Ontology Based PEP Identification

Andrejs Gaidukovs[1], Pavels Grjozs [1], Girts Kebers[2]

[1] Institute of Applied Computer Systems, Riga Technical University, Latvia
{andrejs.gaidukovs; pavels.grjozs}@edu.rtu.lv;
[2] SIA "Lursoft IT", Latvia
{kebers}@lursoft.lv

Abstract. In many countries financial organizations have obligations to identify politically exposed persons. The aim of the research work presented in this paper is to develop the solution that gives an opportunity to identify politically exposed persons using information from the web and several state registers. In particular, the paper focuses on one of the possible solutions that is based on the ontology with the dedicated inference rules. This paper proposes the ontology for identification of politically exposed persons and discusses the problems faced during the development an application of the ontology, showing the limitations of the ontology based approach and pointing to some potential solutions of the identified problems.

Keywords: ontology, politically exposed person, PEP, identification, compliance, anti-money laundering, countering the financing of terrorism

1 Introduction

Politically exposed person (PEP) is defined as an individual who is or has been entrusted with a prominent public function. Due to their position and influence, it is recognized that many PEPs are in positions that can be abused for the purpose of committing money laundering offences and related predicate offences, including corruption, as well as conducting activity related to terrorist financing [1].

This is a research-in-progress paper to report on the preliminary results of the ongoing research regarding the use of the ontology based approach in PEP identification. The project is held in Latvia, thus, there are some country specific regulations incorporated in the solution that can be different in other countries.

The following are three main sources that regulate and describe the rules for PEP identification:

- Latvian Anti-Money Laundering / Countering the Financing of Terrorism (AML/CFT) law [1];
- recommendations [2] issued by Financial and Capital Market Commission (FCMC) that regulates financial organizations' operations and compliance in Latvia, among other, on how to identify politically exposed people;

- Financial Action Task Force (FATF) [3] – recommendations of international governmental organization that stipulates 40+9 policies on AML/CFT .

Politically exposed persons are divided into three main categories:

1. A PEP itself – individual who is or has been entrusted with a prominent public function [3];
2. Family members of a PEP;
3. Persons who are closely related to a PEP – the person that has co-owned private companies with the PEP or has business relations with the PEP.

There are several differences between definitions and rules if various sources of regulations are compared. For instance, time period when the person is considered as politically exposed even after he/she has left PEP indicative position from FCMC definition is 12 months. If the person was PEP then person stays PEP forever by FATF definition. For the purposes of this paper, the FCMC definition of PEP status expiry is used i.e.– 12 months. Another example of unclear definition is business relation definition. Not all deals have to be taken into consideration, as mentioned in all definitions, but there is no clear threshold of the relevant deal size or any other parameters stated in the above-mentioned regulations.

The following research method was used in the study: 1) possible data sources relevant for PEP status identification were analyzed and types of data were defined and structured; 2) related work on PEP identification was studied; 3) preliminary PEP status identification ontologies were built and tested; 4) after the analysis of the test results the final ontology was built; 5) the graph database was used to solve PEP identification challenges that were not met by the ontology.

The paper is organized as follows. In Section 2, the current situation in PEP identification solutions is described. Section 3 proposes and discusses ontology based solution for PEP identification. Section 4 concludes the paper.

2 State-of-Art in PEP Identification

Identification of PEPs is a significant issue in the field of finance and national security. CaseWare Analytics [4] describes a solution for the global identification of PEPs. CaseWare Analytics study offers the division of PEP status into four classes depending on the probability (risk), as well as provides an insight into the development of a PEP identification system. Possible solutions for the issue of PEP identification and associated data retrieval risks are provided in several patents: Mark. A. Schiffer's patent "Method of Ranking Politically Exposed Persons and Other Heightened Risk Persons and Entities" [5] , David Lawrence's patent "Automated Political Risk Management" [6] , patent "Evaluating Customer Risk" [7]. All these methods use different information acquisition methods from several data sources. To join these data sources, bodies of related notions are used, for example, controlled dictionaries, concept maps, and ontologies.

Ontologies are used in Anti-Money Laundering systems. One of examples for the use of ontologies is an article by Dr. Jerry A. Smith [8] showing a money laundering

scheme from a perspective of an ontology. The scheme shows concepts understandable to both a person and interconnected system that have to be able to operate in this domain of knowledge.

There are number of software products available on the market for financial institutions and insurance companies for maintaining compliance with anti-money laundering requirements, for example: Actico Name Matching Customer [9], FICO TONBELLER IT-based Anti-Money Laundering [10], IBM AML compliance solution [11]. All these solutions use external monitoring lists for comparing with the customer lists. This process includes comparing of similar names, aliases, and spelling variations of names, as well as birthdates, nationality and domicile of the persons concerned against monitoring lists. Official lists such as the EU list or OFAC list can be used for monitoring, as well as those offered by commercial services. The purpose of the solution presented in this paper is to provide a commercial service that contains actual list of PEPs.

3 Proposed solution

Proposed solution analyzes data that is available on the web and in state registers. First, the available data sources were analyzed and Resource register was developed for these data sources. Then the data sources in Resource register were rated by data reliability and credibility. Further, Resource register was used for matching the attributes of the data objects available in various data sources. Actual data was retrieved from data sources using crawlers. The ontology was used for person PEP status identification. The proposed ontology (**Fig. 1**) is created using Protégé [12] tool. The proposed ontology has three major classes “Person”, “Position” and “Organization”.

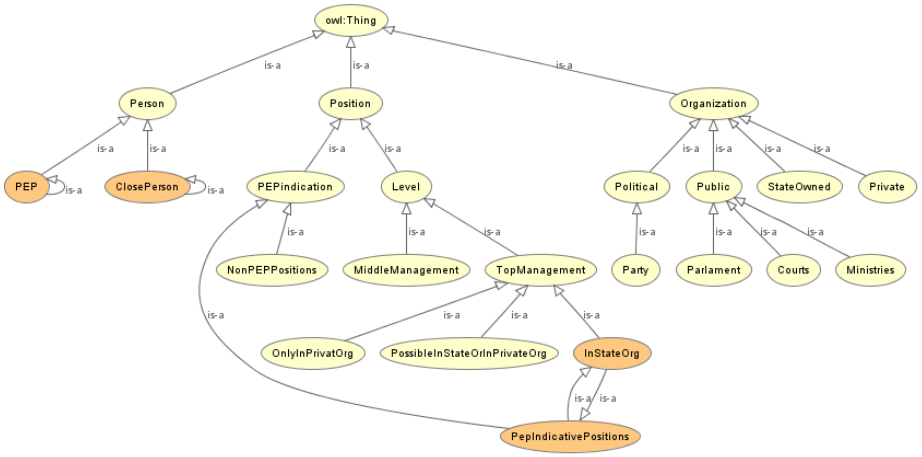

Fig. 1. Proposed PEP identification ontology in Protege

Class “Person” has two subclasses: “PEP” and “ClosePerson”. Two subclasses are needed to have a possibility to stop the reasoning. When PEP is identified, the list of

close persons is defined. (**Fig. 2**) If only one class "PEP" was used then the status PEP would be assigned to each person from ClosePeople list in step 3 (`Step 3: For i=1 to M do ClosePeople[i]=PEP`). Then assigned people would be searched for persons close to PEP and identified persons would again be granted PEP status. Such algorithm would infer that all people are PEPs. If class "ClosePerson" is used then ClosePerson status is assigned to each person from the list in step 3 (**Fig. 2**) and inference stops there. Another option would be using three subclasses as PEP categories, but that would require introducing additional class "FamilyMember". This class would then have the same properties and behaviors as class "ClosePerson". Identification of both subclass instances is the same (and, thus, the information processing for financial institution is also the same). Therefore, there is no need to have three subclasses and it is sufficient to have two subclasses to identify the PEPs.

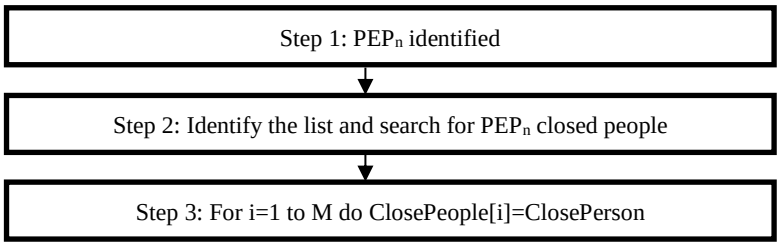

Fig. 2. PEP identification steps

Class "Position" is divided into two subclasses: PEPindication and Level. "PEPindication" describes possible positions: PEP and nonPEP positions. "Level" describes management level and consists of "Middle management" and "Top management". The research focuses on top management positions because PEPs are from top management level. Top management level has three subclasses:

1. OnlyInPrivateOrg – the class that contains positions that refer only to private organizations. For instance, position "owner" is only possible in private organization as the State organisations would not have any private individuals as owners (i.e. state organisations are owned by state).
2. InStateOrg – the class that contains positions referring to public organizations. For instance, minister, deputy, judge – these positions can be held by persons that work only in public sector.
3. PossibleInStateOrInPrivateOrg – a person with such position can work either in private sector or in public sector. For instance, board member or director. State owned organizations have management boards and board members, as well as state agencies hire directors.

The class "Organization" has to identify the type of organization and four subclasses are possible: Political, StateOwned, Public or Private. **Table 1** shows possible combinations when person is politically exposed based on the position held and organization type he/she works for. ("Yes" in the table means that the person is politically exposed.)

Table 1. PEP positions matrix

	Position available only in public sector	Position available only in private sector	Position available in private and in public sectors
Private organization	Not available	No	No
Public organization	Yes	Not available	Yes

PEP identification ontology's object properties are described in **Table 2**. Object property "hasRelation" has tree sub-properties: "isBusinessCoowner", "isBusinessPartners", "isFamilyMember". Sub-properties have the same domain and range as parent property.

Table 2. PEP identification ontology's object properties

Object properties	**Domain**	**Range**
hasJob	Person	Position
hasRelation	Person	Person
worksFor	Person	Organization

The following rules were incorporated into the Ontology:

1. Class "InStateOrg" is equivalent to class "PEPIndicativePositions":

$$InStateOrg \equiv PEPIndicativePositions$$

2. Person who has job from PEP Indicative positions list is politically exposed:

$$\forall x\ (hasJob(x, PEPIndicativePositions)) \rightarrow PEP(x)$$

3. Person who has position available in private or in public sectors and works for state-owned company is politically exposed:

$$\forall x\ (hasJob(x, PossibleInStateOrInPrivateOrg) \wedge worksFor(x, StateOwned)) \rightarrow PEP(x)$$

4. Every person who is a family member or has a co-owned business or has a business relationships with the PEP is close to PEP person:

$$\forall x,y\ ((Familymember(y,x) \vee BusinessCoowner(y,x) \vee BusinessPartner(y,x)) \wedge PEP(x)) \rightarrow ClosePerson(y)$$

Created rules and classes were tested successfully. Ontology is consistent and reasoning works correctly.

In further development the challenge to incorporate the period when a person is considered as a PEP after the person has left the PEP indicative position was addressed (in this study this period is 12 months). It was needed to add to object property time period as attribute and later to define rules that use this attribute in inference process. The only discovered possibility to implement this was to use additional class "Interval" and use instances of this class in reasoning. Such solution would mean that for every person–position relation class "Interval" instance must be defined. The solution would have as many Interval instances as the number of aforementioned relations.

As described earlier, another challenge is that PEP identification regulations have no clear guidance on possible deal parameters and their threshold value. During the project there were defined some attributes based on deal value and available related information in Latian law. While in experiments different limits were tested, the following two potential limits were investigated for the practical usage:

1. Limits defined in Latvian AML/CFT law [1] – Euro 15 000. If a financial organization notices money transaction with sum equal or larger than Euro 15000, the financial institution has an obligation to pay higher attention to and control customers who have made this money transaction;
2. Limit defined in Latvian public procurement law [13] – Eur 10000. This limit is starting sum when the state organization has to issue public procurement procedure.

In order for the proposed ontology to evaluate the deal criteria for the identification of persons closely related to a PEP based on their business transactions, the proposed ontology needs to be extended to incorporate additional object property "isBusinessPartners" and attribute "DealValue". Possibility to add the attribute to the object property and using of this attribute in the reasoning process was not found.

Another problem faced during development of ontology was of ontology was as follows. **Fig. 3** shows a screenshot from Protégé tool [12] with person's object properties. If the person has one position and one employer the inference works correctly. However, if the person has two positions in two types of organizations, the inference result is not satisfactory. In the example in **Fig. 3**, the person works in L-energo (state owned company) as a project manager and holds board member position in ABC Ltd (private company). In real life the person is not a PEP, but Protégé reasons that this person is the PEP because the person works in the state-owned company on a top management level. Actually all instance properties are defined as independent rules. But in this study there is the need to combine two rules into one.

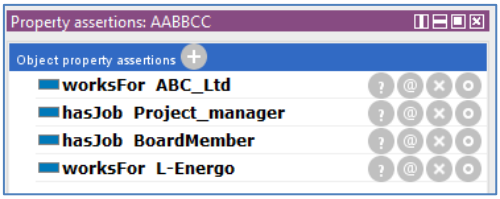

Fig. 3. Person with different positions

After investigation, all three above-mentioned problems were identified as N-ary relation problems. Possible solutions of this problem are described in [14]. All solutions offer to add a class and define the range as instances. However, in this study there is no possibility to define the range of time intervals or moments for every query that is processed. Possible solution for property chains is described in [15], but we did not succeed to implement it in Protégé. Other ontology editing tools were investigated in order to satisfy the needs of current ontology. Only solution that handles n-ary relations is x-Protégé [16]. This is add-on to Protégé tool. But after many experiments and due to lack of x-Protégé tool manual there still was no result.

Experiments were switched from ontology to graph database. The graph database tool OrientDB [17] was used. OrientDB tool allows definition of attributes for relationships and there are no restrictions on the number of attributes for relationships. The only limitation is that only one attribute can be visualized for a relationship. Defined attributes can be used in querying. **Fig. 4** shows realization of "hasDeal" relation were deal value of 20 000 Euro is defined. The time attribute is added to the relation "hasJob" that depicts the time moment when a person starts working as CEO. Also the chain of relations hasJob->worksFor can be defined in OrientDB. Thus, the combination of ontology approach with graph database allows solving all three above mentioned challenges.

Fig. 4. Relations with attributes in OrientDB tool

4 Conclusion

This paper has been developed with an aim to evaluate the possibility of using ontology approach for the identification of politically exposed persons (PEPs), their family members and close associates.

To ensure the compliance of the offered solution with requirements set forth in laws and regulations of the Republic of Latvia; laws and regulations defining the PEP status have been analyzed within the framework of this paper, thus defining the amount of search tasks and necessary data.

The studies in the field of identification of PEP status have been analyzed in this paper as well. The topic of PEPs is shortly covered in scientific articles; however, several commercial products were analyzed and patent information was studied.

The ontology for PEP identification has been developed and tested. Based on the position held by a person, the ontology applies the laws of logical reasoning to conclude whether the person is to be assigned with a PEP status or not. Likewise, the ontology infers whether a person is closely related to a PEP in cases when this person has close relations with the relevant PEP.

During ontology development there were number of limitations identified for this method. It was not possible to add an attribute to the relation. Possible solutions described in literature were not suitable for PEP status identification. After many experiments with the ontology, as an alternative, the graph database method was tried and gave satisfactory results. Graph database OrientDB tool supports attribute attachment to a relation. Also this tool supports relation chains. The graph database method is suitable to be combined with the developed ontology.

The ontology shows good results when implementing reasoning. When defining equivalent classes and object properties, the inference works correctly using these defined languages. The research work on combining the ontology with other methods such as information fusion and agent technologies is continuing for finding the most effective solution for PEP identification not only at the national, but also at the international levels.

Acknowledgement. The research has been supported in part by the funding from the research project "Competence Centre of Information and Communication Technologies" of EU Structural funds, contract No. 1.2.1.1/16/A/007 signed between IT Competence Centre and Central Finance and Contracting Agency, project No. 1.14.

References

1. Latvian Anti-Money Laundering / Countering the Financing of Terrorism (AML/CFT) law, https://likumi.lv/doc.php?id=178987.
2. FKTK, http://www.fktk.lv/.
3. FATF: FATF Guidance. (2013).
4. Webinar: Reduce Time Spent Screening Politically Exposed Persons | CaseWare Analytics, https://www.casewareanalytics.com/events/webinar-reduce-time-spent-screening-politically-exposed-persons.
5. Schiffer, M.: Method of ranking politically exposed persons and other heightened risk persons and entities. (2008).
6. Lawrence, D.: Automated political risk management. (2002).
7. Evaluating customer risk. (2006).
8. Deep Learning Intelligence Platform – Addressing the KYC AML Counter Terrorism Financing Challenge – Data Scientist Insights, https://datascientistinsights.com/2014/12/17/deep-learning-intelligence-platform-addressing-the-kyc-aml-terrorism-financing-challenge/.
9. Smarter Decisions with ACTICO, https://www.actico.com/en.
10. FICO TONBELLER - Anti-Money Laundering | Fraud Prevention | FATCA Compliance | Know Your Customer, http://www.tonbeller.com/en/.
11. IBM AML compliance solution, https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=IIS03009USEN.
12. Protégé, http://protege.stanford.edu/.
13. Public procurement law, https://likumi.lv/doc.php?id=133536.
14. Defining N-ary Relations on the Semantic Web, https://www.w3.org/TR/swbp-n-aryRelations/.
15. Hitzler, P., Krötzsch, M., Rudolph, S.: Foundations of Semantic Web technologies. CRC Press (2010).
16. Willms, C., Krieger, H.-U., Kiefer, B.: ×-Protégé An Ontology Editor for Defining Cartesian Types to Represent n-ary Relations.
17. OrientDB - Distributed Multi-Model and Graph Database, http://orientdb.com/orientdb/.