Semantic Data Governance for GDPR Compliance

Natasa Varytimou and Jack Spivak, TopQuadrant Inc.

May 18, 2017
Semantic Data Governance for GDPR Compliance

- 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.
- Quick Poll
- Questions and Answers
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
ENTERPRISE CUSTOMERS
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.
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
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
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.

Attribution: PII Chart Trademark, Identity Theft 911 LLC, Scottsdale, AZ
GDPR complexity 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?
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
Need simple answer, but this isn’t it.
GDPR - What do we need to talk about?

Discovering the path between personal data ...

... and specific GDPR obligations
GDPR - What do we need to talk about?

Personally Identifiable Information (PII)

Galactic Investments: Merger and Restructure

GDPR Regulation

- Collection
- Storage
- Hosting
- Transformation

GDPR Regulated Data Activities

- Data Subject
- Data Controller
- Data Protection Officer (DPO)

GDPR Regulated Roles

- Pacific Data Regulations?
- Country Data Regulations?
Graph Representation: Value of Relationship

- 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
GDPR - What do we need to talk about?

Regulated Data Actions
- Transport Outside EU
- Disclosure By Transmission
- Processing
- Archiving
- Storage
- Collection
- Alteration
- Adaptation

Regulatory Obligations
- Consent Request
- Consent Review
- Consent in Plain Language
- Consent Preservation
- Consent Withdrawal
- Data Erasure
- 72 Hour Notification

Recording
Alignment
Hosting
Backup
Combination
Storage
Archiving
Processing
Collection
Alteration
Adaptation
Semantic graph =

Things + Relationships

Person works for Organization
Organization provides Product
Product has offices Location

GDPR – Saying What We Mean
Standards for Meaningful Expression

Classification: What is this thing?

Vocabulary: Shared terms can we use

Statements: Saying things

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

Query: What did you say?

*W3C = World Wide Web Consortium led by Tim Berners-Lee
• 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
TopBraid EDG - Standardizing “Regulation”

Regulatory Compliance Graph

Things Relationships

Regulation

NormativeRegulationType

hasNormativeRegulationType

Right

NormativeRegulationType

hasBeneficiaryOfRegulation

Beneficiary

RegulatedParty

RegulatedActivity

regulatedParty

subjectTo

transfersOfPersonalDataToThirdCountriesOrInternationalOrganisationsProcedure

intendedBenefitOrGoal

name

document 2.6

requiredComplianceTask

hasRequiredComplianceActivity

requiredComplianceActivity

hasRequiredComplianceResource

Paragraph 7a

stipulatedBy

Article 7

hasParagraph

Chapter II

hasArticle

hasChapter

GDPR Document

Transfers of Personal Data to Third Countries or International Organisations Procedure
TopBraid EDG - Standardizing “GDPR”

A Process “in Context”

- SituatedProcess
  - sourceLocation
  - targetLocation
  - accessedBy
  - controlledBy
  - usesInProcess
  - situates
    - Transmission Outside EU

- Belgium
- United States
- Third Party
- Controller
- SCALA 001 SoftwareExecutable
GDPR Regulation in TopBraid EDG
TopBraid EDG - Standardizing "Governance"

Data, Technical and Enterprise Knowledge Graphs

Governance

Things Relationships

Personally Identifiable Information (PII)

CustomerName DatabaseColumn

Loan DatabaseTable

Sensitive Personal Data

ProtectionMethod class

medium (Five Level Scale)

isPersonallyIdentifiableInformation

criticality

obfuscatedBy

columnOf

raci:responsible

tableOf

Relational Database

XYZ Person

Responsible for the app but it can also be responsible for the data element or responsible for the dataInProcess
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.
GDPR Demo Example: “Transmission Outside EU”
DEMO: TopBraid EDG Semantic Data Governance for GDPR Compliance
Who are my data partners?
What data do I share with them?
What countries are they in?

Do I have data regulation assets in my system for those countries?

What 3rd country jurisdictions have regulatory authority for what data and/or what data processing?
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
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
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.
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.
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.
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

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

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

And also export to TopBraid EDG Explorer module

3rd party systems
- Data Warehouse
- Portals
- In-House Apps
- Data Stores

Real Time Access

TBL Server
- SPARQL
- Rules
- Data transforms
- Import & Export
- Web Services
- Data Integration services

govern information assets

TopBraid EDG

Direct RDF/SPARQL access

Sharepoint

Data Integration services

Data transforms

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

... Questions?
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:**

**Other EDG demos/webinar recordings:**
  Webinar: Data Governance for the Connected Enterprise: TopBraid EDG in Action
  Webinar: Metadata Management is Key to Data Governance Initiatives
TopBraid EDG GDPR Solutions

Generate GDPR Knowledge Base
Generate GDPR Knowledge Base based on Compliance asset triples

Compliance Asset as a Graph
Exports the entire Compliance Asset in one of the formats below.

- JSON-LD
- N-Triples
- Turtle

Captured Process

Identifiers Metadata
- type: SituatedProcess
- label: "CUSTOMER_NAME_TRANSITION_OUTSIDE_EU-Belgium"
- Situated Process Details
- uses in process: "SOX, PCI"
- located in: "Belgium"
- Uses in other Asset Collections
SHACL – A W3C Standard for powering Compliance Engines

SHACL Shapes

Targets
(Mechanism to select focus nodes)

- JavaScript Targets
- SPARQL Targets
- Core Targets
  (sh:targetClass, etc)

Constraints
(Restrictions on focus nodes)

- Core Targets
  (sh:targetClass, etc)
- JavaScript Constraints
- SPARQL Constraints

Constraint Components
(Define validation instructions)

- Core Constraint Components
  (sh:minCount, sh:class, etc)
- JavaScript Constraints
- SPARQL Constraints

Rules
(Inferences, mappings)

- JavaScript Rules
- SPARQL Rules
- Triple Rules

Node Expressions

- use
- use

Functions

- use
- use

Compact Syntax

can implement

declare

declare

declare

SHACL Core
Advanced SHACL
SHACL-SPARQL
SHACL-JS

© Copyright 2017 TopQuadrant Inc.