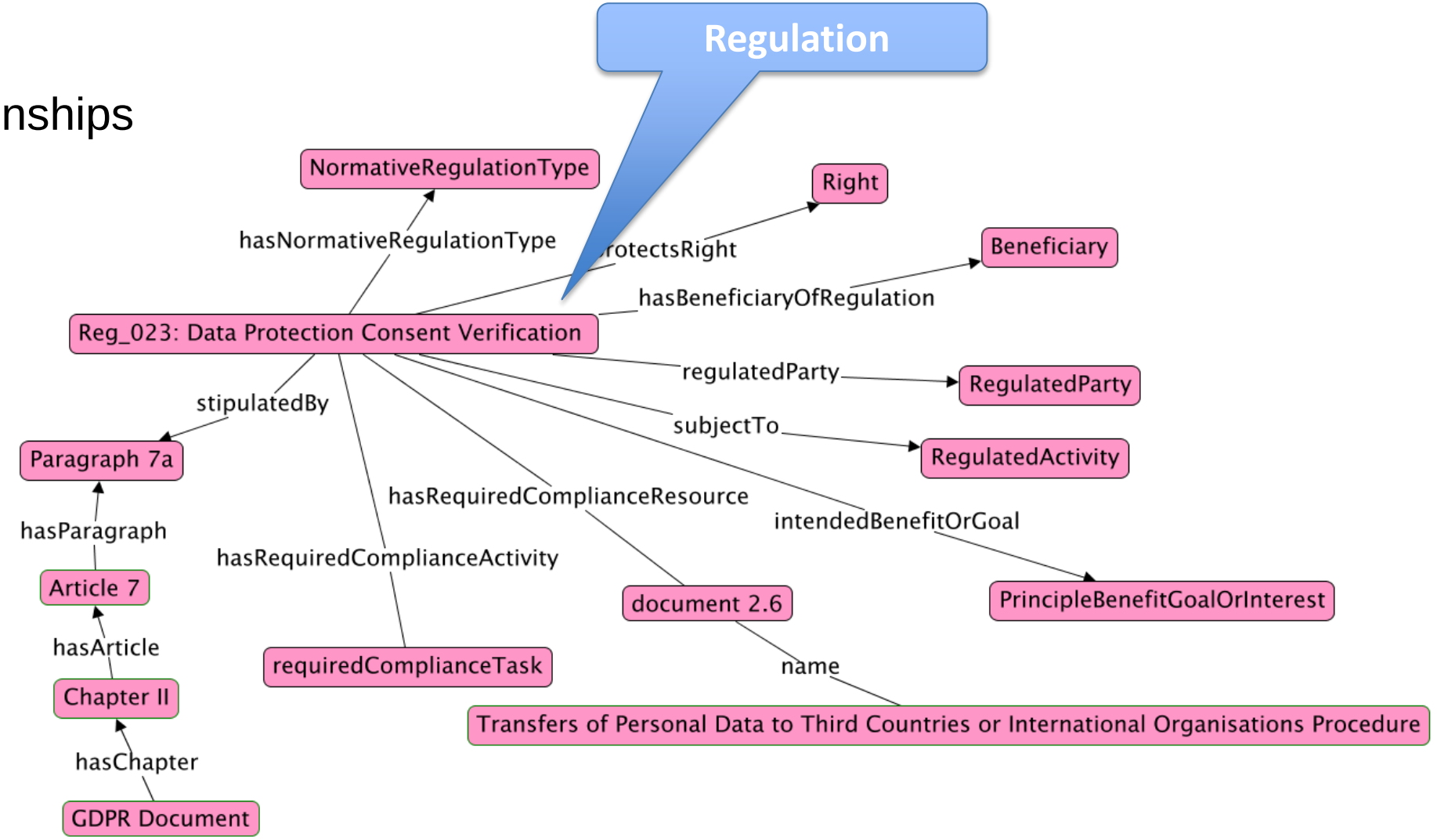
# TopBraid EDG - Standardizing “Regulation”

Regulatory

Things
Relationships



**Regulatory Compliance Graph**

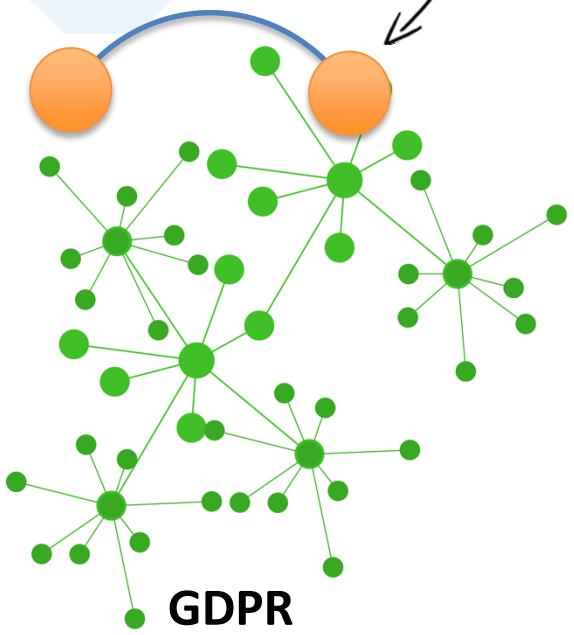



© Copyright 2017 TopQuadrant Inc.

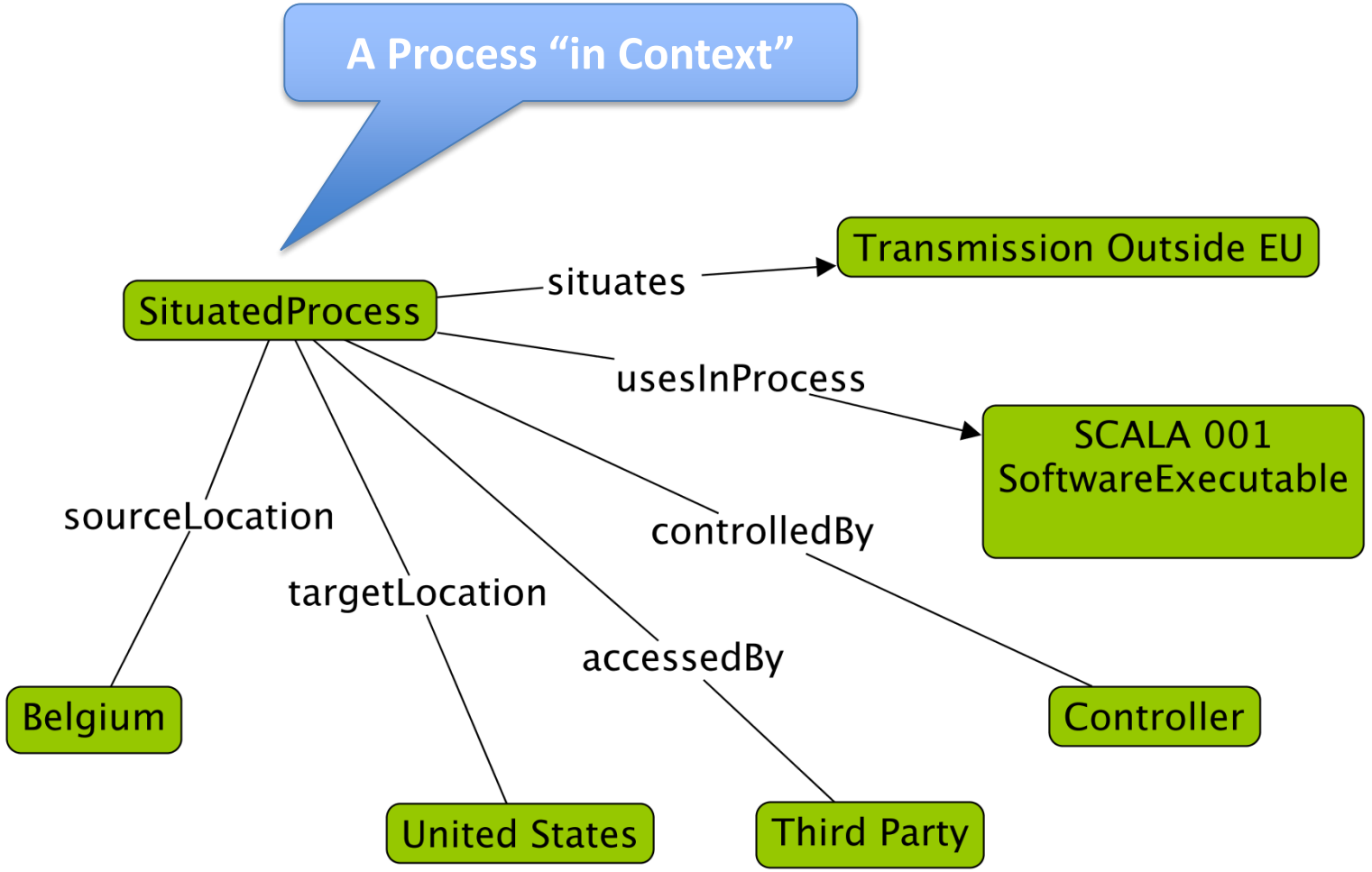
# TopBraid EDG - Standardizing “GDPR”

GDPR

Things
Relationships

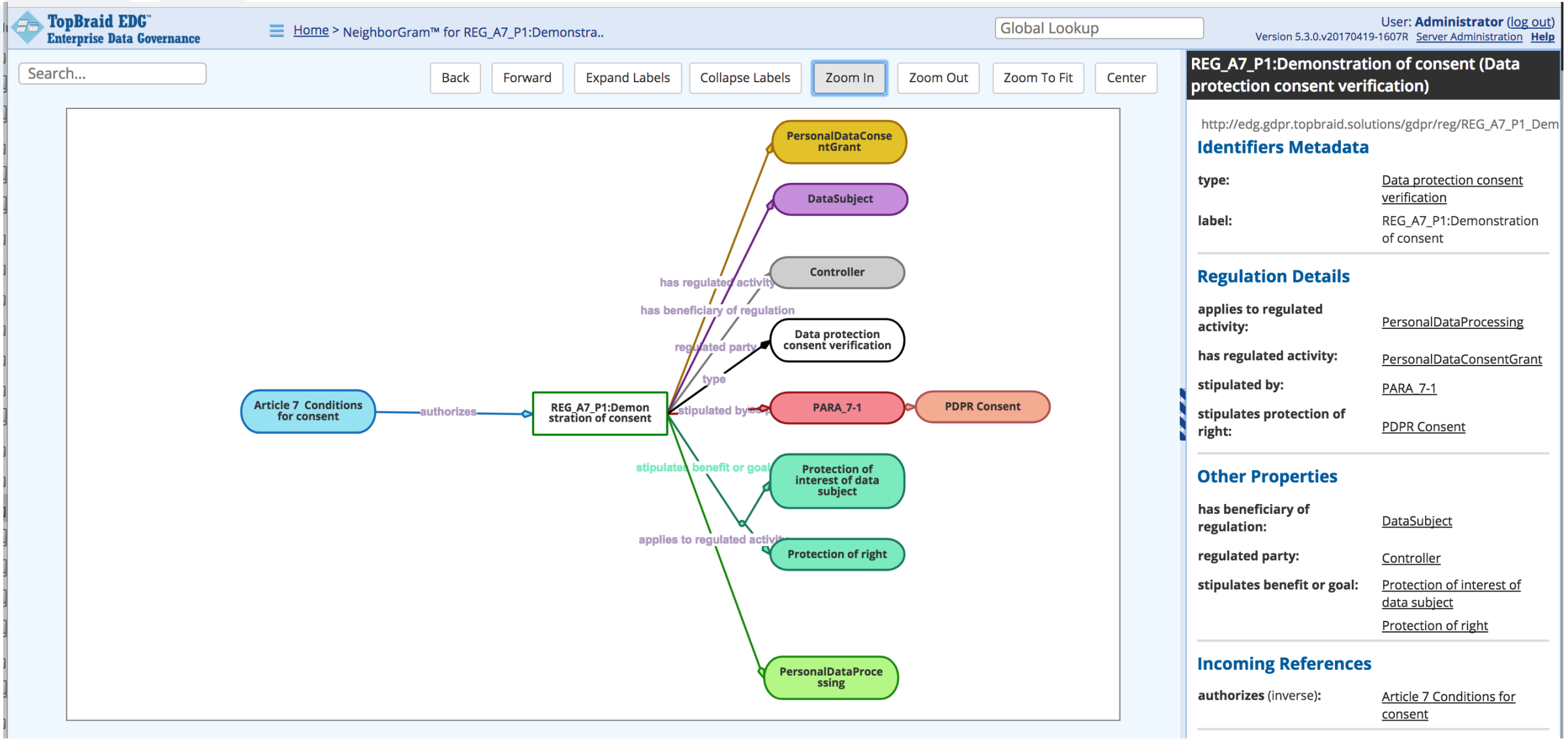


**GDPR Compliance Graph**




© Copyright 2017 TopQuadrant Inc.

# GDPR Regulation in TopBraid EDG

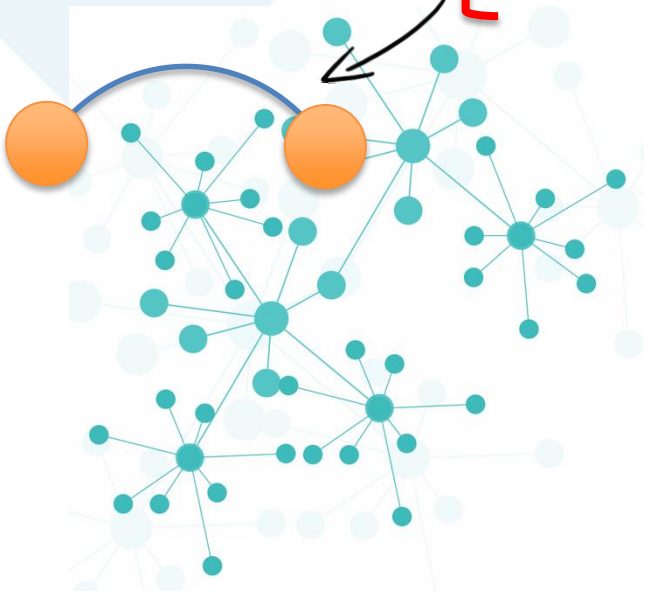



© Copyright 2017 TopQuadrant Inc.

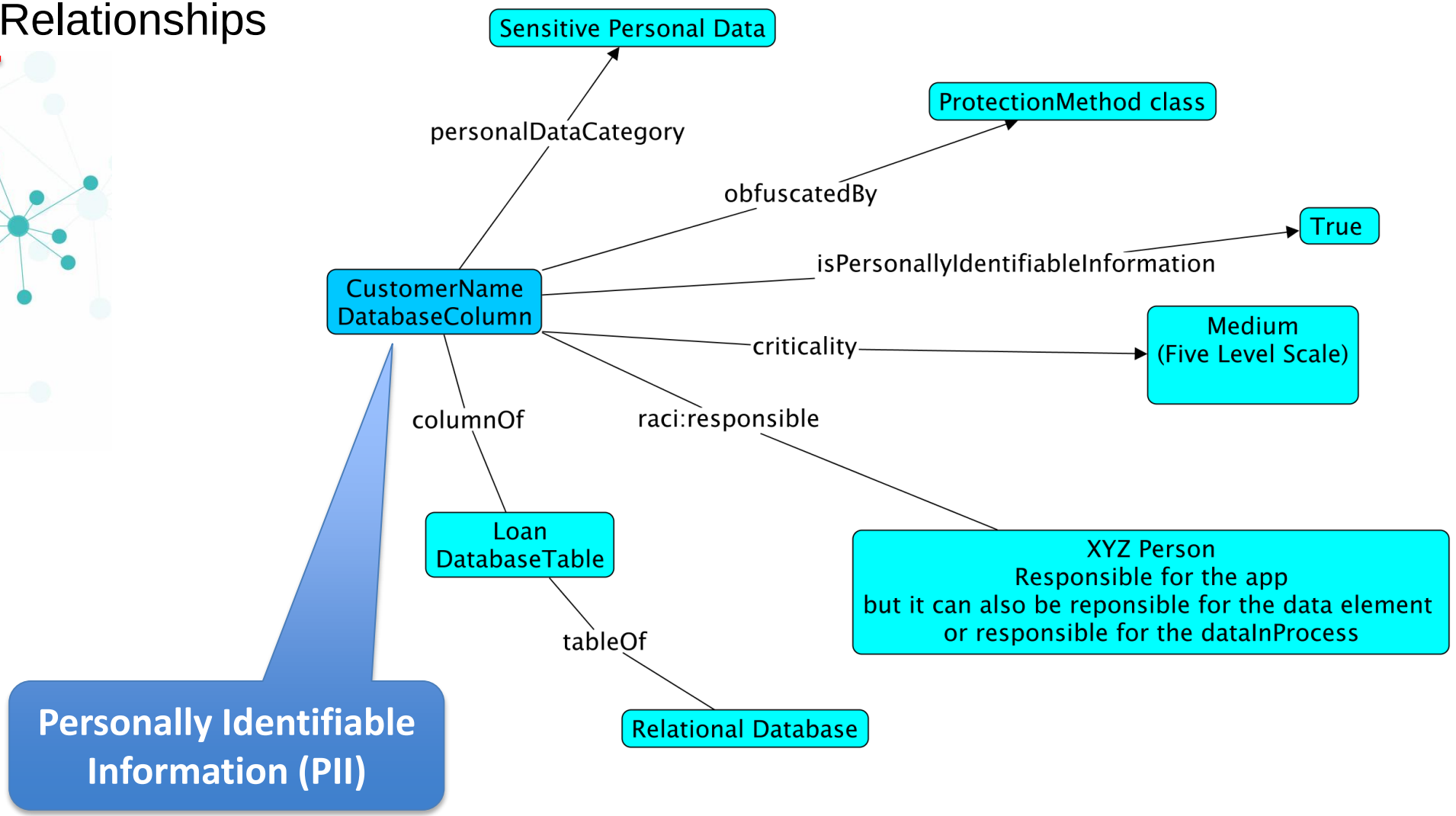
# TopBraid EDG - Standardizing “Governance”

Governance

Things
Relationships

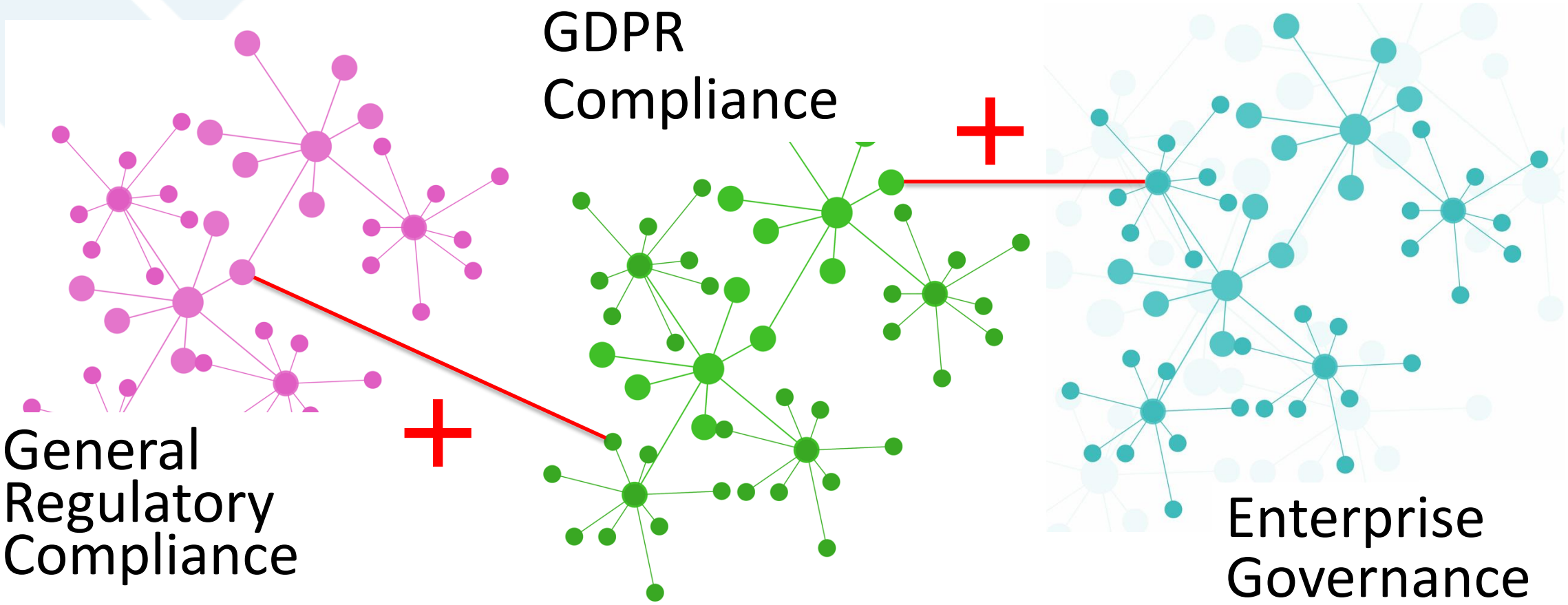


**Data, Technical and Enterprise Knowledge Graphs**




© Copyright 2017 TopQuadrant Inc.

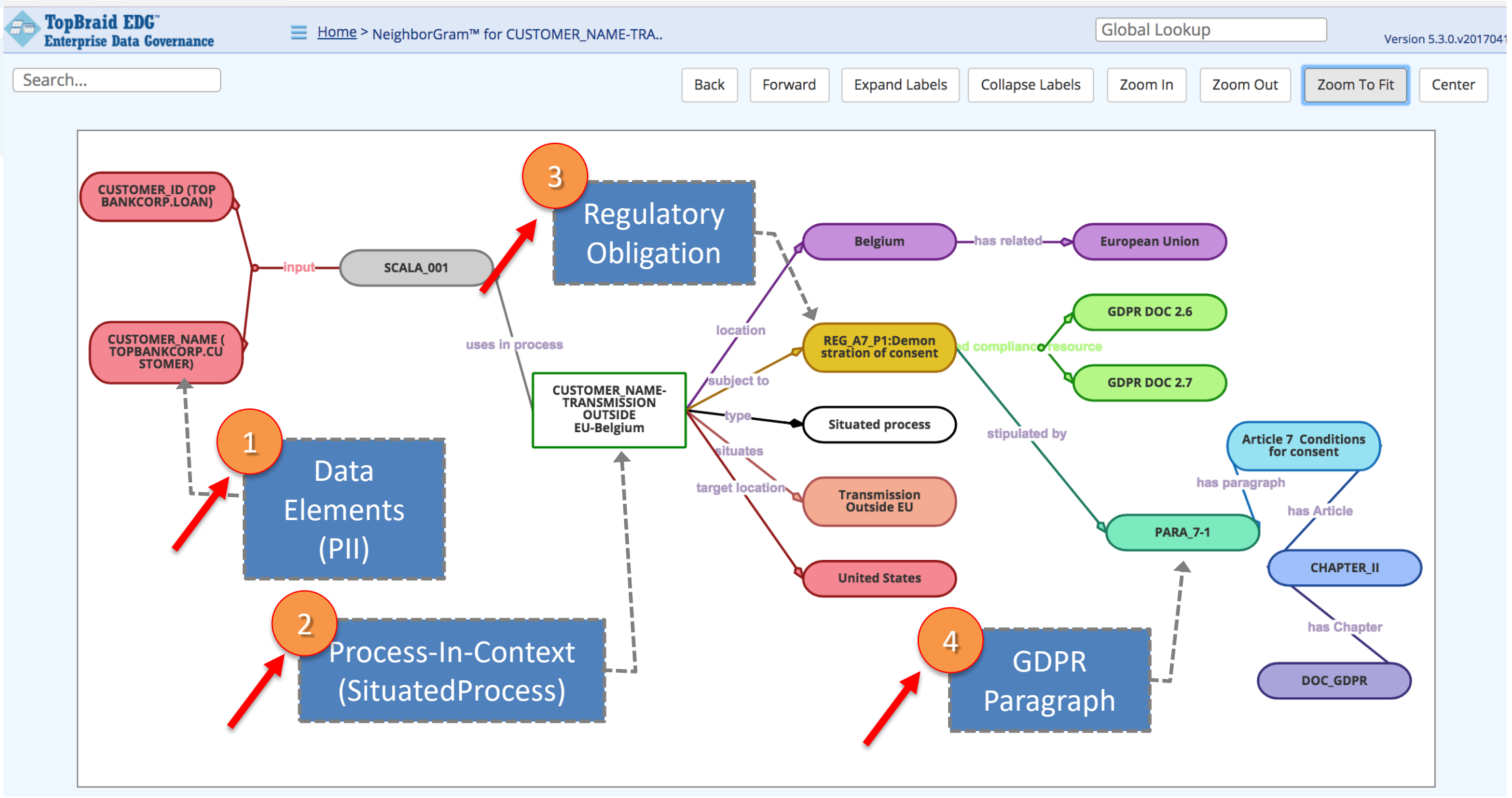
# The Power of TopBraid EDG …




*… is in bringing this all together into a connected knowledge base that can be queried for insights, reports and decision support*

© Copyright 2017 TopQuadrant Inc.

# GDPR Demo Example: “Transmission Outside EU”




© Copyright 2017 TopQuadrant Inc.

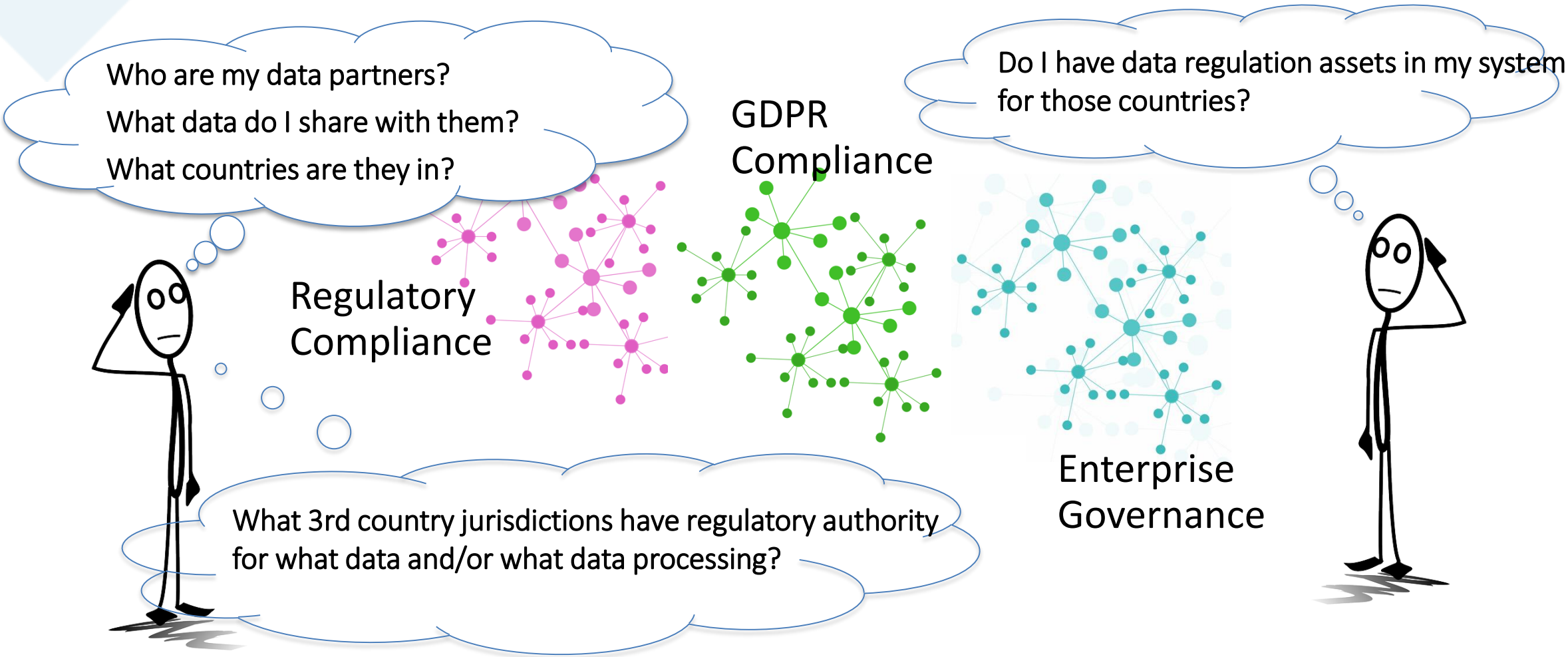
TopQuadrant™

# DEMO:
# TopBraid EDG Semantic Data Governance for GDPR Compliance

© Copyright 2017 TopQuadrant Inc.

# TopBraid EDG Knowledge Base

## *… Helps answer compliance questions*




© Copyright 2017 TopQuadrant Inc.

# Semantic Engineering for Regulatory Compliance

## Regulatory Mandates: Governance, Infrastructure, Quality

## EDM Council Position: THE DATA PROBLEM CAN BE SOLVED

"Ensuring a Unified View of Meaning and Achieving Flexible, Query-Based Analysis Can Be Solved Using Semantic Engineering"

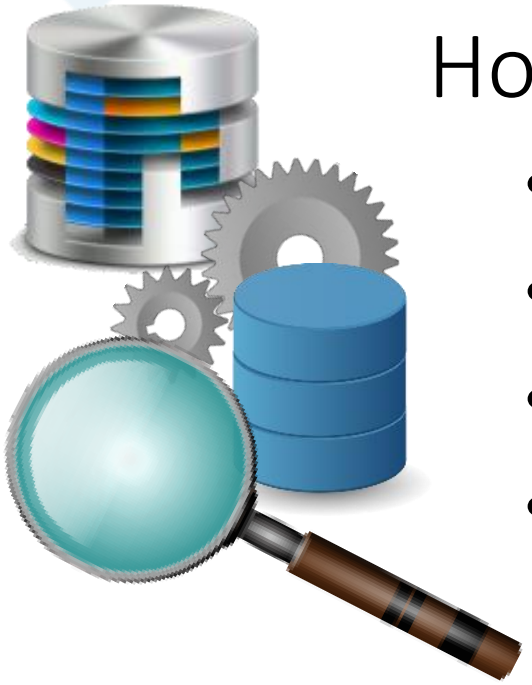
| Regulatory Actions |
|---|
| Fundamental Review of Trade Book (FRTB) |
| Dodd-Frank: Title I (systemic risk) and Title VII (derivatives) |
| European Market Infrastructure Regulation (EMIR) |
| BCBS 239: Principles of Risk Data Aggregation & Reporting |
| Comprehensive Capital Analysis and Review (CCAR) and Basel III |
| General Data Protection Regulation (GDPR) |
| Investment Book of Records (IBOR) |
| Bank Integrated Reporting Dictionary (BIRD) |
| Financial Data Standardization Project (EC) |
| Regulatory Fitness and Performance Program (REFIT) |
| Common Data Template for Systemically Important Banks (FSB) |
| Data Gaps Initiative (FSB) and Common Reporting (COREP) Template |
| Inventory of Data Reporting Requirements (DRR) |
| Markets in Financial Instruments Directive (MiFID2) |
| Capital Requirements Regulation & Directive (CCD/CDR IV) |
| Alternative Investment Fund Managers Directive (AIFMD) |
| Directive on Undertakings for Collective Investments in Transferable Securities (UCITS) |
| Solvency II (EIOPA) |

Attribution: Enterprise Data Management Council

© Copyright 2017 TopQuadrant Inc.

# TopBraid EDG Knowledge Base

## *… Helps understand*



How enterprise contexts for…

- Data Assets
- Software and systems
- Processing locations
- Third party processors

… relate to compliance

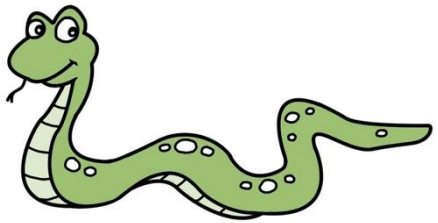
- responsibilities
- obligations
- actions needed

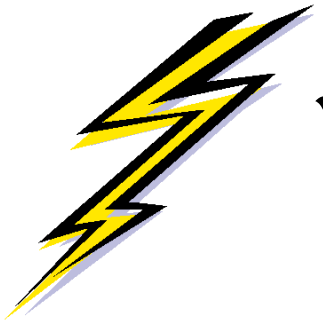
© Copyright 2017 TopQuadrant Inc.

# TopBraid EDG Knowledge Base

*… Captures knowledge about business, processes and data*

✓ Know where PII lives





✓ Know what changes it

- Uses & Intentions
- system structure
- process changes

✓ Know where it moves

- Backup, restore, disaster recovery may be done by 3rd party
- in different location

© Copyright 2017 TopQuadrant Inc.

# TopBraid EDG Knowledge Engine

*Helps automate GDPR compliance;*
*assessments, documentation, discovery of obligations, compliance gaps …*

## Machine-Process-able Standards for:

- ✓ policies, methods, procedures and workflows for performance of required actions/tasks
- ✓ informational resources language, documents, forms, templates used in workflows
- ✓ supporting systems for compliance validation & verification, change tracking, audit, etc.

© Copyright 2017 TopQuadrant Inc.

# GDPR and Data Governance

## *Article 25 – Data protection by design and by default*

… the controller shall, …

… implement ***appropriate technical and organizational measures*** …

… designed to implement data-protection principles …

… and to integrate the necessary safeguards into the processing …

… to meet the requirements of this Regulation and protect the rights of data subjects.

© Copyright 2017 TopQuadrant Inc.

# TopBraid EDG – Achieving Data Protection By Design

*Semantic Engineering and Graph Representation*

Pre-Built Infrastructure:

- Insight
- Analysis
- Governance
- Quality
- Protection
- Audit

© Copyright 2017 TopQuadrant Inc.

TopQuadrant™

TopBraid EDG™
Enterprise Data Governance

- Business Glossaries
- Reference Data Management
- Basic Metadata Management
- Advanced Metadata Management
- Vocabulary Management

© Copyright 2017 TopQuadrant Inc.

# TopBraid EDG Features

- Role-based access control
- Audit trail of change history
- Sandbox working copies
- Multi-lingual content
- Hierarchies
- Data quality rules
  - With validation, using SPIN and SHACL
- Event notifications
- Tasks and comments (optional integration with JIRA)
- Traceability within and across different asset types e.g.,:
  - Glossary terms to data elements to reference data to applications to business processes to …
- Modularity
- Custom extensions

- Configurable dashboards
- Imports/Exports
  - Some common across all assets collections, some asset collection-specific e.g., DDL import for data asset models or MultiTes import for taxonomies
- Search (parametric, faceted)
  - Within and across assets collections
- Visualization
  - Varies per asset e.g., UML-like class diagrams, NeighborGram, Lineage diagrams
- Configurable web services
- Model-driven edit widgets
  - Model driven auto-complete, in-line editing, specialized widgets and forms for OWL Manchester syntax, SPIN rules and SHACL shapes
- Extensive configurable metadata at the asset level
  - E.g., where used, version number, etc.

© Copyright 2017 TopQuadrant Inc.

TopQuadrant™

# TopBraid EDG Capabilities

- **Business glossaries** – governing business terminologies
- **Reference data management** – governing and delivering to applications code sets such as countries, currencies, etc.
- **Metadata management** – connecting technical metadata across different data sources, mapping physical to logical to conceptual models and (optionally) to glossary terms
- **Data lineage** – tracing lineage of data as it flows across data stores processed by transformation pipelines and jobs
- **Enterprise architecture** – connecting business capabilities, business processes, applications and data
- **Regulatory compliance** – addressing regulations such as GDPR, BCBS, etc.
- **Search and information retrieval** – managing terminology as ontologies and SKOS taxonomies, auto-classifying documents, enhancing search and navigation
- **Big data management** – governing big data lakes configurations and registering, connecting and locating datasets in the data lakes
- **Data catalogs** – registry of enterprise datasets

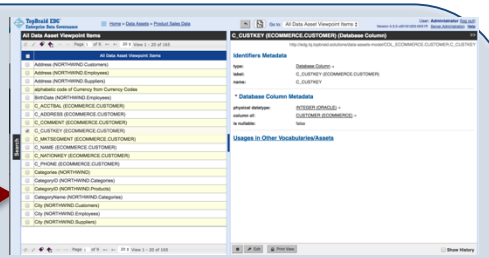
# Integration with Other Systems

TopQuadrant™

Import formats
- Spreadsheets (xls, csv, tsv)
- RDF
- XML
- UML
- RDBMs and SQL DDL
- System specific outputs e.g., Collibra, IBM Bus Glossary, MultiTes

TopBraid EDG

govern information assets



TBL Server

EDG Repository

SPARQL

Rules

Data transforms

Import & Export

Data Integration services

Web Services

Export formats
- HTML
- tab-delimited files
- RDF
- Avro, JSON
- Sharepoint, etc.

*And also export to TopBraid EDG Explorer module*

Real Time Access

direct RDF/SPARQL access

JMS

Web Services

3rd party systems

Data Warehouse

Portals

In-House Apps

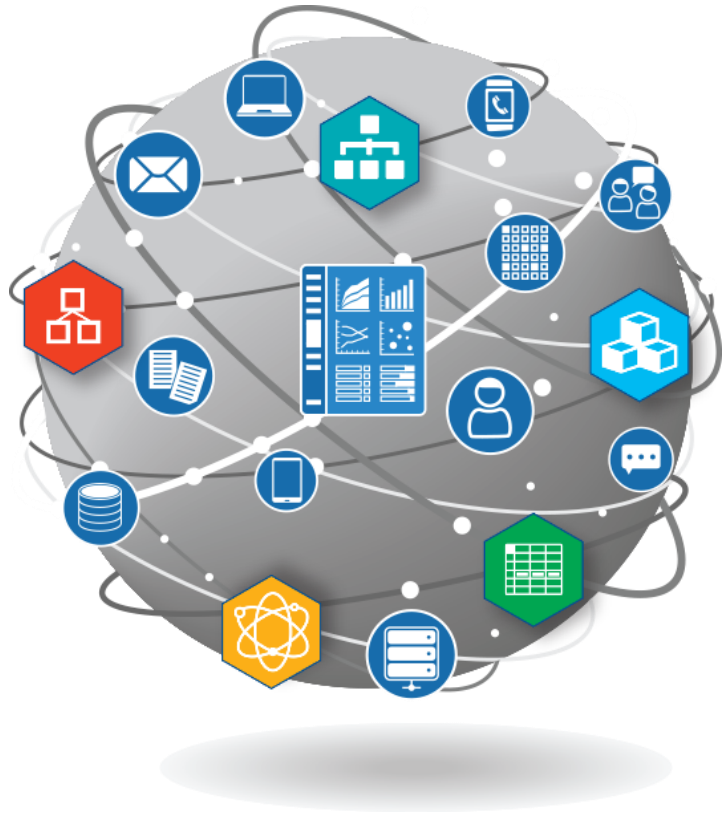
Data Stores

Import/Export formats
- RDF (Turtle, RDF/XML, JSON-LD)
- relational databases
- XML, XMLS
- RSS
- Excel spreadsheets
- delimited text (.csv, tsv)
- Web pages (RDFa, Microformats)

© Copyright 2017 TopQuadrant Inc.

# TopBraid EDG: Summary and Benefits for GDPR

- *Core flexibility and extensibility*
  Add user defined models, assets and properties as needed (model-driven)
- *Models: pre-built and user defined*
  Support multiple types of governance assets
- *Connections:*
  Can be made between any types of assets
- *Flexible Connections Enable*:
  - People (UI) and software (APIs/web services) to view, follow and query the connections to answer core questions, e.g. *"Where did this come from?"*
  - complete data governance vs. siloed data governance, i.e "reference-ability"



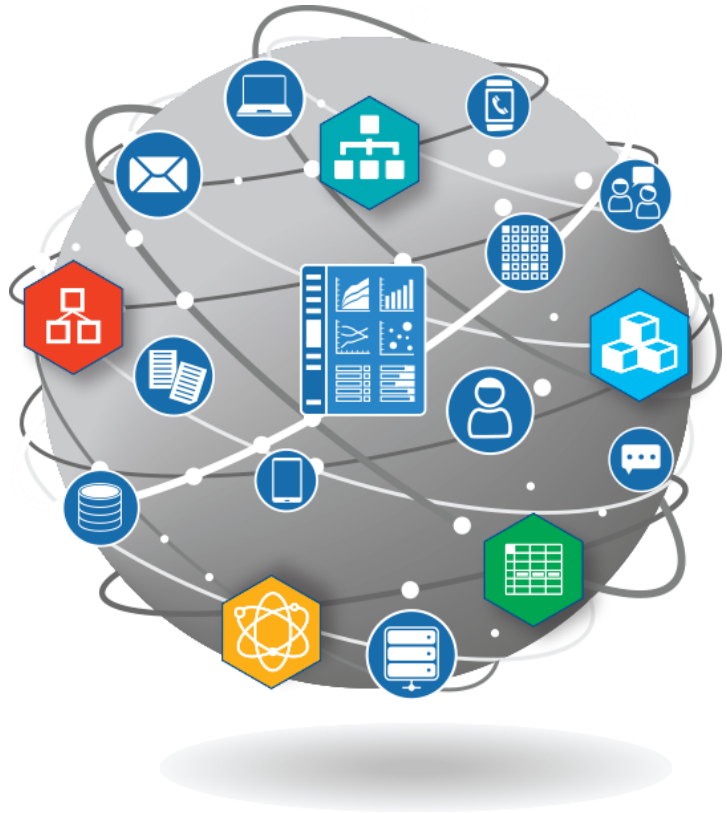
© Copyright 2017 TopQuadrant Inc.

*Flexible Connections Enable*:

# … Questions?

# TopBraid EDG: Summary and Benefits for GDPR

- *Core flexibility and extensibility*
  Add user defined models, assets and properties as needed (model-driven)
- *Models: pre-built and user defined*
  Support multiple types of governance assets
- *Connections:*
  Can be made between any types of assets
- *Flexible Connections Enable*:
  - People (UI) and software (APIs/web services) to view, follow and query the connections to answer core questions, e.g. *"Where did this come from?"*
  - complete data governance vs. siloed data governance, i.e "reference-ability"



© Copyright 2017 TopQuadrant Inc.

# Thank You !

## To Learn More …

*Contact us:* at info@topquadrant.com to:

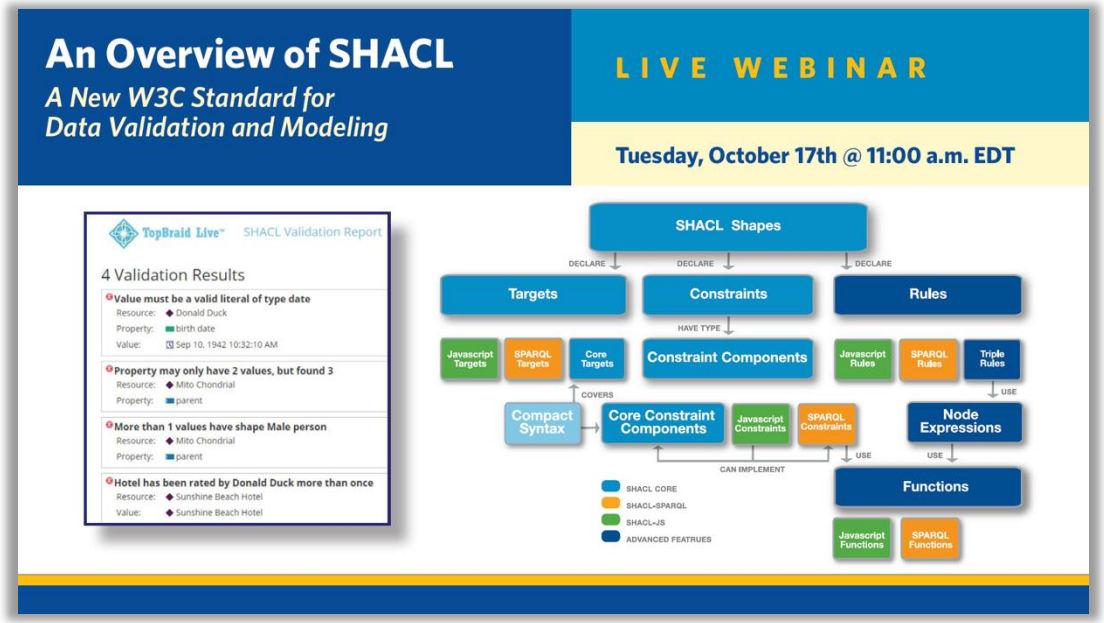
- Discuss our GDPR compliance solutions
- Request a more targeted demo of TopBraid EDG
- Ask for a free EDG evaluation account

## *EDG Product Info:*

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

## *Other EDG demos/webinar recordings:*

- 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




**Register now!**: **topquadrant.com/webinar-SHACL-overview**

© Copyright 2017 TopQuadrant Inc.