Master Thesis: A Process for Automated Creation of Records of Processing Activities Using Enterprise Architecture Models

Ahmet Tanakol, 08.04.2019, Final Presentation, Munich

Chair of Software Engineering for Business Information Systems (sebis)
Faculty of Informatics
Technische Universität München
wwwmatthes.in.tum.de

Outline

1. Motivation
   - GDPR
   - Record of Processing Activities
   - The state-of-the-art Approach
   - Problem
2. Approach
   - Enterprise Architecture
   - Using EA Models as an Information Source
   - Analyzing EA Models
3. Evaluation
4. Conclusion

Motivation

GDPR

GDPR: General Data Protection Regulation

- Enforce privacy and protection over data for all individuals within the EU
- Allow individuals to have control over their personal data

What is personal data?

Article 4: ‘personal data’ means any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person [1];

1. "EUR-Lex – 32016L0680 – EN – EUR-Lex". eur-lex.europa.eu.

Motivation

Article 5 - Accountability

According to Article 24 (Responsibility of the controller), the GDPR not only demands compliance, but also requires you to demonstrate compliance.

2. "Privacy notices under the EU General Data Protection Regulation". ico.org.uk.

Motivation

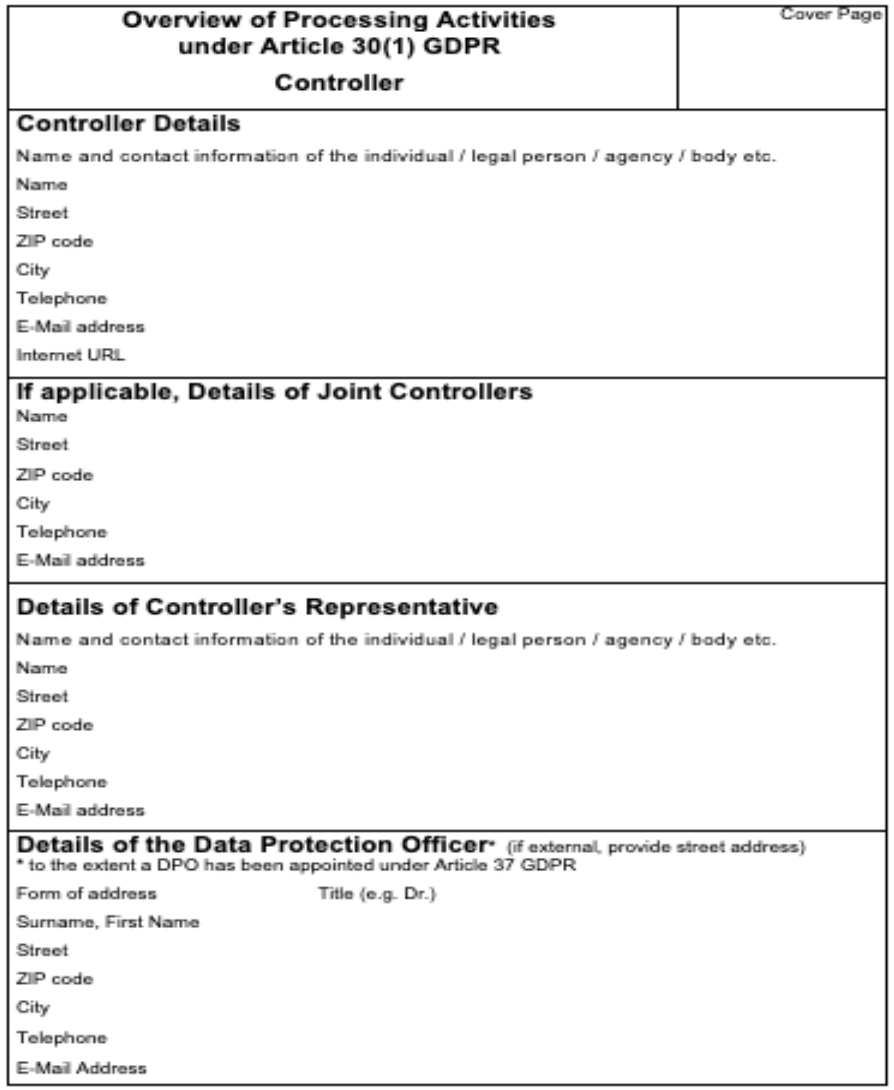
Article 30 - Records of Processing Activities

Under Article 30, organizations (process personal data) have to maintain a record of their processing activities.

A document to show GDPR compliance.

It gives information about what you do with personal data.

- How personal data is used
- Why it was collected
- How it is processed
- Who has access
- Where it is stored



Overview of Processing Activities under Article 30(1) GDPR

Controller

Cover Page

Controller Details

Name and contact information of the individual / legal person / agency / body etc.

Name
Street
ZIP code
City
Telephone
E-Mail address
Internet URL

If applicable, Details of Joint Controllers

Name
Street
ZIP code
City
Telephone
E-Mail address

Details of Controller's Representative

Name and contact information of the individual / legal person / agency / body etc.

Name
Street
ZIP code
City
Telephone
E-Mail address

Details of the Data Protection Officer* (if external, provide street address)

* to the extent a DPO has been appointed under Article 37 GDPR

Form of address Title (e.g. Dr.)
Surname, First Name
Street
ZIP code
City
Telephone
E-Mail Address

3. Feltz, Daniel 2018. German DPAs Publish Model GDPR Processing Records

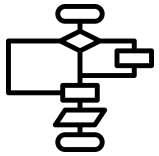
Motivation

The state-of-the-art Approach

Identify

Identify the data processing activities, technical and organizational measures to protect data (e.g. interviews with the stakeholders within the organization)

Visualize



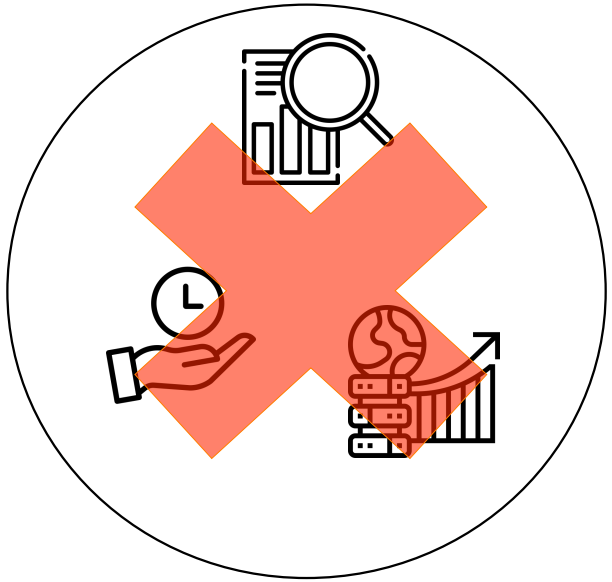
Optionally, visualize these identified processing activities to understand how the personal data moves across the organization

Store

Create an Excel registry to store all the information that has to be included in a record of processing activity. Fill the registry row-by-row.

Motivation

Problem



Error-prone and time consuming process.

Hard to maintain the registry.

Not feasible for large enterprises.

Research Goal:

The goal of this research is to propose an automated process to create and maintain records of processing activities by using Enterprise Architecture models

Outline

1. Motivation
   - GDPR
   - Record of Processing Activities
   - The state-of-the-art Approach
   - Problem
2. Approach
   - Enterprise Architecture
   - Using EA Models as an Information Source
   - Analyzing EA Models
3. Evaluation
4. Conclusion

Enterprise Architecture

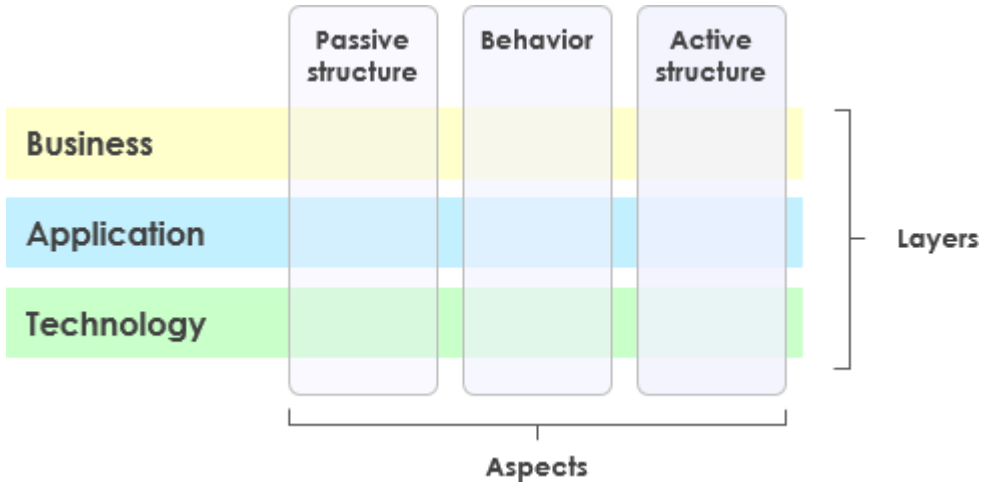
A model representing the elements of an enterprise (business, organization, application, information, infrastructure, data)

The current state of the enterprise can be analyzed and documented to to fulfill the obligations that the GDPR sets upon the enterprises.

ArchiMate®

Modeling Language

Layers and Aspects to describe EA



Archi
archimate modelling

ArciMate Modeling Toolkit

Supports

ArchiMate®

4. The Open Group. ArchiMate 3.0.1 Specification, 2017.

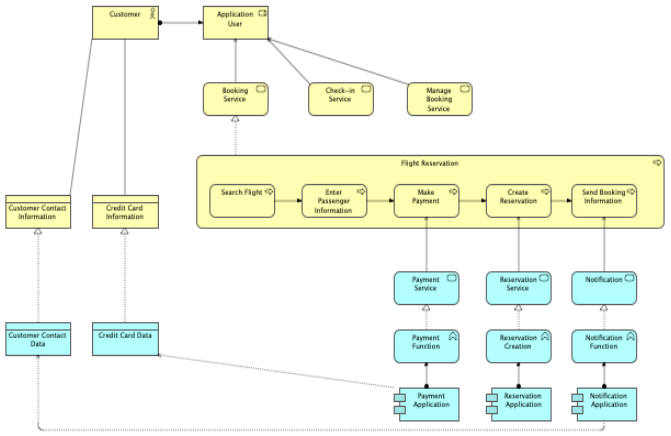
5. Archimatetool. https://www.archimatetool.com/. accessed: 26.03.2019

Approach

Using EA Models as an Information Source

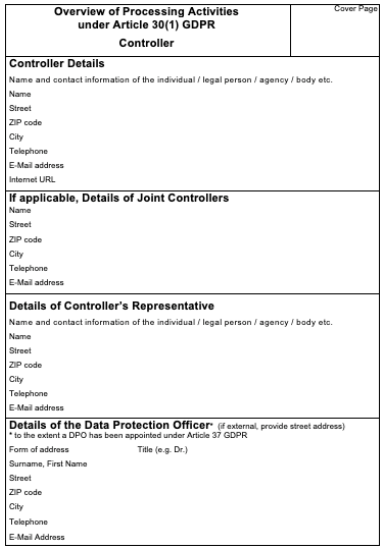
Architecture and architecture models as the major source of information to understand data processing activities within an organization

- ➢ How personal data is used
- ➢ Why it was collected
- ➢ How it is processed
- ➢ Where it is stored



EA Models

- Model record of processing activities as a requirement
- Customize EA elements and create relationships between requirements and elements from other layers
- Use machine-readable format of the EA models



Overview of Processing Activities under Article 30(1) GDPR
Controller
Cover Page

Controller Details
Name and contact information of the individual / legal person / agency / body etc.
Name
Street
ZIP code
City
Telephone
E-Mail address
Internet URL

If applicable, Details of Joint Controllers
Name
Street
ZIP code
City
Telephone
E-Mail address

Details of Controller's Representative
Name and contact information of the individual / legal person / agency / body etc.
Name
Street
ZIP code
City
Telephone
E-Mail address

Details of the Data Protection Officer* (if external, provide street address)
* to the extent a DPO has been appointed under Article 37 GDPR
Form of address Title (e.g. Dr.)
Surname, First Name
Street
ZIP code
City
Telephone
E-Mail Address

Record of Processing Activities

Approach

Information About Data Processing Activity

EXAMPLE

Approach

Parsing the XML format

1) Find the requirements where “isRequiredByRPA” is true from the motivation folder

```
<element xsi:type="archimate:Requirement" name="Get Name of Processing Activity" id="
c3925484-d56e-4dea-bd63-c4be62228c93">
  <property key="isRequiredByRPA" value="true"/>
</element>
<element xsi:type="archimate:Requirement" name="Get Department of Owner of Processing Activity" id="
34825a1c-d826-4401-aa0d-249c1f5f089d">
  <property key="isRequiredByRPA" value="true"/>
</element>
```

2) Find “RealizationRelationships” from the relations folder and filter the relationships by “target” value

```
<element xsi:type="archimate:RealizationRelationship" id="0c2ebc26-a15d-44f4-a377-e32bbc2ecc11" source="
7ba47558-0f42-4806-8bed-e7191930ae29" target="c3925484-d56e-4dea-bd63-c4be62228c93"/>
```

3) Use “source” value to find the EA element from which the RealizationRelationship comes

```
<element xsi:type="archimate:BusinessProcess" name="Order Placement" id="7ba47558-0f42-4806-8bed-e7191930ae29">
  <property key="Processing Activity" value="Order Placement"/>
  <property key="Description" value="Customers can give order"/>
  <property key="Lawful Basis" value="Contract"/>
  <property key="Purpose" value="Customer Service"/>
</element>
```

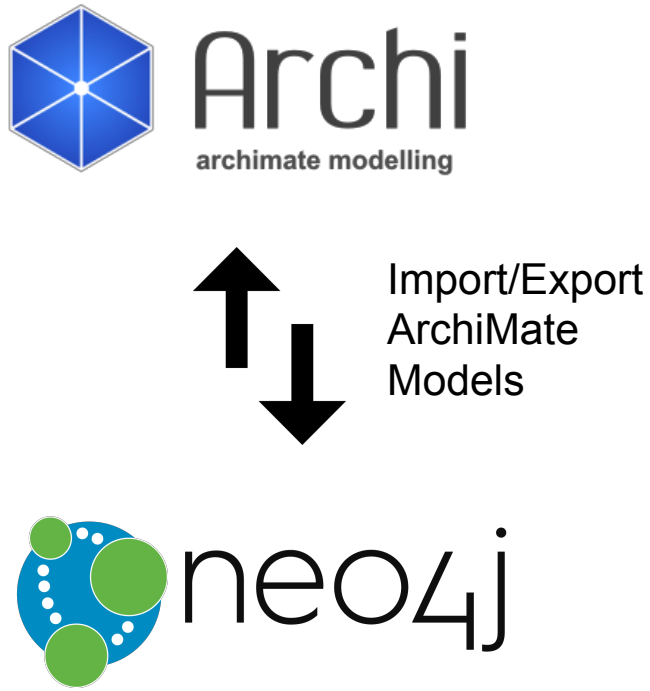
Approach

Analyzing EA Models

EA models can be converted into simplified visualizations which can be easier to understand.

For further analyses, we should be able to run queries over the data represented in the models similar to what we do in relational or document-based databases.

Graph databases are quite promising for wide range of analysis.



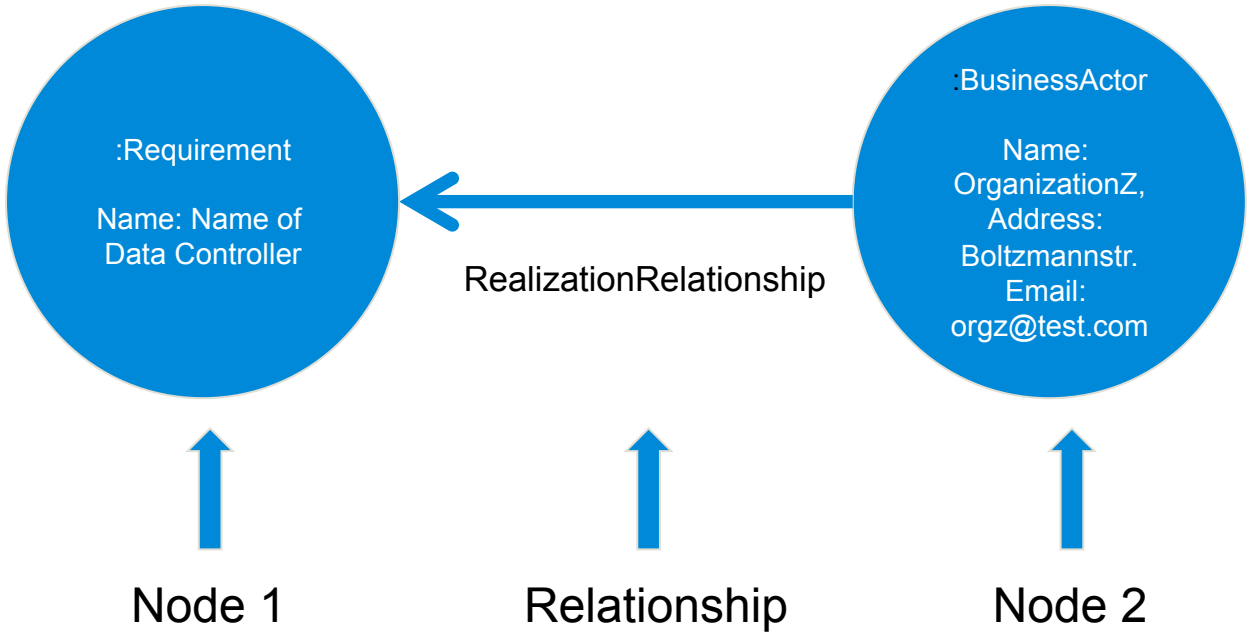
5. Archimatetool. https://www.archimatetool.com/. accessed: 26.03.2019

6. Neo4j. https://neo4j.com/style-guide/. accessed: 26.03.2019

Approach

Analyzing EA Models

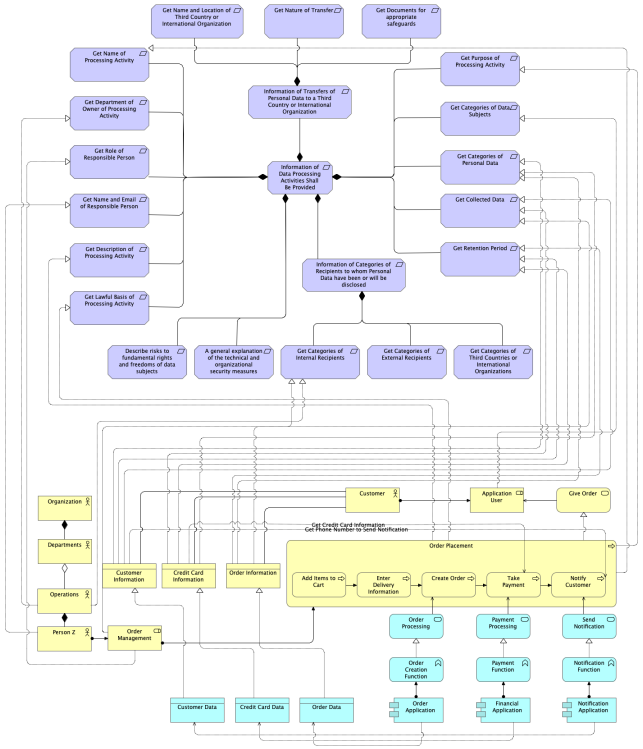
Graph Data Model

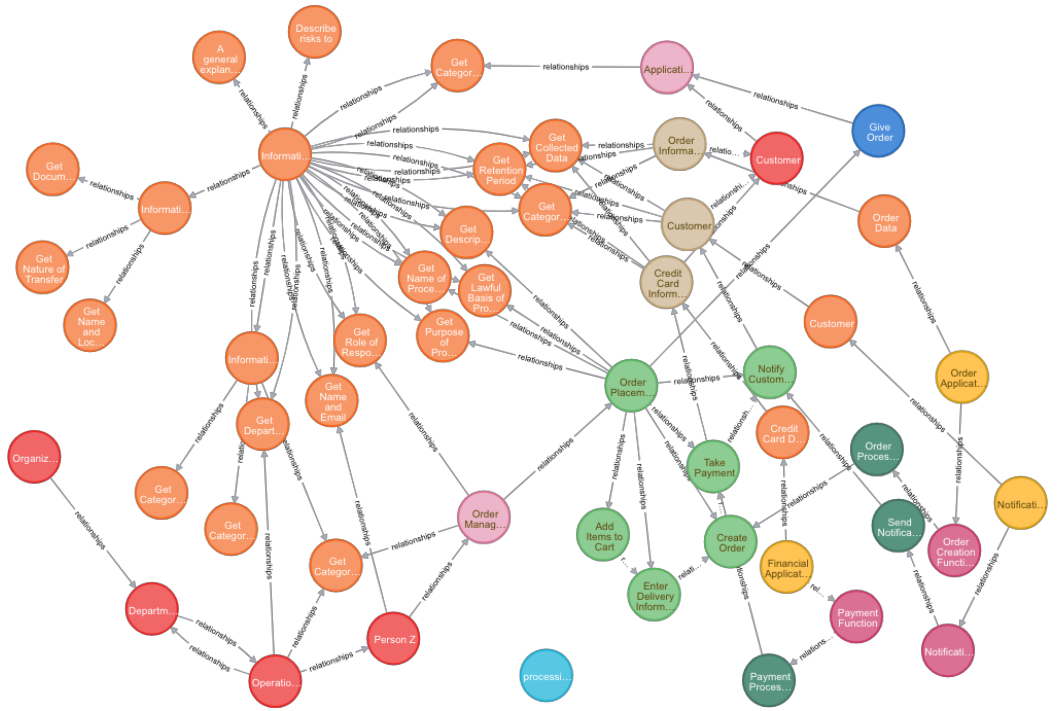


- Nodes represent entities or other domain components.
- Relationships are connections between nodes and they can be used to find related nodes.
- Both nodes and relationships include properties and labels.
- Properties : key-value pairs
- Labels: Use to group nodes into sets (more targeted queries)

Approach

Analyzing EA Models





Transfer EA models into a graph database by running queries on the graph database.

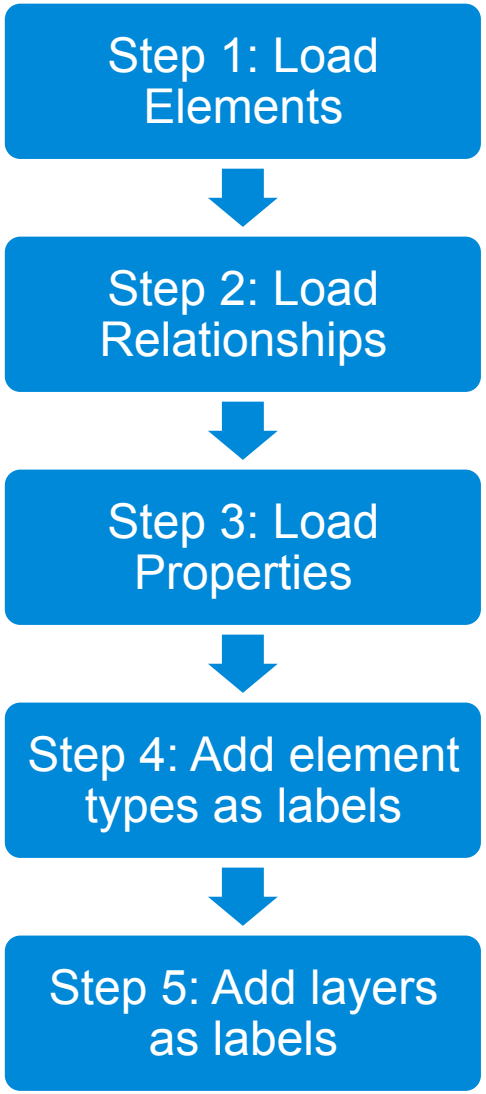
It is possible to export EA models in CSV format. We can load data from CSV file in Cypher (provided by Neo4j) query language.

Customize nodes and relationships by adding labels to represent layers from the ArchiMate full framework.

Labeling can help us do extensive analyses.

Approach

Analyzing EA Models



Load Elements

```
LOAD CSV WITH HEADERS FROM 'file:///elements.csv' AS line
CREATE (:elements {class:line.Type, name:line.Name,
documentation:line.Documentation,
    id:line.ID })
```

Load Relationships

```
LOAD CSV WITH HEADERS FROM 'file:///relations.csv' AS line
MATCH (n {id:line.Source})
WITH n, line
MATCH (m {id:line.Target})
WITH n, m, line
CREATE (n)-[:relationships {id:line.ID, class:line.Type,
documentation:line.Documentation, name:line.Name}]->(m)
```

Add Labels

```
MATCH (n:elements)
CALL apoc.create.addLabels(n,
[ n.class ]) YIELD node
RETURN node
```

Approach

Analyzing EA Models

Query graph data to retrieve the elements that realize the requirements

```
MATCH (r:Requirement)<-[:relationships *1..1{ class:
"RealizationRelationship" }]-(n:BusinessLayer)
WITH n,r
WITH apoc.map.removeKeys(n,
["name","id","documentation","class"]) as info, r
RETURN info,r.name
```

"info"	"r.name"
{"Department of Owner of Processing Activity":"Operations","Category":"Operations"}	"Get Categories of Internal Recipients"
{"Department of Owner of Processing Activity":"Operations","Category":"Operations"}	"Get Department of Owner of Processing Activity"
{"Email":"personz@organization.com","Name":"Person Z"}	"Get Name and Email of Responsible Person"
{"Retention Period":"365","Category":"Credit Card","Collected Data":"Card Holder, Credit Card Number, CVV, Expire Date"}	"Get Collected Data"
{"Retention Period":"365","Category":"Credit Card","Collected Data":"Card Holder, Credit Card Number, CVV, Expire Date"}	"Get Retention Period"
{"Retention Period":"365","Category":"Credit Card","Collected Data":"Card Holder, Credit Card Number, CVV, Expire Date"}	"Get Categories of Personal Data"
{"Retention Period":"365","Category":"PII","Collected Data":"Name, Email, Home Address, Phone"}	"Get Categories of Personal Data"
{"Retention Period":"365","Category":"PII","Collected Data":"Name, Email, Home Address, Phone"}	"Get Retention Period"
{"Retention Period":"365","Category":"PII","Collected Data":"Name, Email, Home Address, Phone"}	"Get Collected Data"
{"Retention Period":"365","Category":"Order","Collected Data":"Items"}	"Get Collected Data"
{"Retention Period":"365","Category":"Order","Collected Data":"Items"}	"Get Retention Period"
{"Retention Period":"365","Category":"Order","Collected Data":"Items"}	"Get Categories of Personal Data"
{"Processing Activity":"Order Placement","Purpose":"Customer Service","Description":"Customers can give order","Lawful Basis":"Contract"}	"Get Lawful Basis of Processing Activity"
{"Processing Activity":"Order Placement","Purpose":"Customer Service","Description":"Customers can give order","Lawful Basis":"Contract"}	"Get Name of Processing Activity"
{"Processing Activity":"Order Placement","Purpose":"Customer Service","Description":"Customers can give order","Lawful Basis":"Contract"}	"Get Description of Processing Activity"
{"Processing Activity":"Order Placement","Purpose":"Customer Service","Description":"Customers can give order","Lawful Basis":"Contract"}	"Get Purpose of Processing Activity"
{"Category":"Customer"}	"Get Categories of Data Subjects"
{"InternalRecipients":"Organization"}	"Get Categories of Internal Recipients"
{"InternalRecipients":"Organization"}	"Get Role of Responsible Person"

Outline

1. Motivation
   - GDPR
   - Record of Processing Activities
   - The state-of-the-art Approach
   - Problem
2. Approach
   - Enterprise Architecture
   - Using EA Models as Information Source
   - Analyzing EA Models
3. Evaluation
4. Conclusion

Evaluation

Direct e-mails, posting on LinkedIn Groups, Meetup

- Conducted 3 interviews with data privacy experts from different organizations

- Meetup at the main campus:

  Organized with the help of Datenschutz München

  4 data privacy experts attended

13 MAR

Past Meetup

Themenvortrag: Automatisches VVV/VVT durch Enterprise Architecture Mgmt

Evaluation

Questions about the state-of-the-art approach and problems	
How is a record of processing activities created?	Excel spreadsheets
How often is it updated?	Best case scenario: When it is necessary (e.g. new processing activity) Reality: Once a year or before important situations (e.g. compliance audits)
What kind of impediments are faced in the process of creating this document?	Too many manual processes (e.g. Data collection, keeping up-to-date) Hard to verify legitimacy of processing activities
How do you define a valid record of processing activities?	Hard to achieve 100% completeness To mitigate risk, document every single detail

Evaluation

Questions about the concept	
Is the proposed approach feasible for organizations?	Valuable if EA models exist Only few processes are modeled What happens if the organization uses different ways to document EA
What are the challenges and disadvantages in this approach?	EA models are must-have Creating relationships between requirements and other EA elements Creating EA models from scratch if they don't exist Requires knowledge about modeling language
What else can be done from your perspectives?	More information sources can be integrated Reduce effort to create models Create a bidirectional link between EA models and RPA A concrete implementation of the concept
What should be the functionalities of a tool that automatically creates a record of processing activities?	One platform to manage privacy-related issues Different information sources should be supported by the platform One click to generate RPA Retrieve data without knowing any special query or modeling language Automatically update RPA when it is necessary Check legitimacy of processing activities

Outline

1. Motivation
   - GDPR
   - Record of Processing Activities
   - The state-of-the-art Approach
   - Problem
2. Approach
   - Enterprise Architecture
   - Using EA Models as Information Source
   - Analyzing EA Models
3. Evaluation
4. Conclusion

Conclusion

Future Work

- Study needs to be conducted to address the usability and the acceptance of the proposed EA models and the feasibility of the approach
- How the communication should be conducted between data privacy experts and Enterprise Architects.
- Further investigation about the needs and concerns of data privacy experts and how EA can support them.

Using EA models is a promising solution to maintain records of processing activities;

- Understand how data is collected and processed
- Do extensive analyses over data
- Create automated processes

Ahmet Tanakol

B.Sc.

Technische Universität München
Faculty of Informatics
Chair of Software Engineering for Business Information Systems

Boltzmannstraße 3
85748 Garching bei München

ahmet.tanakol@tum.de
wwwmatthes.in.tum.de

References

1. European Commission. REGULATION (EU) 2016/679 OF THE EUROPEAN PAR-LIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natura lpersons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). L119:1–88, 2016.

2. ICO. Guide to the General Data Protection Regulation (GDPR).https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/.accessed:2018-11-12.

3. Daniel Felz. German DPAs Publish Model GDPR Processing Records –Translations Provided.https://www.alstonprivacy.com/german-dpas-publish-model-gdpr-processing-records-translations-provided/, 2018. accessed: 2019-02-04.

4. The Open Group. ArchiMate 3.0.1 Specification, 2017.

5. Archimatetool. https://www.archimatetool.com/. accessed: 26.03.2019

6. Neo4j. https://neo4j.com/style-guide/. accessed: 26.03.2019