MASTER'S THESIS

Enabling Business Rule Traceability as Guidance to Regulation Origination in Finance

Berghuis, B.

Award date:
2022

Enabling Business Rule Traceability as Guidance to Regulation Origination in Finance

Master:	Open Universiteit, faculteit Bètawetenschappen
	Masteropleiding Business Process Management & IT
Programme:	Open University of the Netherlands, faculty of science
	Master of Science Business Process Management & IT
Course:	IM0602 Voorbereiden Afstuderen BPMIT IM9806 Afstudeeropdracht Business Process Management and IT
Student:	Brent
Berghuis Identification number:	
Date:	16-06-2022
Thesis Supervisor	Lloyd Rutledge
Second Reader	Ella Roubtsova
Third Reviewer	<If Applicable>
Version number:	3.0
Status:	Final draft

Table of Contents

Abstract

We evaluate and assess the applicability of a Reference Architecture for Traceability (RA4T) to Finance regulations, and assess if it can assist with traceable implementation of regulations. This enables tracing back business rules to their source in regulations. This evaluation is conducted in a thesis circle in which other members assess the architecture in other domains. This reference architecture aims to guide developers in making rule-based systems which provide traceability. Traceability ensures that end users in their interface can trace back the full origin of system conclusions to their policy documents.

The reference architecture applies a Controlled Natural Language (CNL) to rule-development to ensure consistent interaction between humans, processable by computers. This research addresses two areas: First, we will evaluate the effects on traceability by applying RA4T to Finance Regulations and see whether traceability changes for end users. The context is European Anti Money Laundering Directives, as they stipulate traceable implementation. Then, we will structure the scenarios in CNL and transform to code. Logical conclusions on rules are tied to data to accommodate basic explanations by providing snippets of the source regulation. The challenge is keeping intact source-related meta-data needed for traceability. This contribution is sufficient when the Architecture is evaluated via scenarios, and constitutes an improvement in the domain context.

Keywords: Business rule traceability, CNL, Reference Architecture for Traceability, Design Science, Anti Money-Laundering, Virtual Currencies.

Summary

In this thesis, we assessed and evaluated a reference architecture developed to increase traceability in rules-based systems. The concept of traceability implies the ability to trace back individual components of rules to their origin in the source documentation. This architecture, dubbed as Reference Architecture for Traceability (RA4T), is set up as facilitator for tracing back business rules to their origination in source documentation or policy. It is made up of three distinctive stages: Development, Automation and End User. These stages are linked together via one End to End chain (E2E) which is comprised of four traceability links for implementation. If the chain of traceability stays intact, and all four links have been implemented, this leads to traceability across the stages. A novel approach is to include Controlled Natural Languages (CNL) for rule-development to ensure consistent interaction between humans and machines. This thesis is written with the Semantics for Business Rules and Vocabulary (SBVR) in mind, but recent amendments made the RA4T CNL agnostic.

Within a thesis circle the RA4T is evaluated and assessed on its applicability to several case domains. We did so via a scenario-based evaluation strategy within Design Science Research (DSR). This thesis research was conducted in the Finance domain, more specifically the Anti-Money Laundering Directives (AMLD) brought forward by the European Union. The case regulations are AMLD, ancillary documentation and official guidance.

In our research we wanted to find out how, and to what extent the RA4T facilitated the development of business information systems (BIS) for the Finance domain, and if, after careful evaluation, any improvements to the architecture can be proposed.

Therefore, we had to discuss more in depth the relations between business-rules, regulations and AMLD. But also the workings of a CNL, exclusively SBVR, and tools that assist with setting up the ontology in Web Ontology Language (OWL) like Protégé.

We set up an experiment, in which we, via scenario-based evaluation, assessed the effects of the RA4T on traceability during BIS development. We approached this with the AMLD as object of study in mind, which we treated with the RA4T. This was the validation model to assess effects on traceability. We then formulated a strategy to find the modality within the regulations and to structure them into SBVR. We modeled the input data to resemble a type of transaction used regularly by banks, called the MT103. The scenario rules were inputted into the ontology with Protégé and then validated via data covering parts of the scenario.

The scenarios were chosen on their ability to demonstrate and cover the breadth of AMLD's, and displayed variations in end users, transactional object – and type. We effected traceability from the end-user interface all the way up to the source in AMLD for all scenarios, according to the RA4T. We could cycle through the traceability links within the chain. Cycling through revealed much of the process of thought that is behind these rules. We identified a potential extension to the architecture to include date / time functions. These can allow for automated regulation validity checking.

Summarizing the results from the scenarios, we concluded that RA4T facilitates the development of BIS with rule traceability by giving clear indications on where traceability should be implemented. It proposes methods to do so and serves as lay-over to quickly check on stages within the process. Furthermore, by proposing a CNL it ensures a smoother interaction between humans, domain experts and machines alike, and allows for automation opportunities by enabling the use of automated CNL to OWL mappings. We propose to extent the RA4T with a date / time repository to enable improvement exercises on development, and for automated regulation validity checks. Future research could be initiated on the development of a tool that can cover E2E development.

1. Introduction

1.1 Finance and Regulations Area

The financial sector post 2008, widely considered the banking crisis of the 21st century, has experienced a tightening of regulatory constraints. With the societal implications of the failure of an industry, which touches almost all human life on this planet, the fall-out was immense. With a new focus on banking industry regulation, more areas than just strictly financial regulations came into the spotlight. Major scandals involving dirty money, corruption, embezzlement and tax-evasion, considered predicate offenses for money laundering (FATF, 2019); lead to a call for better legislation to combat Money Laundering and enhance cooperation in this field.

The Anti Money-Laundering industry

The sudden growth of the Anti Money-Laundering sector converges with the excessive growth in available financial products and their increased complexity (Rysin & Rysin, 2020). The recent development of crypto currencies led to a paradigm shift in the AML sector. Most current designs of AML legislature have a centralized intermediated financial system as object, thereby enabling the regulation of financial intermediaries in the jurisdictions where they operate territorially speaking (Poskriakov, Chiriaeva & Cavin, 2020). Whilst Virtual Assets and Currencies (VA) like cryptocurrencies remain decentralized and therefore escape the impact of most AML regulation (Poskriakov, Chiriaeva & Cavin, 2020)

It did so in such a way that the European Union had to amend their fifth directive on Anti Money-Laundering and Terrorist Finance (AMLD5) to include recent key developments in the payments and transactional sector as well, resulting in AMLD6 in November of 2018. This directive is better equipped to regulate Virtual Currencies and assets like cryptocurrencies and their custodian wallets. The AMLD6 directive acknowledges the decentralized nature of transactional objects and their custodian wallets and the limited influence a central banking authority can exert over them in their amendment of AMDL5 in Art. 1, 2 (d) 18 of AMLD6.

1.2 From Regulations to Traceable Business Rules

Finance Regulations act as guideline for policy of companies that are regulated by it. These companies then translate this policy into a structured language, adapted to the business

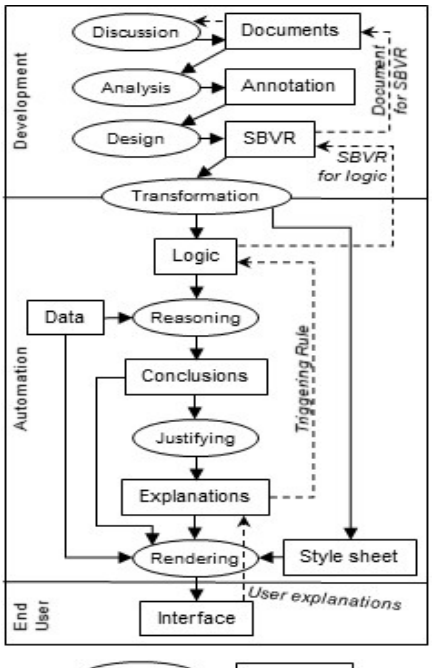

Figure 1 The Reference Architecture for Traceability (Rutledge & Italiaander, 2021)

environment from which implementable business rules for their processes affected by applicable regulations are derived. Transactions executed by banks are an example of such a process. These rules are linked to the source documentation via the annotation process, where domain experts link snippets of source documentation to the resulting rule. In turn, the regulator expects these obliged entities, entities regulated by the regulator, to be able to have oversight on the implementation of the regulations into policy and business rules. Furthermore, they must be able to trace back to policy origins of triggered business rules. However, when a rule is triggered, this only reflects that rule, but not where the rule originates from (Rutledge & Italiaander, 2021). Taking the example of a transaction: we now know it is not allowed, but not to what regulation it is linked. We also do not know whether all these business rules are consistent to each other and whether we can compare them. The ability to trace back to the origins of a business rule in regulations and cycle through the triggered rule(s) increases transparency and business rule traceability, just as the regulator expects. Business rule traceability resembles the ISO 9001:2008 clause 7.5.3. standard for traceability in the sense that all steps require their own unique identification code to be able to distinguish the individual components of the rule; to trace the individual components back to the source regulations. The ISO is oriented towards physical objects, whilst business rules are non-physical. Rutledge and Italiaander, 2021 address the need for traceability with a Reference Architecture for Business Rule Traceability (RA4T), which you can find in figure 1. They propose using a structured format for Business rules in Semantics for Business Rules and Vocabulary (SBVR) which is a controlled natural language, to align business rules with their respective business vocabulary and increase consistency across the domain (Rutledge & Italiaander, 2021)(Njonko et al., 2014). The dotted lines in the reference architecture pinpoint the steps in which traceability should be implemented (Rutledge & Italiaander, 2021). A proposed addition to the architecture is not only the triggering rule, but to also include the triggering data, and implement traceability there (Rutledge et al., 2022). Following these steps increases general business rule traceability, and enables back tracing from the rule programme to the data that triggered the rule, all the way to the SBVR and finally to the regulatory source of the rule. We will evaluate if we can also apply this to the Finance Domain, its effects on traceability, and potential improvements to the initial architecture.

1.3 Increasing traceability within Finance Regulations

For this we need to answer the following questions:

'How and to what extent does the Reference Architecture for Traceability facilitate developing business information systems for the Finance Domain where users can trace back to the full origin of the systems' decisions?'

Here we address the need for an improvement in Traceability.
To come to an answer to the main research question, we must formulate an approach to evaluating of, and, potentially proposing improvements to the Architecture. Captured in the following sub-question.

'Which, if any, improvements to the Reference Architecture for Traceability can be proposed after evaluation?'

1.4 A general approach to increase Traceability within Finance

To explore the applications of business rule traceability, the reference framework developed by (Rutledge & Italiaander, 2021) is used. This reference architecture is then evaluated via Design Science methodology. This reference architecture not only aims to helps end users in their understanding, as they are not necessarily domain experts, but also improves the transparency of

the rule origination. (Rutledge et al., 2021). Applying this RA4T to AMLD requires a conversion to a Controlled Natural Language. This is a commonly seen approach, when it comes to finance regulation as they are often set up in business-rules (Bellomarini et al., 2020). The RA4T uses SBVR to accomplish this. This is not an exclusively compatible CNL to the architecture. Scenarios will be used to evaluate business traceability for this domain. In this scenario-based evaluation, we will investigate whether these scenarios can be captured in a CNL effectively, and if the overall business rule traceability improves by following the reference architecture. The scenarios must be diverse enough for validation of the reference architecture. Whilst the goal of AMLD' are broad, we will focus on variations in users, transactional object, and transaction types. To ensure the angle of virtual currencies and assets is incorporated, they will be included in the transactional objects. These scenarios will be set up from the viewpoint of a Bank as regulated object. This limits the transactional objects, types of transactions and some end user to what a Bank can process. Possible answers to the research question will be explored in the theoretical framework. Predicted effects and possible evaluation methods of artefacts derived from the literature within Design Science Research will be discussed. The specific appliance to the case study to measure the effects on Business Rule Traceability and the evaluation methods will be further discussed in the Methodology chapter.

2. Theoretical Framework

2.1 Introduction to the theoretical framework

This chapter will be divided in two main sections. First, we will describe the search strategies utilized to come to a literature base to which the theoretical framework is related. Secondly, the approach to, and the actual theoretical framework will be described. The second section will also cover some of the nuance's design science brings to research. The first part is made up of strategy, whilst the second part goes more in-depth content wise.

2.2 Literature Search Strategies

There is an extensive body of knowledge covering design science and information systems available. Just starting with an article can be a flawed approach to finding literature relevant for research. By coming up with strategies and using overarching concepts to link articles to one another can give some structure to the literature search section (Booth, 2008).

Snowballing

There is an apparent split between literature, conceptual thinking, and practical application of technology. This notion also serves the approach in which data and information is gathered. By 'snowballing' in literature and keeping track of technological recommendations, we aim to gather as much relevant data as possible. A first candidate is the subject of design science in the CNL environment and more specifically the body of work containing the reference architecture by (Rutledge & Italiaander, 2021). Bibliographies can lead to more relevant documents, and the articles themselves contain references to tools and means.

Berry Picking

Another strategy used is Berry Picking, namely its three sub-parts; backward chaining, subject searching and author searching (Booth, 2008). Backward chaining implies we use the reference architecture as starting point to trace back to references on which the article was based. Before we can use subject searching, we must split our approach in several key subjects. The key subjects are design science, rule traceability, Controlled natural languages and business rules. Additional subjects

will be searched for. Author searching is used to find authorities within their subject.

Academic Communities

Most topics have their own academic communities, with their own journals. An additional area of interest is where these communities intersect (Wakeling et al., 2019).

2.3 The concept of a Theoretical Framework

The theoretical framework functions as the blueprint of a paper. Driven by a possible contribution to the knowledge base (Gregor & Hevner, 2013), the framework tries to make visible the relations between theories.

A Design Science Literature Framework

(Wieringa, 2014) states design science as the design and investigation of artefacts in contexts. The context in this case is the problem context in which the artefacts are set loose to study, and hopefully improve in that context. The artefact in this paper being the Reference Architecture for business rule traceability (Rutledge & Italiaander, 2021) and the context being the business rules set and the traceability thereof. This captures the problem context and the artefact will be assessed on it improving of the traceability. The initial premise of this paper is that this concerns a knowledge question: it should have one answer that is evaluated by the truth (Wieringa, 2014). We are investigating whether we can answer knowledge questions about the design, or artefact, in context. The investigation has a symbiotic relation with the knowledge context. The knowledge context can provide us with existing answers from the literature and if new answers to these questions arise, they are added to the body of work.

The Knowledge context consists among others of scientific theories. The set of scientific theories used in the research are considered the theoretical framework (Wieringa, 2014). The body of knowledge used beforehand is considered the existing- or prior knowledge. The aim of doing design science research is to add something to the existing knowledge context called Posterior knowledge (Wieringa, 2014). This is what is called the contribution to the scientific theory.

Literature Requirements

The theoretical framework is based on scientific theory, composed of two criteria: it must be a theory, describing patterns of a phenomena; or observable occurrence (Wieringa, 2014). This theory must be scientific, for which it must be: peer reviewed, empirically tested, and repeatable (Wieringa, 2014).

2.4 Consequences of a Theoretical Framework

The conceptual framework of a theory serves to frame research problems, describe occurrences, and analyze the structure of these occurrences and state generalizations. The scientific generalization in turn can explain or predict the effects of artifacts in context (Wieringa, 2014). Any theoretical framework should be able to guide you to answers for the Problem, Purpose, Significance and research questions (Grant & Osanloo, 2014).

2.5 Problem Context and Research Question

As DSR is somewhat different from other research paradigms, it is important to ensure that the research question(s) fit the DSR-approach. The three basic type of research questions in DSR are: problem solving, gap spotting and problematization (Thuan et al., 2019). In this paper, problem solving is identified as key driver for the research question. Traceability is considered the problem context in which the RA4T (Rutledge & Italiaander, 2021) tries to contribute. This paper however is oriented towards a scenario-based evaluation to assess said reference architecture's application to the finance domain. (van Aken et al., 2016; Wieringa, 2014) split the approach in design problems and knowledge questions; the first tries to improve a context by (re)designing an artefact the latter calls for knowledge without necessarily improving the context. For this paper the research question is a combination of what Wieringa, 2014 calls an Effect Question and sub questions that relates to the evaluation of an artefact and potential suggestions for improvement (Thuan et al., 2019).

2.6 Architectural Explanations

Explanations to knowledge questions come in a few variations. One of those is the architectural explanation, which states that an effect was produced by the interactions among several architectural components. These interactions are the mechanism that produced the effect (Wieringa, 2014). The RA4T by Rutledge and Italiaander, 2021 is an architectural explanation for the phenomena of traceability .The reference architecture aims to facilitate rule based development and the traceability of the decisions that led to these rules (Rutledge & Italiaander, 2021).

2.7 DSR Evaluation Strategies

A design artefact must be critically evaluated over the following properties: Utility, in what sense does the artifact help improving the context, the quality in which it does so, and how efficient. (Hevner et al., 2004). The literature around DSR evaluation underscores the key importance of evaluating artefacts, in this case the RA4T by (Rutledge & Italiaander, 2021). Whilst the reference architecture itself is being and will be evaluated against the overarching DSR evaluation principles; observational, analytical, experimental, testing and descriptive (Hevner et al., 2004), it is important to limit the scope of this paper. The reference architecture will be assessed on its utility, quality, and efficacy on traceability on several specific scenarios of end user interference. Scenario evaluation falls under the descriptive evaluation methods for DSR (Hevner et al., 2004). Evaluation by way of scenario can serve as a linking pin between abstraction and more tangible situations on which the artefact can be assessed (Venable et al, 2012).

2.8 Controlled Natural Languages and their Formalization

SBVR is a formal controlled natural language for business rule data models (OMG, 2013). The standard introduced by the Object Management Group allows representation of informal specifications in a given natural language to formal logic (Bajwa et al., 2011). This formal representation not only allows for machine processing; it enables the setup of business vocabularies, rules and facts in a specific domain (Bajwa et al., 2011). The domain specifies the exterior context of the rule set. In general, SBVR consists of a business vocabulary tied to a specific domain and sets of business rules (OMG, 2013). The vocabulary describes the specific business domain via two elements: concepts and fact types (Bajwa et al., 2011). Concepts are key representations of a business entity in a domain. One business entity in this context is Virtual Assets in the domain of Finance regulations. These concepts in SBVR are made up of noun concepts, individual concepts and verb concepts (Bajwa et al., 2011). Fact types are a combination of a verb and a noun concept and specify the relations between several concepts in a set of business rules (Bajwa et al., 2011). In the strict sense, SBVR is not a language, but a metamodel for describing the business knowledge base of organizations (Sierhuis, 2013). One important enabling feature of this metamodel is improving the communication between experts and engineers (OMG, 2013). By using operators, like Boolean operators, concepts and facts can be expressed in relation to each other. These operators can be logical, like equal to, modal, which specify the value of a fact and quantifying operators. (Sierhuis, 2013).

2.9 Formalization of Scenario-Related Legal Text

The primary angle for scenarios is the AML sphere within the Finance domain. The initial documentation picked for the evaluation of the RA4T are the Anti Money-Laundering Directive 5 and 6 of the European Commission (AMLD). For evaluation purposes AMLD will be considered as the source documentation specified by (Rutledge & Italiaander, 2021) on which the SBVR rule set is based. As this is a legal document, any CNL used or interpreted must facilitate this for legal text as well. Whilst the object is regulatory, the interpretation and annotation require a legal viewpoint still. Additionally, legal text requires at least two domain experts, one for the legal text itself and a domain expert for the formal representation in machine readable rules (Libal & Novotná, 2021). To

eliminate this dependency (Libal & Novotná, 2021) propose to invert any legal translation back to its original text, which enables the legal expert to interpret and evaluate the translation to logic. Moreover, there is currently no general consensus on a single representation for expressing legal or regulatory information (Gaur et al., 2015). This notion expresses the need for a wide view on what kind of formal language can be used to investigate the domain. The literature presents us with several approaches to this problem. For one, the Natural Language to Knowledge Representation (NL2KR) platform can be investigated which translates legal text in variety of formalizations (Baral et al., 2013). Natural Language to SBVR (NL2SBVR) is another potential solution (Bajwa et al., 2011). Another solution is proposed by (Azzopardi et al., 2018), they suggest using CNL's for automated compliance checking for Financial Service Regulations (FSRCNL) including the AML environment as well. They use CNL to create a model of promised behavior in the payments ecosystem in the context of sender/receiver verification (Azzopardi et al., 2018). Whilst we consider this the most specific application of CNL to the case domain, there are some limitations; the most important is that FSRCNL is more oriented towards impact of European regulation on the United Kingdom (Azzopardi et al., 2018). The development of Regelspraak follows an interesting path. for Dutch tax regulations SBVR was not seen as a good fit, which led to the usage of RuleSpeak, whilst a better fit, the absence of certain constraints and operators led the authors to develop RegelSpraak (Corsius et al., 2021). This CNL is adapted to the Dutch Tax regulations, but went through a proof of concept for health care insurance, expressing flexibility in which type of regulations it can cover (Corsius et al., 2021).

2.10 Defining scenarios and end users in the AML environment

Scenario's need to be granular and well defined before they can evaluate the utility of an artefact (Hevner et al., 2004). Therefore, the case scope, domain experts and end users need to be addressed per scenario. The scenarios will be based on three indicators: the type of transaction, the object of this transaction and the end user of the interface relating to this transaction. These are the parameters considered the ensure scenarios differ enough to facilitate evaluation of the RA4T.

Transaction type

AMLD specifies a wide range of possible type of transactions. For scenario development, we will zoom in to three types and their specific patterns. For instance, a high-risk third country transaction indicates that a transaction originates from a pre-defined high-risk third country. Or a transaction that involves a sanctioned beneficiary, which connects sanctions regulation to the AMLD. Lastly, a suspicious transaction is picked, since it is accompanied by a specific set of end users but a general set of transactional objects.

Transactional object

A transactional object for this purpose can be currency transfers, virtual currency transfers and assets. The intrinsic nature and description of virtual currencies addresses some parts of the problem. AMLD describes them as a digital representation of value, which is not guaranteed by a central bank or authority. This exposes the decentralized nature of the currency.

End user or actor

Central in the scenario and the RA4T alike are the end users per scenario. The following are chosen for the scenario evaluation. The split is made on three levels of involvement: (1) insider, one who processes these rules at a financial institution. (2) Regulatory Outsider, one who interprets these rules from a regulatory perspective and (3) Outsider Person, one who experiences the rules but does not necessarily have access to them.

2.12 The Reference Architecture for Traceability in relation to the Scenarios

In that sense, the reference architecture could provide for a way to trace back business rules around finance regulations to their original source documentation to AML professionals working the field or explain to an Ultimate Beneficiary of a transaction why his or her transactions falls through. This enables insights in the knowledge base for professionals and non-professionals alike and guides them into the original train of thought that led to this ruleset. Tying back to the improvement of a certain condition specified in the theoretical framework.

As the development stage in RA4T, see figure 1, focuses on an agile way of working, we feel the need to shortly address the key characteristics of agile development as well. Agile at its core is software development in a unpredictable and complex environment by way of iterations instead of 'big bang' solutions (Cowan & Grushka-Cockayne, 2021). Organizationally this means teams are self-steering and many roles are interchangeable (Cowan & Grushka-Cockayne, 2021). The Agile approach ties back to the nature of Design Science on the notion of problem solving; whereas design is seen as Complex, Emergent, or Iterative and Integrative of multiple perspectives critical to problem resolution (Nerur & Balijepally, 2007).

3. Methodology

3.1 Introduction to the methodology

We are trying to answer the research questions by applying a scenario-based evaluation of the RA4T in an experimental situation. Whereas we try to compare the concrete effects by either applying or not applying the reference architecture to the specified scenarios. We then compare all four implementation steps of the architecture between the various scenarios.

To come to a good scenario-based evaluation of the RA4T we must transform snippets of identified source documentation and translate them into an unambiguous state as SBVR. Annotations link the rules to the source documentation. SBVR is widely used and standardized, and a lot of guidance is available (OMG, 2013). We then capture these rules in Web Ontology Language 2 (OWL). The scope of this paper is that the rules and vocabulary in SBVR and its results in OWL are limited to the scenarios defined in chapter 2. The rules ensure a yes/no answer from the set, but the ontology gives insights in the relations and explanations for the concepts within the scenarios. The combination of these two; clear rules and relations can assist with interpreting the source. The scenarios act as a funnel in which AMLD is pressed. These scenarios will be used to assess whether the RA4T by (Rutledge & Italiaander, 2021) can facilitate traceability throughout the layers for this domain.

3.2 Contribution Types in Design Science Research (DSR)

It is important to specify the type and impact of the contribution that it is expected to make. The contribution and impact from this paper is inherently linked to the RA4T from (Rutledge & Italiaander, 2021), for which we refer to figure 1. To assess a contribution's maturity, (Gregor & Hevner, 2013) propose a split on type and example artefacts that accommodate these types. They distinguish three levels of maturity of a contribution, ranging from more case specific to a higher level of abstraction.

Level one is considered the situated implementation of an artefact. A level higher defines the contribution as a nascent design theory. Finally the highest level of maturity is reached when the theory describes embedded phenomena (Gregor & Hevner, 2013). A phenomenon is considered embedded when it is persistent over time and their resulting information channels asynchronous with respect to the accompanying instructions (Moher, 2006). If traceability is considered the case in point and that notion is considered an embedded phenomenon, the reference architecture's contribution will zero in a third level of contribution maturity. (Rutledge & Italiaander, 2021) describe rule traceability as the enablement of an end user to trace interface behavior back to decisions that lead to it. Conforming to the requirements of an embedded phenomena, this concept

is persistent over time, traceability requires a complete chronological link to function, and it is asynchronous to the instruction that accompany the end-user interface (Moher, 2006).
Within this reasoning, the contribution of this paper will be limited to the first level of maturity, as it entails a situated scenario-based implementation of the RA4T.

3.3 Design Science Research Context and Methods

Research Question

This thesis follows the Design Science Research (DSR) Methodology, to answer the Research Questions.

Empirical Cycle

More specifically, it follows the road map for the empirical cycle proposed by (Wieringa, 2014), this roadmap provides guidance on how to approach knowledge questions within DSR. This Research Question is expected to have a descriptive answer. Furthermore, this question is related to a task in the engineering cycle, as it is part of the treatment validation (Wieringa, 2014)(Wieringa et al., 2012). We can tie this back to the improvement goals of the RA4T as artifact itself; it tries to improve traceability in interface behavior for rule-based systems (Rutledge & Italiaander, 2021). This combination leads to Utility-Driven Research (Wieringa, 2014). Driven by Utility, in this context means usefulness to improving the problem context, which gives it focus and connection to other research (Gill & Hevner, 2013). Defining utility as usefulness provides focus at the expense of narrowing implications, therefore, (Gill & Hevner, 2013) propose to complement usefulness with fitness, to come to a Fitness-Utility model for DSR evaluation. As a novel approach to evaluation in DSR, it will be used in its complementary nature. Generalizing the approach to Utility-Driven Research creates oversight in the steps to take and the methods to follow.

Treatments of the Objects of Study

This research covers a treatment of the Object of Study (OoS), in this case the AMLD by way of the artifact and by the RA4T to validate its usefulness on Business-Rule Traceability (Wieringa, 2014),(Rutledge & Italiaander, 2021). The Independent variable in this design is considered **the experimental treatment** with variable levels consisting of: Treatment with the reference architecture, and no treatment with this architecture. The dependent variable of this treatment design is **the measured variable** defined as Business-Rule Traceability. We therefore need to operationalize Traceability in such a way we can pin expected changes to it. Causal inference is the characteristic to measure with regards to the treatment, as exactly only the treatment is applied. This poses some problems as one must rule out all possibilities of confounding influences (Wieringa, 2014)(Shadish & Cook, 2002).

Method Selection

As usefulness is a driver for doing the research, rigor is the driver for method selection (Gregor & Hevner, 2013). The degree of rigor ascertains which methods are viable for selection and is defined in its expected validity (Straub & Ang, 2011). The method selected for this paper is the Single-Case Mechanism Experiment, with the goal of describing and explanations of the action-reaction occurrence of the OoS (Wieringa, 2014). Also defined as single-case causal experiments (Wieringa, 2014)(Wieringa, 2014). By applying the experimental treatment of the RA4T, we can investigate its effect on the measured variable in Traceability.

Validation Model and Treatment Design

A validation model is constructed by way of the OoS. Preparing the RA4T from (Rutledge & Italiaander, 2021) for appliance to AMLD in business rules makes it technically a prototype of that artifact and the context is modelled after AMLD.
Before a validation model is considered valid, we must adhere to two standards set forth by (Wieringa, 2014): It should support several inferences, and it must be repeatable by other researchers. With a strong link to expected causal relations, the validation model must support

architectural explanations. This means that any change in traceability within the original architectural components of the OoS need to be visible when the reference architecture is applied. It is also important the approach to the OoS is repeatable. Relevant for assessing architectural inference support is whether the information is exact enough to do analysis and if variations in the validation model can limit answers to the Research Question (Wieringa, 2014).
For treatment design we apply scenarios to the validation model, comprised of artifact and context (Gregor & Hevner, 2013; Venable, 2006).

The Reference Architecture for Traceability

We will try to closely resemble the proposed steps from the RA4T to ensure the OoS is operationalized for this framework. The steps are as followed: First we need to develop the business rules and code from AMLD in a CNL, then we need to compile this via an applicable editor or knowledge manager, and finally, come to an interactable user interface for end users (Rutledge & Italiaander, 2021).
The validation model is then treated to the scenarios specified in the theoretical framework. Effects on business-rule traceability are then investigated. The RA4T will function as a lay-over to investigate whether the solid lines that imply the traditional view without traceability are there in the experimental situation (Rutledge & Italiaander, 2021). Furthermore, looking at the reference architecture (Figure 1), the dotted lines will illustrate places where it can assist with implementing traceability (Rutledge & Italiaander, 2021). As there are four layers of traceability facilitated by the reference architecture, these will be part of the validation model, and full traceability culminates to exactly four implementation areas (Rutledge & Italiaander, 2021).

3.4 Data Analysis Methods and Tools

To analyze the effects of the treatment on traceability we must be able to: annotate documents and design corresponding CNL code and then transform this code to machine processable language. Furthermore, we need to edit any resulting ontology and model. We need to transform these rules to a formal logic representation; SBVR lends itself for automated transformation to OWL2 (Reynares et al., 2014) and is used for the RA4T as well (Rutledge & Italiaander, 2021). To be able to analyze and edit the ontology, we used the open source software Protégé (Musen & Team, 2016) which covers all blocks from the automation stage (Rutledge & Italiaander, 2021). When SBVR and the source documentation is stated in Hypertext Markup Language (HTML) and accompanied with the URLS for the Semantic Web then this should allow for analysis and cycling through all layers of traceability for the user, achieving the desired effect in traceability (Rutledge & Italiaander, 2021). This cycling through traceable layers can establish the causal relation of the reference architecture to the traceability, underscoring the importance of the editor and Graphic User Interface (GUI) in relation to the analysis of the data.

3.5 Plan of Approach

Based on the predefined scenarios, we will annotate the AMLD into business rules. For the input data, the Society for Worldwide Interbank Financial Telecommunication (SWIFT) transaction type messaging will be used as overlay to structure the scenario characteristics. The message type used will be that of single credit transfer MT103, which is one of the most used payment messages worldwide (Cook & Soramäki, 2015). Whilst it does not fully capture the specifics of virtual asset transfers and the related blockchain technology, we will use their naming convention in the MT103 type like messages. Making public and merging some parts of these message types is also proposed by (Hardjono et al., 2021). We will confirm the MT103 message fields needed to capture the scenarios, based on: their transaction type, object, and end user. These we will then structure into one specific input data format used for all possible scenarios. This is a transaction message particular to this paper. The source documentation is stored in HTML and made accessible via their semantic web URLS. While s2o is a welcome solution for conversion into OWL2, we will also keep track of the

ontology in a manual way via the editor Protégé. For style sheet display Fresnel could be used, as the authors of the RA4T have used this in similar approaches as well (Rutledge et al., 2021).
To set up the evaluation environment, we must identify the regulations that cover the scenario best. An AML expert could be helpful, but there is a lot of official guidance material available for AML which is the discussion phase in text. Then we must annotate the applicable regulation into objective indicators, also here the guidance documents are very helpful. Furthermore, jurisprudence can also assist with pinpointing the specifics in practice. This annotated regulation we will then translate to SBVR to get rid of any remaining ambiguity. The SBVR in turn must be converted to OWL, this can be done automatically via s2o, but we opted to manually set up and edit the ontology in Protégé. The ontology is set up in accordance with the business rule in SBVR and the properties are checked against the guidance documentation. Inferred classes based on the rule are annotated with an HTML link that links to the rule in SBVR first and the source document in HTML second. After set up we need test data to evaluate the scenarios, we opted to utilize the individual's matrix plug-in. This way we can easily input object and data properties for various scenarios. Finally, we let the reasoner run on the test data to infer the next course of action for the test transactions. For these conclusions we will check if the end user can come to explanations, which rule and or data triggered this rule and from the triggered rule to the logic from the SBVR all the way to the source regulation in the applicable AMLD.
For this scenario-based evaluation, we assess how well we can implement traceability on the traceability links of the RA4T. All these steps require the implementation of traceability according to the reference architecture (Rutledge & Italiaander, 2021).

3.6 Converting regulations to unambiguous business rules in SBVR

The previously defined scenarios all tie back to one or multiple pieces of text in any AMLD. The way European directives work, is that they prescribe the legal implications, to be implemented by member states and translate into national law. The Dutch central bank (DNB) then gives guidance on how to operationalize the national law derived from the AMLD. Therefore the (*Leidraad Wwft En Sw*, 2020) is used to quickly come to official interpretations and implications of the national law, and then in turn link this to the AMLD. With help from the (*Leidraad Wwft En Sw*, 2020) we will identify the scenario-related parts of AMLD, and the conversion to SBVR can begin.

3.7 SBVR Construction of AMLD-derived Business Rules

Within this paper, SBVR is used in an operative rule context, in which they describe business processes (Reynares et al., 2014). Research by (Al Khalil et al., 2016) states that SBVR behavioral (operative) rules correspond to legal regulative norms, which in turn are characterized by their deontic modality. This implies that for regulative norms behavioral rules in SBVR are a good fit. To indicate the deontic modality of the rule in SBVR, the word obligatory is used. In research by (Joshi & Saha, 2020), an automated approach to semantically enriching regulations is used to quickly extract the deontic expression via certain words. We used these words to come to the deontic expression in AMLD. In that sense; should, must and shall, are words used to define obligations. We use these words as they are similarly used in AMLD. Within all three scenarios, the deontic expression of obligation is the starting point. The rules in SBVR therefore all start with ''It is obligatory". However, to come to the rule in SBVR an effective summarization of the legal text must be distilled based on the annotations, deontic expressions, and the existing guides.
(Al Khalil et al., 2016) propose to declare an SBVR rule in SBVR vocabulary except for the modality and keywords. Except for modality and keywords, we can break up the rule in noun concepts (NC), verb concepts (VC), verb symbols (VS) and verb concept role (VCR). Critical are the verb concepts as they function as the conditions of the rule (Al Khalil et al., 2016)

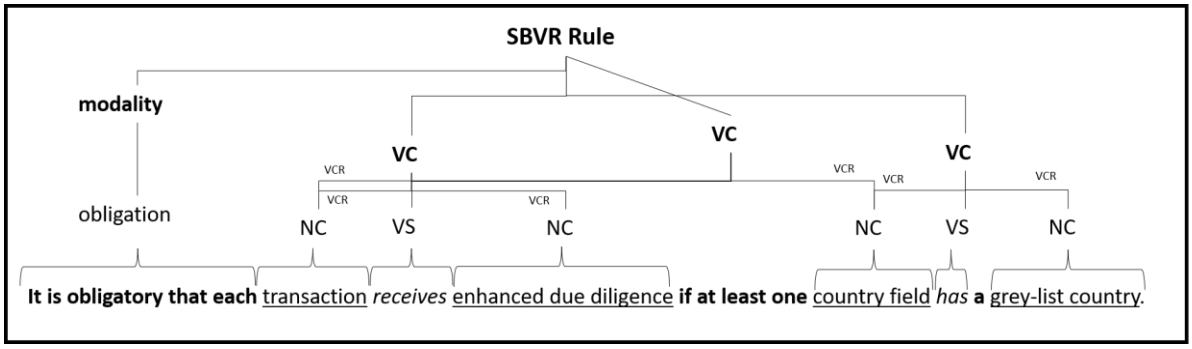

Figure 2: SBVR rule for High Risk Third Country, deconstructed in verb concepts, noun concepts, verb symbols and verb concept roles.

3.8 From SBVR to OWL

Whilst more automated approaches are available, such as the previously mentioned s2o, it is meaningful to convert the rules manually. We deconstructed the rule above in the corresponding concepts, which primed it for conversion to OWL. (Reynares et al., 2014) state that the noun concepts (NC) above translate to classes in OWL. Verb concepts (VC), or fact types, are mapped depending on their binary relation to a class domain and range. Fact type roles (VCR) specializes its instance when they involve certain fact types (Reynares et al., 2014). Also, the quantifications need mapping, in the example above we consider an existential quantifier (Reynares et al., 2014); each transaction that receives enhanced due diligence will involve grey-listed countries, and each transaction that involves grey-listed countries will receive enhanced due diligence. To convert the concept of segmentation in SBVR, we need to map to the OWL concept of Disjoint Union (Reynares et al., 2014). In the first rule this ensures that due diligence levels, country fields and certain country lists are segmented, and transactions cannot involve other segments.

4. Results

4.1 Introduction to the results

In the following paragraphs we will discuss the results from the experiments, where we applied the RA4T to the defined scenarios. We will experiment on three scenarios, where a diverse set of focus angles is chosen: We will incorporate various actors in the process, look at the type of transactions and we will consider the object, or currency. This approach ensures a varying set of scenarios to cover geographic locations, natural persons, and currencies, be it regular or virtual. Together the scenarios stretch over the core breadth of AML directives. We will describe the process in the following sections.

4.2 High Risk Third Country Scenario

For the high risk third country scenario the triggering rule for which an explanation can be requested is that each transaction which involves a high risk third country must receive enhanced due diligence. In Protégé, we set up a disjoint relation between standard, enhanced and unacceptable due diligence and we inserted the country list for grey-list countries as identified by the competent bodies. When an explanation is requested, Protégé will stipulate that standard due diligence is not allowed due to some countries being on the grey-list. The end user can then in turn click on the annotation of this rule to be linked to a webpage which contains the rule itself and a hyperlink to the marked source document in AMLD5. Appliance of the reference architecture not only ensure tracing back to the source documentation, but also enriches the rule with knowledge as to how it came to be. Without appliance, we would have just a rudimentary explanation for the triggered rule.

Guidance type: Operative Business Rule
Description:

''Business relationships or transactions involving high-risk third countries should be limited when significant weaknesses in the AML/CFT regime of the third-countries concerned are identified, unless adequate additional mitigating measures or countermeasures are applied. When dealing with such cases of high-risk and with such business relationships or transactions, Member States should require obliged entities to apply enhanced customer due diligence measures to manage and mitigate those risks. Each Member State therefore determines at national level the type of enhanced due diligence measures to be taken with regard to high-risk third countries''. (DIRECTIVE (EU) 2018/843: Article 12(1))

It is obligatory that each transaction receives enhanced due diligence if at least one country field has a grey-list country. For this rule, a few axioms in OWL are considered, namely the country grey-list as subclass of the country field, and these grey-lists being disjoint with their class siblings.

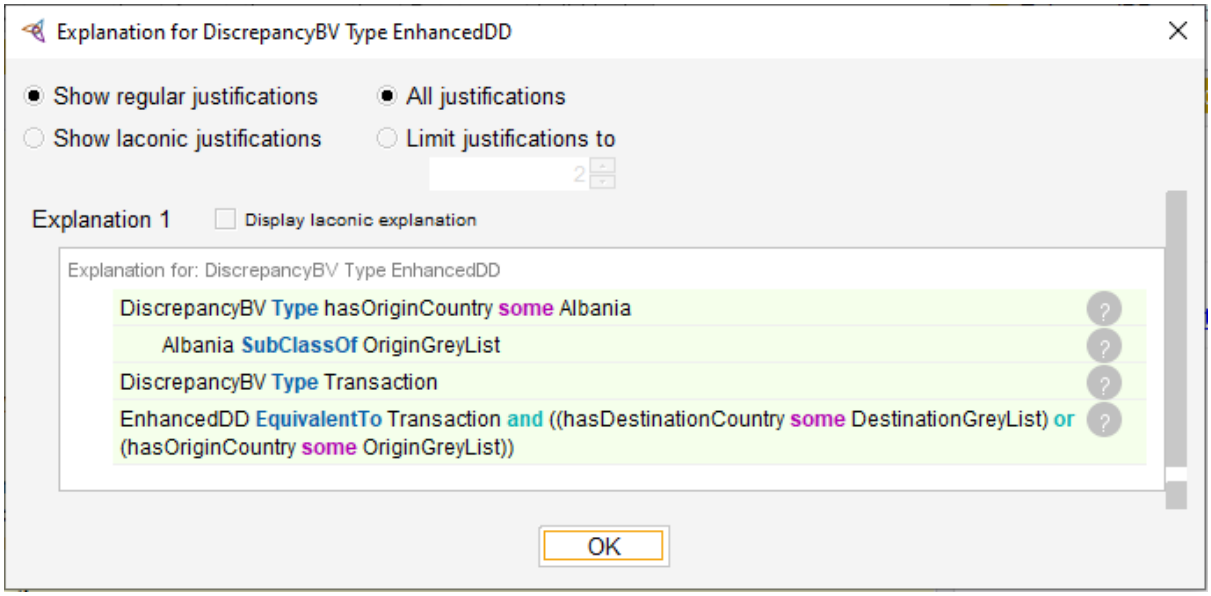

Figure 3: Protege explanations for applied Due Diligence

In the above example, Protégé explains that before we can consider Enhanced Due Diligence, there needs to be an instance of a transaction, and this transaction must involve some grey-listed country.

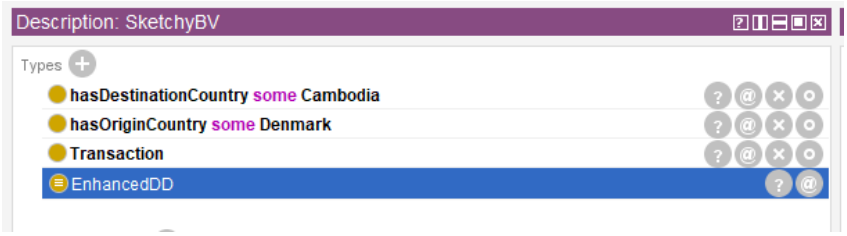

Figure 4: Inferred Due Diligence level for transaction

The other way around, we can also have Protégé reason that this transaction requires EDD based on the specified rules.

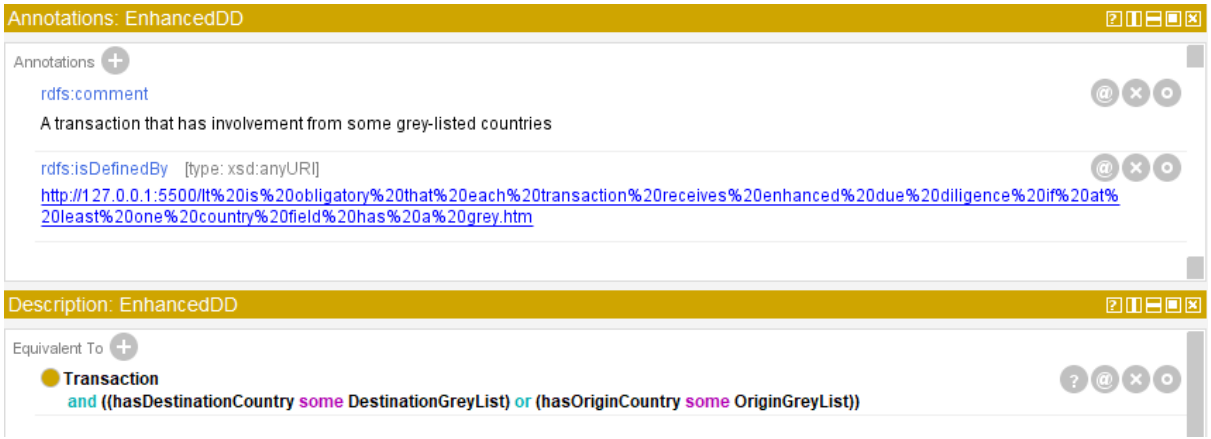

Figure 5: Rule for Enhanced DD and Link to SBVR, annotated with rdfs:isDefinedBy

To facilitate an understandable environment for domain experts and developers alike, we use the Resource Description Framework (RDF) Schema for annotation. This offers mechanisms for describing resources, and the relationships between resources (Brickley et al, 2014).
We then use the rdfs:isDefinedBy to specify that a resource, in our case the business rule, is defined by another resource: the link to the document which contains the SBVR. In turn, the SBVR is defined by the source documentation. The rdfs:isDefinedBy points to the source of a definition, by defining it with an Internationalized Resource Identifier (IRI) (Kara et al, 2022). This standardized way of describing the relationship between resources provides for an unambiguous and transparent approach and helps RDF browsers to distinguish between main resources and ancillary resources (Sauermann et al, 2007). This is especially relevant for development, and more so development instructions due to a well formatted and widely used schema such as RDFs. The rdfs:isDefinedBy can also be queried on, granting additional means of querying on traceability within vast amounts of data. We could propose for this very specific ontology to add a sub property to the rdfs:isDefinedBy to specify further the origination in the source documentation. We propose to use in this ontology the OriginatedFrom sub property. This leads to the explanation that is defined by the rule in SBVR, which in turn originated from the marked text in AMLD.

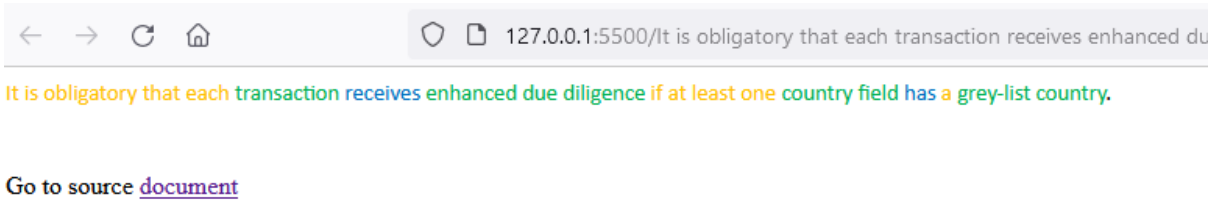

Figure 6: HTML SBVR and link to Source

Above the SBVR is specified, and to enable full traceability to the source documentation we can now click on the link ‘Go to source document’. We are then linked to the highlighted portion of the AMLD source document from which the rule is derived.

(12)Business relationships or transactions involving high-risk third countries should be limited when significant weaknesses in the AML/CFT regime of the third-countries concerned are identified, unless adequate additional mitigating measures or countermeasures are applied. When dealing with such cases of high-risk and with such business relationships or transactions, Member States should require obliged entities to apply enhanced customer due diligence measures to manage and mitigate those risks. Each Member State therefore determines at national level the type of enhanced due diligence measures to be taken with regard to high-risk third countries. Those different approaches between Member States create weak spots on the management of business relationships involving high-risk third countries as identified by the Commission. It is important to

Figure 7: Highlighted sections in source documentation

In the described scenario the rule is triggered due to a transaction being marked as "requires Enhanced Due Diligence", the other way around, we could also have the reasoner decide what level of Due Diligence a transaction requires, based on the involved countries. If an explanation is requested, Protégé will link to the defined class Enhanced Due Diligence which is annotated with an URL which leads to the rule in SBVR, this same page has another link to the source documentation in AMLD5 where the relevant text is marked in yellow. At first glance traceability is implemented and realized at the 4 points. Whilst it is already meaningful to have this current solution, proposed stylesheets which include additional legal and policy insights can increase the knowledge an end user has at its disposition. Legal documentation often refers to lists and annexes, enabling browsing through parts of the business vocabulary would be helpful. The first scenario can be considered a success as a triggered rule or automated inference can already give insights in the rules and documentation that led to them.

4.3 Suspicious Transaction involving Cryptocurrency Scenario

This scenario involves an FIU investigator and AML professional as end users where the inferencing capabilities of Protégé can reason whether a transaction should be marked as suspicious. The rule is

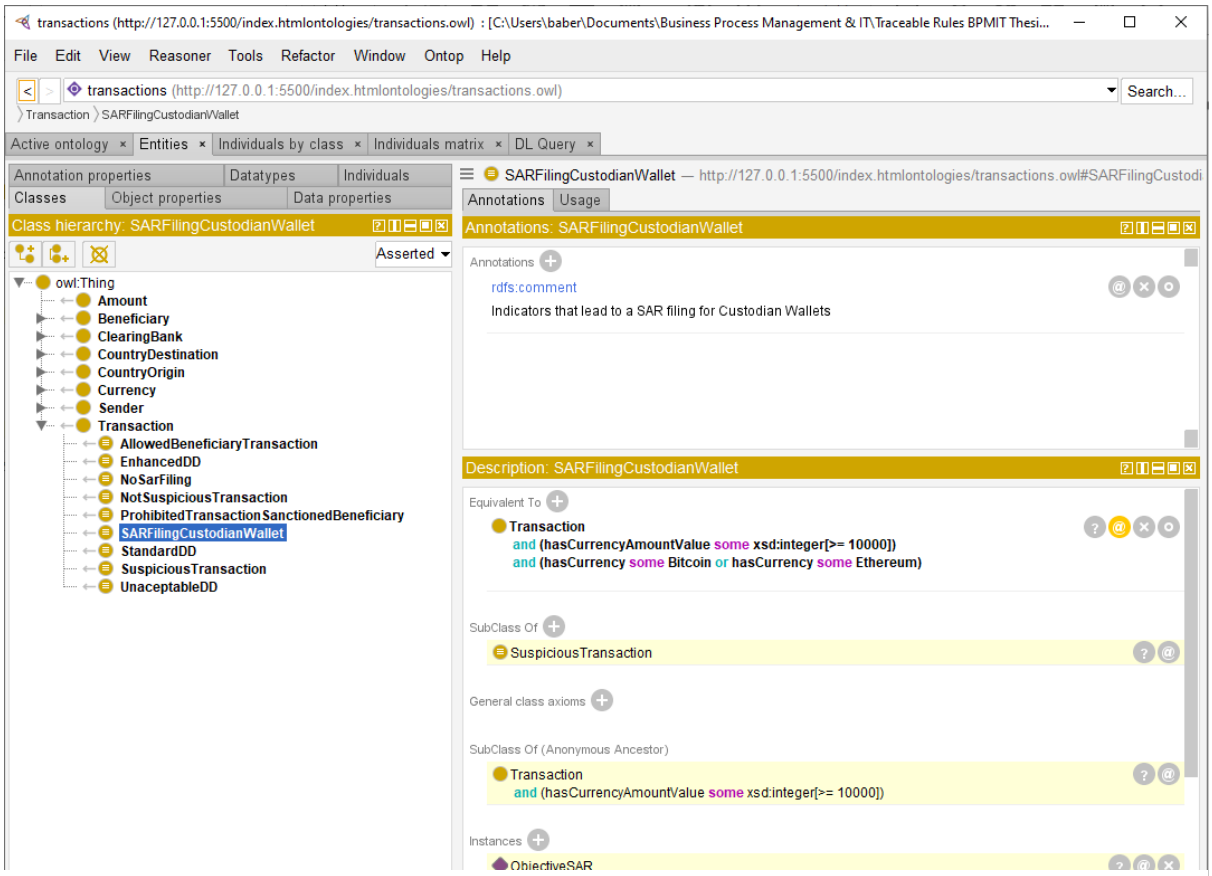

Figure 8Description SAR Fileable Transaction

derived from guidance on AMLD5 on how to objectify certain loosely defined scenarios. Furthermore, this scenario involves Cryptocurrency as well. The combination of currency and value specify if a transaction should be considered for a Suspicious Activity Report (SAR) filing involving custodian wallets. For the scenario to be successful, we must be able to designate a transaction as suspicious, give basic explanations of why it is considered as such, and link to the rule in SBVR and finally into the source documentation. The SAR fileable transaction is composed of an integer data value to designate the value and in which cryptocurrency it is. We reasoned on test data named 'Objective SAR' which was denominated in Ethereum and has a value of 10.000. These thresholds could lead to the filing of a SAR, and only if they are both reached. Once the threshold is hit, and it involves cryptocurrency, the reasoner will infer that this is a SAR fileable transaction. The Test data functions as expected in the scenario, and in its annotation gives insight into the rule. You can request Protégé to give an explanation, and it will look like the following:

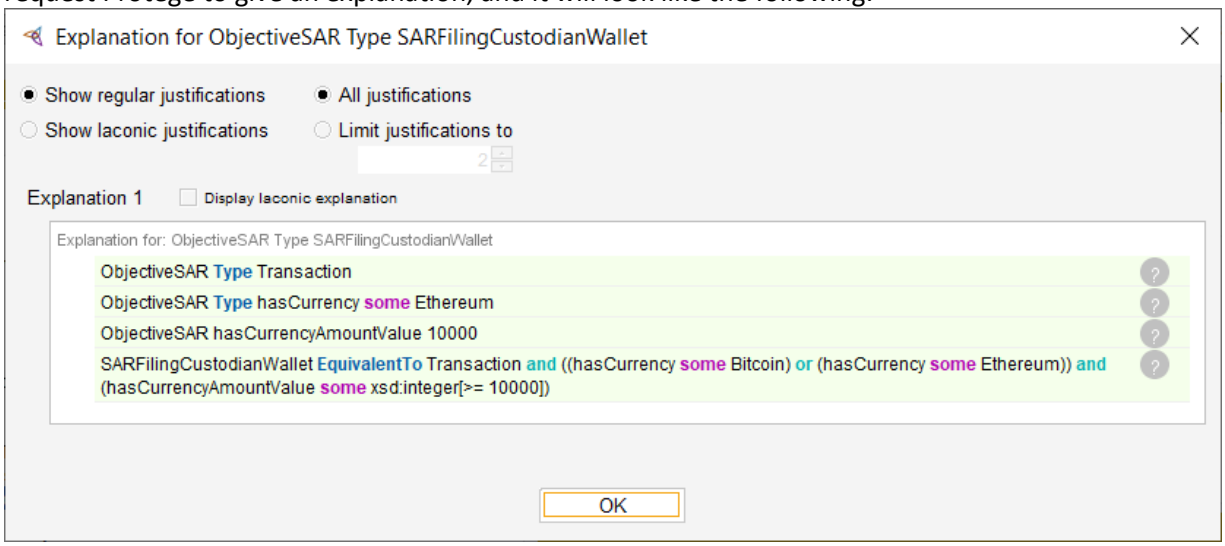

Figure 9 Explanation for SAR fileable Transaction

The SAR filing is annotated with an HTML link to the following rule in SBVR: *It is obligatory that each transaction that has a currency integer value of at least 10.000, and is denominated in a cryptocurrency, results in a Suspicious Activity Report Filing.*

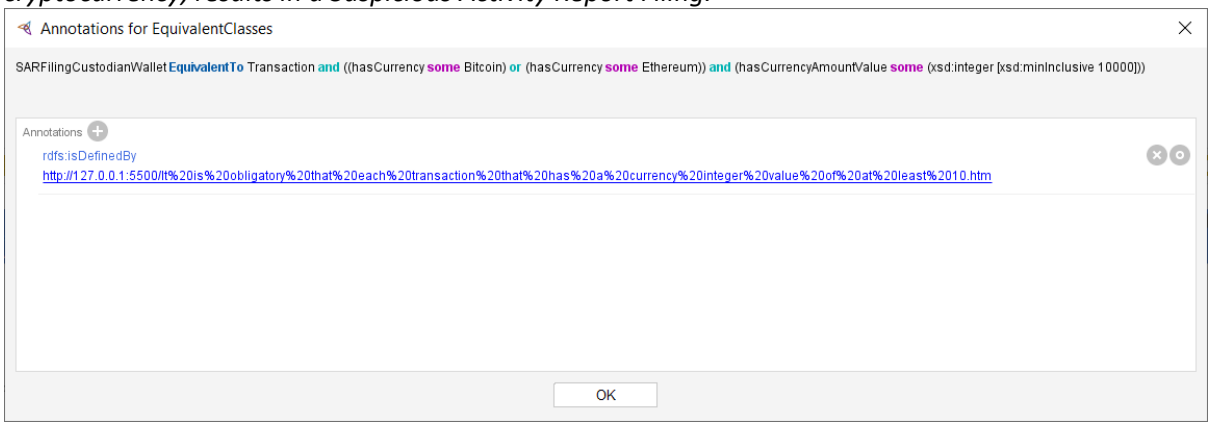

Figure 10 rdfs:isDefinedBy a link to the SBVR rule in HTML

We can deconstruct the rule based on the modality, which is obligatory. We identified this modality by the keyword shall. The obligation is to file a SAR if certain criteria are met. There is guidance available for these criteria. However, an obliged entity has a lot of decision room on what it considers as suspicious. The criteria above are chosen due to the ability to demonstrate data properties in Protégé. To continue, we broke the rule further down in NC, VC, VS and VCR and tied that back to their corresponding OWL. These manual mappings give good understanding of OWL, but more importantly, reflect the use of a CNL in the RA4T. Legal modality can be expressed more

concisely, and annotations in turn can specify what parts fall under the legal obligation, and which parts are under the obliged entities discretion. This moves any discussion an obliged entity could have with the regulator straight to the rule itself as it makes transparent the reasoning. Finally, development speed is increased because after the rule is set up, development and discussing the rule can be done in parallel. Especially with a more automated approach to SBVR to OWL conversion. The final link in the SBVR HTML document links to the documentation from which the rule was derived. By now all traceability links from the RA4T have been implemented and cover the full scenario. The scenario is a good fit to the architecture and provides for valuable information to end users, even more so for communication between end users. It not only explains why a system marked a transaction as suspicious, it also gives insight into the origination and substantiation of the rule to other bodies like the Financial Intelligence Unit (FIU). Furthermore, developments within the world of custodian wallets are moving fast. A date time and validity check of the regulation could prevent actioning on old information, this can be a proposal for extension to the RA4T. As the RA4T currently does not cover the concept of date / time, nor improvement efforts based on event data. The applicability and match of the RA4T to the domain exemplifies the how and to what extent RA4T assist with developing BIS, where it can improve traceability over the breadth of AMLD.

4.4 The Sanctioned Beneficiary Scenario

The last scenario covers the transaction process from the originating party, the one sending the money or virtual currency. This also shifts the understanding of the end user from an expert to a layperson. Therefore, it is important to avoid any yes/no explanations, as the end user does not necessarily understand what is blocking the transaction. The rule and general policy are straightforward; sending money to a sanctioned individual is prohibited.

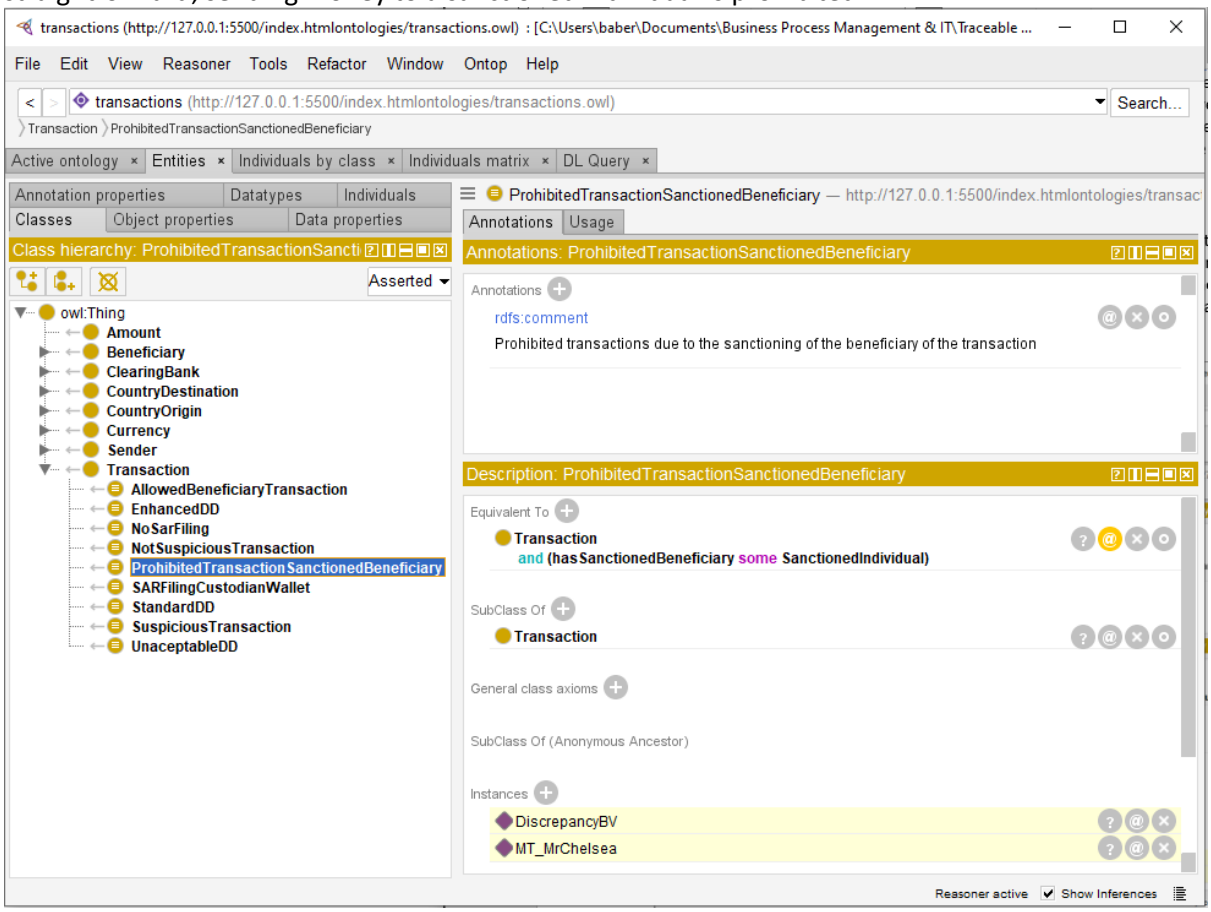

Figure 11 Prohibited Transaction to Sanctioned Beneficiary

However, for the end user it will only be a blocked transaction if the name is an exact match, a partial match will still require a review from a professional at the side of the clearing bank. Fuzzy

matching is out of scope for this exercise, but is something to be investigated for these types of scenarios. The originating party can request an explanation of why the transaction is considered prohibited, and this will lead to the description that some sanctioned individual is involved. This description is annotated with an HTML link that's links to the rule in SBVR: *It obligatory that a prohibited transaction involves at least one sanctioned individual.*

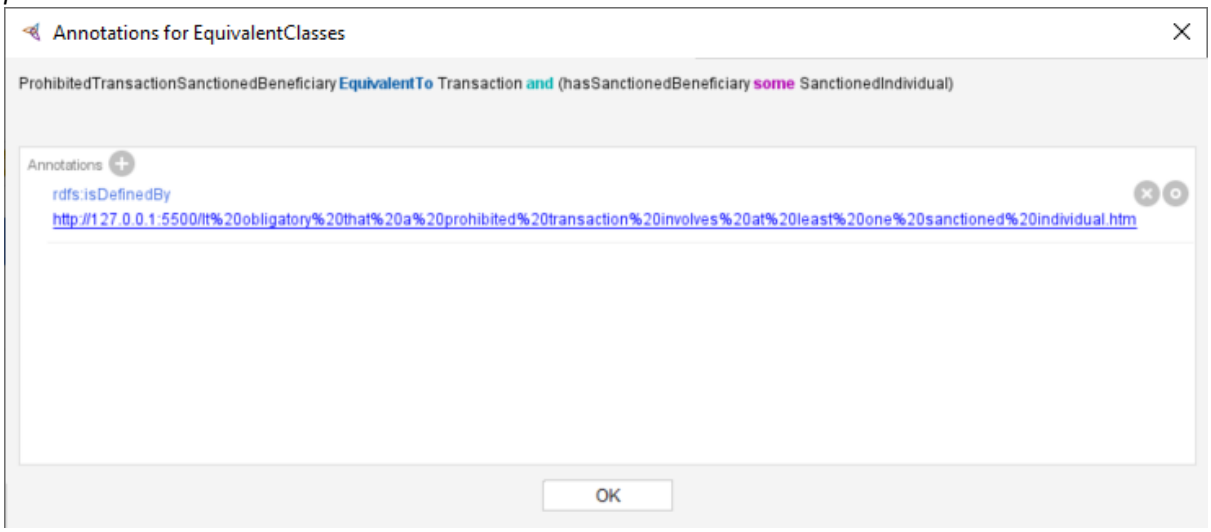

Figure 12 HTML link to the rule in SBVR

The modality of this rule is obligatory due to the word shall in *DIRECTIVE (EU) 2018/843: Article 30 a: i.* The modality covers from an obliged entity perspective the facts that transactions that involve sanctioned persons are prohibited. The brevity of this rule is useful as it gives layperson end users enough information, but also any professional who gets this information has a general understanding of what happened. The noun concept sanctioned individual is interesting as this could be linked in real time to the sanctions list data from the EU. This shows the flexibility a CNL can offer, and fits an agile environment. Protégé, when requested, will give the following explanation for the rule.

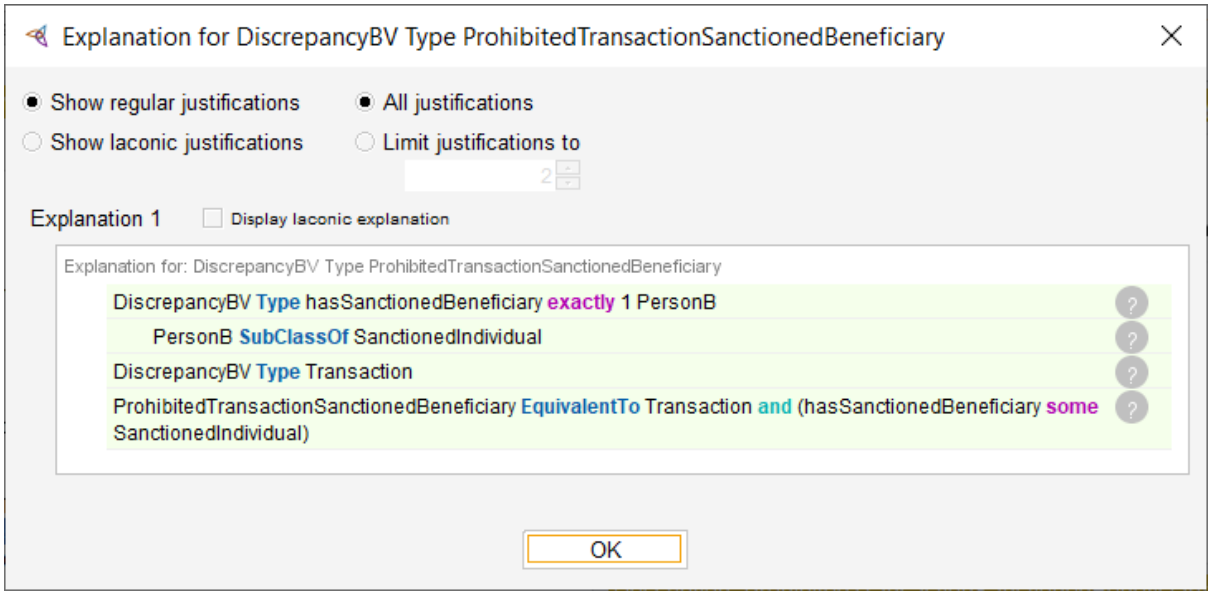

Figure 13 Explanation for the Sanctioned Individual

This allows us to see that any transaction involving a sanctioned individual is prohibited. Traceability is intact due to the ability to click on the explanation, and be guided straight to the HTML link to the rule in SBVR. In turn the linked to page contains a link to the source documentation. Whilst this covers the traceability links from the RA4T, it can still leave anyone who is not an expert, longing for

more information. By providing a link to the current and public list of sanctions, the end user now has all the relevant information at its disposal. This also implies any future transactions can be checked against the lists, instead of brute forcing the transaction system itself. An additional consideration is the approach; should a transaction to a sanctioned individual be considered an inconsistency, or does it need automatic inferencing to the label of a prohibited transaction.

4.5 Summarizing Scenario Outcomes

The scenarios are a good fit to the reference architecture, as most regulations for transaction are rule-based anyway. It also helps those transactions are formatted automatically as event data, and do not require conversion to standards. This ensures that the scenarios can be set up from similar source data, and a single transaction in this format can cover for multiple scenarios. While a singular format is no prerequisite for the architecture to function, in this evaluation, it was very valuable, as it helped successfully implementing the traceability links from the RA4T. The format also assists with transparency with regards to triggering data as the fields are accompanied with a basic explanation of what the field contains. Lastly, in this format, it is easier to generate a lot of test data. This can provide for potential parametrization within the system. During scenario evaluation, we identified that the potential of Date/time information from the transaction event data is not used. Ideally the metadata of system time is held against the validity of the regulation itself, to see whether the rule is still current. This allows for versioning and automatic audit trail. This could serve as a useful extension to the reference architecture.

5. Discussion, Conclusions and Recommendations

5.1 Remarks before discussion and conclusion

Looking back on the experiments and the scenario-based evaluation, revisiting the research questions and approach is a valuable and necessary step in evaluating this thesis. Also, the implications and implementations using the rdfs:isDefinedBy resource annotation as a technical solution for backward linking will be discussed.

5.2 Discussion and Reflection

In the experiments, we knew based on RA4T, there were four area of implementation for the traceability links. These areas are the critical parts to find out in what sense the RA4T facilitates BIS development that support full traceability in the finance domain, as these links function as check-in points for traceability across the E2E chain. Successfully implementing the traceability links on the four areas says a lot about the match the RA4Tcould have with the domain, but it is just as important as to investigate to what extent these traceability links facilitate development of systems that support traceability, and if the traceability chain stays intact throughout the whole development. Annotating the source documentation to make it fit for conversion to SBVR posed some difficulties due to articles referencing other articles. This required the following work-around: Our angle was to start off with the deontic modality and then refine the rule itself. Identifying the modality within legal rules can be straightforward, but regulations themselves can pose a challenge due their ambiguity regarding practical implications of the legislation. This holds true for AMLD, as directives are intended to be translated to national law of member states. The document-derived rule therefore must lean also on guidance documentation and discussion between member states. However, this is not necessarily a bad thing. Linking specific parts of the rule, whereas some parts are derived from the directive, and linking other to guidance documents, provides a transparent overview into decisions. These rules can therefore be the anchor points in discussions, either in company or with external stakeholders such as the regulator. It is worth noting that the literature is divided on whether SBVR is the best possible fit for regulations, or that a more legislative-oriented

CNL is better fit for this purpose. Recent work by (Rutledge et al., 2022) addresses this by making the architecture CNL agnostic.
The SBVR to formal logic conversion in OWL posed dual options for approaches. The s2o tool is an automated approach, and can speedily map SBVR to OWL on the press of a button. As the traceability link is there in HTML the original rule before OWL mapping remains. Via the automated way, traceability is therefore successfully implemented. We chose a second approach, by deconstructing the rules and mapping to OWL manually in line with the work of (Reynares et al., 2014). This has the advantage of accumulating a feeling for the mappings in OWL themselves, which in turn can lead to a better understanding of how concepts in SBVR and OWL interact. This might prove valuable for communication at domain expert level. In a sense this also facilitate traceability as professionals now can infer how OWL relates to the SBVR concepts within the rule. A lesson learned is that the plan of approach did not fully cover an automated approach, as we did not get the s2o tool to work with this research' SBVR. This automation implementation is important to increase development speed and is a better fit for an agile software development environment than the manual approach as specified above. This increased development speed is reflected the latest paper that involves the RA4T by (Rutledge et al., 2022).
Modelling the rules in OWL was done via expressions and quantifiers, for which Protégé has very ontology-oriented interface. As the SBVR is already in an OWL environment, setting up the rule according to the SBVR is quite easy, if the existential quantifier is in place, and annotated with an rdfs:isDefinedBy link to the SBVR in HTML. Whilst due to the RA4T, the traceability improved in the absolute sense, as in the natural state traceability is normally not catered for, we can also zoom into whether transparency improved. The rdfs:isDefinedBy adds additional transparency by giving another level of information about the relation between rules, their rule in SBVR and all the way up to the source. It does so in multiple ways: first, the annotations are anchored in the code and clarify for developers the relation between resources. Secondly, the annotation semantically enriches a more formal model, as agents in the process generally understand the concept of something being defined by. Finally, once the system fills up with data, one could query the annotation model as well for ad hoc needs.
When the model is reasoned upon, it will be able to show you inconsistencies, or will infer the next course of action. To implement traceability, it was crucial to implement the HTML link to the expression. Protégé facilitates users in requesting explanations on inconsistencies across the model, and can provide explanations based on expression level. The traceability links on both levels must be in place to ensure traceability. This was successful over all three scenarios, and from requesting explanations, we were able to trace back to the pieces of AMLD that led to the rule. It is interesting that we can go about it from two different angles: firstly, we can opt for an inconsistency model, whereas the model tells you what cannot be done in the model based on the rules, and which data breached the rules. Another approach is to have the reasoner infer what the boundaries, or next steps, based on the rules are. This is something which requires a more coherent strategy in the plan of approach, which was not in place during this experiment.

5.3 Conclusion

Concluding this research, the RA4T facilitates the development of business information systems in the finance domain by: giving clear indications on traceability implementation areas, whereas these process steps for implementation map quite well to the actual steps in the domain. It specifies the specific steps to be taken and the order in which to do so. Furthermore, it can serve as a layover for fixes when the traceability link is broken and point to potential problem areas. The RA4T also serves as a high-over conceptual model to use when developing any rule-based systems, and splits it into three comprehensible blocks (Development, automation, and end user). Picking SBVR ensures automation options are available, but recent changes to the architecture can impact automation potential, especially CNL to OWL mapping. Coupled with the fact that SBVR is not always the best fit

for regulations, and alternatives like RegelSpraak have been proposed (Corsius et al., 2021), the balancing act between automation and CNL fit for legislative text remains.
The extent to which RA4T facilitates developing BIS for the Finance Domain that support traceability is far reaching. By introducing the possibility of tracing back from end user to regulation in a structured way has the potential to drastically change the approach of rule-based system development when it comes to regulations. The significant difference is that it is possible to trace back to source documentation, whereas in the natural state this is not taken into consideration from an end-to-end perspective. The true extent though, is facilitating transparency into the reasoning behind the rules, and making that reasoning available to the correct audience. Whether this audience is a natural person looking to find out why a transaction bounced, or the regulator that wants to have an overview of their prescribed implementations. The RA4T currently does not cover the concept of date time, by which it cannot facilitate event data or automated checking of regulation validity. Therefore, extending the RA4T to include date / time functionality, is our proposal for improvement to the reference architecture. How to implement this is specified further in the recommendations.

5.4 Practical Implications and Recommendations

The RA4T serves as a reference architecture, and successfully so. It can be positively used to ensure developing BIS which provide for traceability to the origin of system decisions. If used on the finance domain, compliance is a recurring theme. The RA4T can be utilized for ensuring compliance with the applicable regulations, regardless of whether it is concerning AML or any other regulation.
Businesses are not necessarily aware of their want for traceability within their rule-based systems, and this requires a pitch on the advantages of introducing the concept. Finally, after evaluation, we propose to implement the Date / Time dimension in the reference architecture as well. By logging and incorporating in the model, one opens the possibility to check against regulation validity, and accrue relevant event data derived from the system. This data is available via the numerous steps, but does require consolidation in one repository. If done so it can indicate bottlenecks in the process and open a plethora of automated process mining options. Its relation to the RA4T could be overarching the three process blocks of development, automation and end user and serve as shadow repository from which improvement exercises can be initiated. This fits the continuous improvement idea from the agile mindset.
For researchers looking into this topic for this domain, I recommend to refine some methodological concept in more detail.
A clearer approach to document gathering for SBVR conversion, and describing the relations between legislative text and guidance documents can facilitate a faster and more transparent way to come to rules in CNL. Furthermore, incorporating an automation strategy for SBVR to logic conversion can increase the speed in which development can be done. Finally, it is important to pre define the conceptual boundaries of the model, meaning; to know upfront whether it will be based on explaining inconsistencies after adding data, or that it will infer, based on the data what the course of action within the system should be.
Exploring the requirements of tools that can cover the process End to end is an interesting direction for future research about this topic.

6. Bibliography

Al Khalil, F., Ceci, M., Yapa, K., & O'Brien, L. (2016). SBVR to OWL 2 mapping in the domain of legal rules. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9718*(July), 258–266. https://doi.org/10.1007/978-3-319-42019-6_17

Azzopardi, S., Colombo, C., & Pace, G. J. (2018). A controlled natural language for financial services compliance checking. *Frontiers in Artificial Intelligence and Applications*, *304*(666363), 11–20. https://doi.org/10.3233/978-1-61499-904-1-11

Bajwa, I. S., Lee, M. G., & Bordbar, B. (2011). SBVR business rules generation from natural language specification. *AAAI Spring Symposium - Technical Report*, *SS-11-03*(June 2014), 2–8.

Baral, C., Dzifcak, J., Kumbhare, K., & Vo, N. H. (2013). The NL2KR system. *CEUR Workshop Proceedings*, *1044*, 37–47.

Bellomarini, L., Laurenza, E., & Sallinger, E. (2020). Rule-based Anti-Money Laundering in Financial Intelligence Units: Experience and Vision∗. *CEUR Workshop Proceedings*, *2644*(September), 133–144.

Booth, A. (2008). Unpacking your literature search toolbox: On search styles and tactics. *Health Information and Libraries Journal*, *25*(4), 313–317. https://doi.org/10.1111/j.1471-1842.2008.00825.x

Cook, S., & Soramäki, K. (2015). The global network of payment flows. *The Journal of Network Theory in Finance*, *1*(2), 21–52. https://doi.org/10.21314/jntf.2015.014

Corsius, M., Hoppenbrouwers, S., Lokin, M., Baars, E., Sangers-Van Cappellen, G., & Wilmont, I. (2021). {R}egel{S}praak: a {CNL} for Executable Tax Rules Specification. *Proceedings of the Seventh International Workshop on Controlled Natural Language (CNL 2020/21)*. https://aclanthology.org/2021.cnl-1.6

Cowan, A., & Grushka-Cockayne, Y. (2021). Agile Development. *Ssrn*, *50*(3), 79–83.

FATF. (2019). *Virtual Assets and Virtual Asset Service Providers - Guidance for a Risk-Based Approach*. *June*. www.fatf-gafi.org

Gaur, S., Vo, N. H., Kashihara, K., & Baral, C. (2015). Translating simple legal text to formal representations. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9067*, 259–273. https://doi.org/10.1007/978-3-662-48119-6_19

Gill, T. G., & Hevner, A. R. (2013). A fitness-utility model for design science research. *ACM Transactions on Management Information Systems*, *4*(2). https://doi.org/10.1145/2499962.2499963

Grant, C., & Osanloo, A. (2014). Understanding, Selecting, and Integrating a Theoretical Framework in Dissertation Research: Creating the Blueprint for Your "House." *Administrative Issues Journal Education Practice and Research*, *4*(2), 12–26. https://doi.org/10.5929/2014.4.2.9

Gregor, S., & Hevner, A. R. (2013). P OSITIONING AND P RESENTING D ESIGN S CIENCE Types of Knowledge in Design Science Research. *MIS Quarterly*, *37*(2), 337–355. https://doi.org/10.2753/MIS0742-1222240302

Hardjono, T., Lipton, A., & Pentland, A. (2021). Toward a Public-Key Management Framework for Virtual Assets and Virtual Asset Service Providers. *The Journal of FinTech*, *01*(01), 2050001. https://doi.org/10.1142/s2705109920500017

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75–105.

Joshi, K. P., & Saha, S. (2020). A Semantically Rich Framework for Knowledge Representation of Code of Federal Regulations. *Digital Government: Research and Practice*, *1*(3), 1–17. https://doi.org/10.1145/3425192

Leidraad Wwft en Sw. (2020). *december*.

Libal, T., & Novotná, T. (2021). Towards Transparent Legal Formalization. *Lecture Notes in Computer*

Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12688 LNAI, 296–313. https://doi.org/10.1007/978-3-030-82017-6_18

Moher, T. (2006). Embedded phenomena: Supporting science learning with classroom-sized distributed simulations. *Conference on Human Factors in Computing Systems - Proceedings*, *2*, 691–700.

Musen, M. A., & Team, P. (2016). *The Protégé Project: A Look Back and a Look Forward*. *1*(4), 4–12. https://doi.org/10.1145/2757001.2757003.The

Nerur, S., & Balijepally, V. (2007). Theoretical reflections on agile development methodologies. *Communications of the ACM*, *50*(3), 79–84.

Njonko, P. B. F., Cardey, S., Greenfield, P., & El Abed, W. (2014). RuleCNL: A controlled natural language for business rule specifications. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8625 LNAI*, 66–77. https://doi.org/10.1007/978-3-319-10223-8_7

OMG. (2013). Semantics of Business Vocabulary and Business Rules. *V1.2*, *1.2*(May), 422. http://www.omg.org/spec/SBVR/1.0/

Poskriakov, F., Chiriaeva, M., & Cavin, C. (2020). Cryptocurrency compliance and risks: A European KYC/AML perspective. *Blockchain & Cryptocurrency Regulation*.

Reynares, E., Caliusco, M. L., & Galli, M. R. (2014). SBVR to OWL 2 Mappings: An Automatable and Structural-Rooted Approach. *CLEI Electronic Journal*, *17*(3). https://doi.org/10.19153/cleiej.17.3.2

Rutledge, L., Corbijn, J., Cuijpers, B., & Wondaal, L. (2022). *Rapid Prototyping of Business Rule-based Information Systems with Semantic Web Software*. 1–11.

Rutledge, L., & Italiaander, R. (2021). Toward a Reference Architecture for Traceability in SBVR-based Systems. *Proceedings of the Seventh International Workshop on Controlled Natural Language*, *September*.

Rutledge, L., Mellema, P., & Pietersma, T. (2021). Displaying triple provenance with extensions to the Fresnel vocabulary for semantic browsers. *6th International Workshop on Visualization and Interaction for Ontologies and Linked Data*, 103–114.

Rysin, V., & Rysin, M. (2020). The money laundering risk and regulatory challenges for cryptocurrency markets. *Restructuring Management. Models–Changes–Development*, *December*, 187–202.

Shadish, W. R., & Cook, T. D. (2002). EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGNS FOR GENERALIZED CAUSAL INFERENCE. *Handbook of Industrial and Organizational Psychology*, *223*(0), 623. http://pubs.amstat.org/doi/pdfplus/10.1198/jasa.2005.s23

Sierhuis, M. (2013). Theory, Practice, and Applications of Rules on the Web. In *Multi-agent Activity Modeling with the Brahms Environment (Abstract of Tutorial)* (Vol. 8035, Issue July). http://dblp.uni-trier.de/db/conf/ruleml/ruleml2013.html#GovernatoriORSC13

Straub, D., & Ang, S. (2011). Rigor and relevance in IS research: Redefining the debate and a call for future research. *MIS Quarterly: Management Information Systems*, *35*(1). https://doi.org/10.2307/23043485

Thuan, N. H., Drechsler, A., & Antunes, P. (2019). Construction of design science research questions. *Communications of the Association for Information Systems*, *44*(1), 332–363. https://doi.org/10.17705/1CAIS.04420

van Aken, J., Chandrasekaran, A., & Halman, J. (2016). Conducting and publishing design science research: Inaugural essay of the design science department of the Journal of Operations Management. *Journal of Operations Management*, *47–48*, 1–8. https://doi.org/10.1016/j.jom.2016.06.004

Venable, J. R. (2006). The Role of Theory and Theorising in Design Science Research. *Proceedings of the 1st International Conference on Design Science in Information Systems and Technology (DESRIST 2006)*, 1–18.

Wakeling, S., Spezi, V., Fry, J., Creaser, C., Pinfield, S., & Willett, P. (2019). Academic communities:

The role of journals and open-access mega-journals in scholarly communication. *Journal of Documentation*, *75*(1), 120–139. https://doi.org/10.1108/JD-05-2018-0067

Wieringa, R. (2014). Empirical research methods for technology validation: Scaling up to practice. *Journal of Systems and Software*, *95*, 19–31. https://doi.org/10.1016/j.jss.2013.11.1097

Wieringa, R., Condori-Fernandez, N., Daneva, M., Mutschler, B., & Pastor, O. (2012). Lessons learned from evaluating a checklist for reporting experimental and observational research. *International Symposium on Empirical Software Engineering and Measurement*, 157–160. https://doi.org/10.1145/2372251.2372279

Wieringa, R. J. (2014). *Design science methodology for information systems and software engineering.* Springer International Publishing.