Goal and Risk Analysis in the Development of Information Systems for the Web of Data

Marcio Louzada de Freitas, Allan Araujo Silva, Renata S. S. Guizzardi, and Vítor E. Silva Souza

Ontology and Conceptual Modeling Research Group (NEMO)
Department of Computer Science, Federal University of Espírito Santo (UFES), Brazil
marciolfreitas@outlook.com, allan.lp001@gmail.com,
{rguizzardi,vitorsouza}@inf.ufes.br

Abstract. The publication of Linked Data on the World Wide Web, regarding several application domains, leads to new problems related to requirements engineering, which needs to take into account aspects related to new ways of developing systems and delivering information integrated with the Web of Data. The use of vocabularies is an intrinsic activity when publishing or consuming Linked Data and their choice can be supported by the elicited requirements and domain ontology. In this context, we propose the Goals and Risks Analysis for Linked Data (GRALD), an approach for modeling goals and risks for information systems for the Web of Data.

Keywords: Goal Modeling, Risk Modeling, Linked Data, Vocabularies

1 Introduction

The Semantic Web was presented by Berners-Lee et al. [2] as the Web version that seeks to make content understandable by both humans and machines, improve search engines by giving meaning to published content and take into account contextual information of time, space and states of things. A challenge of the Semantic Web is to ensure the expressiveness and generate inference of the published content, without losing performance in the representation of data on the Web [2].

At the core of the Semantic Web idea is the concept of *Linked Data*. Bizer et al. [3] define Linked Data as a set of data interconnected by *Uniform Resource Identifiers* (URIs) whose contents can be processed by machines, forming a Web of Data. The published content is based on the *Resource Description Framework* (RDF, https://www.w3.org/RDF/) and data can be extracted using *SPARQL* queries (https://www.w3.org/TR/rdf-sparql-query/). Metadata (e.g., Microformats, RDFa) can be used to annotate Web pages with semantic content [17].

According to Heath & Bizer [9], data published in Linked Data refer to several domains, such as Geographic, Media, Social Media, Governmental, Libraries and Education, Life Sciences and so on. With the adoption and implementation of Linked Data in several areas of knowledge by companies, institutions and governments, it becomes necessary to analyze goals and requirements, as well as to identify and analyze the risks of adopting Linked Data in Web-based Information Systems. Given that the Web follows an open and decentralized architecture, connecting an information system with external data sources can lead to potential risks, thus the need to understand their impact on stakeholder goals.

This paper proposes *Goal and Risk Analysis for Linked Data* (*GRALD*), an approach that applies Goal-Oriented Requirements Engineering (GORE) and risk analysis techniques for the development of Information Systems for the Web of Data. *GRALD* is based on the *RISCOSS* approach [6] which seeks to align business goals and risks in the adoption of open source software, modeling risks with the *RiskML* language. The modeling of goals is done with *iStar* [7], aiming to understand the social domain to enable requirements engineering, defining social concepts. GRALD is integrated with our previous work, *FrameWeb-LD* [5], a method for building Web-based Information Systems that publish Linked Data.

GRALD is motivated by the growing publication of Linked Data in various domains, in which goal modeling and risks analysis can be applied. Tasks such as requirements elicitation, creation of a domain ontology, and modeling of system goals can also help in the choice of vocabulary for Linked Data publication. The objective of this paper is to demonstrate the applicability of goal and risk analysis in the development of linked data systems, and to assist in the process of choosing vocabularies.

The remainder of the paper is divided as follows: Section 2 summarizes the baseline of our work; Section 3 presents GRALD; Section 4 describes its evaluation; Section 5 discusses related work; and Section 6 concludes.

2 Baseline

Goal and Risk Analysis for Linked Data (*GRALD*) is based on two existing approaches: RISCOSS and FrameWeb-LD. We chose RISCOSS because it uses two different languages for modeling goals (iStar) and risks (RiskML), which allows one to study how the same risks may affect different strategies or ecosystems [6]. RISCOSS extends the goal analysis support in iStar, allowing us to analyze how risks are propagated in the goal graph. Frameweb-LD was chosen because it is focused on the publication of Linked Data in Web-based Information Systems. Through GRALD, we seek synergy between these two approaches.

With some effort, other approaches related to risks and goals could be adapted to use in GRALD, e.g., the GR Framework [1] for modeling and reasoning about risks during requirements analysis or the KAOS language [12] for goal modeling and obstacle analysis. This is, however, out of the scope of this paper.

2.1 The RISCOSS Approach

The *RISCOSS* project [6] (http://www.riscoss.eu) aims at analyzing risks in the adoption of *Open Source Software* (*OSS*) by organizations, modeling their goals using iStar [15].

iStar modeling seeks to understand social concepts and applies them in systems engineering processes. The central concept is the *actor*, which can be human beings, organizations, hardware, software or a combination thereof. The actor is able to act independently, has autonomy, intention to perform an action and her behavior is not totally controllable. Other concepts such as *tasks*, *resources*, *goal*, *softgoal*, *agent*, *roles*, etc. are part of this approach [7].

In RISCOSS, the modeling of risks is done using *RiskML* [6], a language that uses primitive concepts like *Goal* (something of interest for a stakeholder to obtain or maintain), *Event* (the occurrence of something that may undermine the objectives), *Situation* (circumstances where risks are likely to occur), and *Indicators* of risks (existing data measurements approved by experts [14], which can be simple or composite). The impact relationship between an event E and a goal G indicates how the occurrence of E impacts the satisfaction of G.

The attributes of RiskML's primitive constructs are *Likelihood* (Event), *Significance* (Event), *Exposure* (Event), *Satisfaction* (Situation) and *Satisfiability* (Goal); and relationships are *indicate* (between Indicator and Situation); *expose*, *protect*, *increase*, *reduce* (between Situation and Event); *expose* and *protect* (between Events); and *impact* (between Event and Goal). For more detail, see the integrated Goal and *RiskML* metamodel available in [14].

2.2 The FrameWeb-LD Approach

FrameWeb-LD [5] is an approach for building Web-based Information Systems (WIS) that publish Linked Data. It proposes a process divided in five stages: Analysis, Design, Implementation, Testing and Deployment. The main contributions of this approach are an extension of *FrameWeb*'s metamodel [16] allowing Linked Data mappings to be represented in its design models and a tool for code generation to assist developers in publishing Linked Data.

FrameWeb proposes the creation of an *Entity Model* at the architectural design phase, based on the conceptual models/ontologies built in the preceding Requirement Engineering phases, in order to represent domain classes and their integration to frameworks that are commonly used in the development of WISs [16]. *FrameWeb-LD* adds annotations on top of the basic *FrameWeb Entity Model* to specify Linked Data vocabulary mappings [5].

We illustrate this with a running example that will be used throughout the paper: an academic WIS called Marvin (http://dev.nemo.inf.ufes.br:8180/Marvin/) under development in our department at the university. In particular, we focus on a module of Marvin called C2D, which keeps track of members of our postgraduate program and their respective publications for evaluation purposes. Researchers and their publications are registered in the system, venues are then matched to a list of qualified conferences and journals provided by the federal government and, based on this list, each publication is then assigned a score, which is then used to calculate the score of each researcher.

Figure 1 shows the *Entity Model* for C2D, in which UML Classes about researchers, publications, etc. are linked to popular vocabularies, such as FOAF (http://xmlns.com/foaf/spec/) and DBLP (http://dblp.uni-trier.de/). For instance, `Researcher` is equivalent to `dblp:Person`, given that the scope of the DBLP vocabulary is to represent researchers and their publications. *Subclass* relations between vocabulary classes and domain classes can be represented by inheritance, e.g., `Researcher` is subclass of `foaf:Person` (FOAF has a broader scope and represents not only researchers). In the `User` class, the `ld-ignore` stereotype represents that user data will not be published in Linked Data. Note that, due to space constraints, the figure also displays vocabularies which were added to the model during the GRALD process, explained in Section 3.

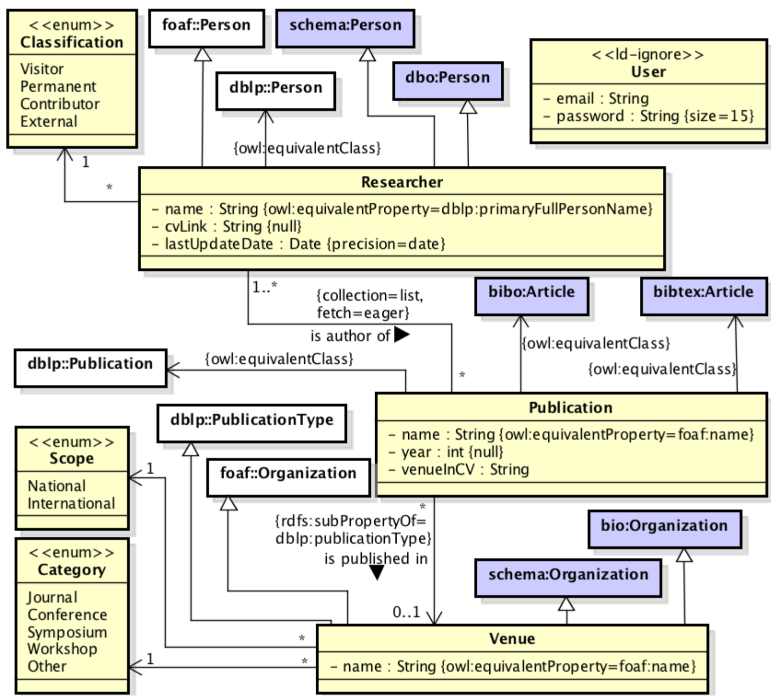

Fig. 1. A FrameWeb-LD Entity Model for C2D [5] with newly added vocabularies using GRALD (dark/blue background).

It is important to note that while FrameWeb-LD allows us to link to external vocabularies, it does not aid developers in finding the most appropriate vocabularies to link. This is very important in the publication of Linked Data, as the objective is to make our data understandable by third party software which has already been programmed to understand some of these popular vocabularies [9]. Linking to unknown vocabularies or to terms that do not properly represent your data can compromise this objective.

In [5], the implementation phase contains three activities: *Encode Operational Ontology in OWL* (which can be automated by tools), *Encode Web Information*

System and *Build Databases*. For the latter, a relational database is created and a Linked Data layer above it is added with the use of D2RQ (http://d2rq.org/) which provides triplestore (a database of RDF triples) features such as a SPARQL endpoint. After Implementation, Test and Deployment phases are carried on, with the deployment of Linked Data done by D2RQ.

3 Proposal

In this section, we present our proposal, named *Goal and Risk Analysis for Linked Data* (*GRALD*). The main contributions of the approach are: (a) modeling of system requirements using a goal-oriented language, with a particular focus on the publication of Linked Data; (b) creation of risks models to support the analysis of risks in the publication of Linked Data; and (c) searching of vocabularies according to the elicited goals and conceptual models of the domain.

An overview of the development process proposed by *GRALD* is presented in Figure 2. The process is divided in three stages (the names of the roles defined in each swimlane). Blue rectangles (light background) represent activities proposed in FrameWeb-LD [5], presented in Section 2.2, whereas gray rectangles (dark background) represent activities proposed by *GRALD* to meet the above contributions. Arrows represent the sequences of activities. The phases of the process are detailed in the following subsections.

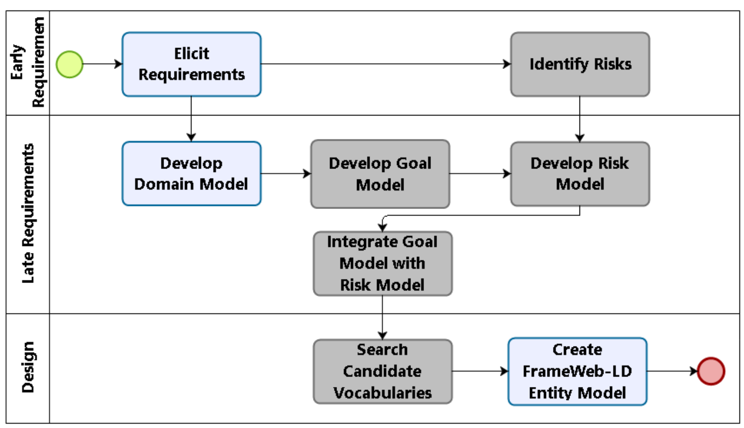

Fig. 2. Overview of the GRALD process.

In our proposal we apply these approaches in a unified way, performing goal and risk modeling with iStar and RiskML (RISCOSS), respectively, for the publication of Linked Data with FrameWeb-LD. Thus, we seek synergy between these approaches to aid in the choice of Linked Data vocabulary to be used by a Web-based Information System, understanding the risks involved in the publication and integration of Linked Data.

3.1 Early Requirements

The first phase starts with the *Elicit Requirements* activity, in which functional and non-functional requirements are captured. These requirements are used to elaborate iStar and related models in the next phase.

Next, the activity *Identify Risks* follows. In RISCOSS [13], risk management is based on a three-layered strategy, the layers cover the gathering of data. In layer I, data about risks is collected from OSS communities, projects and experts that determine the risks drivers; in layer II, risk indicators and models are defined; and in layer III, the risk model is linked with the goal models to represent the impact that the possible risk events have on strategic and business goals. We adapt this strategy to the case of Linked Data publication, collecting data about risks from the bibliography and Linked Data community websites.

According to the W3C [10], best practices for publishing Linked Data should be considered, such as choice of dataset; URI creation; choice and creation of vocabulary; choice of an appropriate license for the publication of content; among others. The adoption of these best practices helps prevent risks and, besides, starting from them we can identify possible risks related to the publication of Linked Data in our projects.

According to [4], traditional Web risks are extended to the Semantic Web, such as *SPARQL* and *SPARUL* injections, etc. Risks related to the creation and maintenance of ontologies and trust and proof of information are also addressed. In our case, we try to consider these risks for systems publishing Linked Data.

3.2 Late Requirements

The second phase of *GRALD* starts with the *Develop Domain Model* activity. Based on elicited requirements, conceptual models/ontologies representing elements from the domain of discourse of our system are built following the guidelines of FrameWeb-LD [5]. In our running example, for instance, researchers, publications, venues, score, etc. are examples of domain elements.

Next, we *Develop Goal Model*. The purpose of this activity is to model the goals (requirements) of the system, with a particular focus on publication of Linked Data. Figure 3 shows the goal model for our running example, C2D.

The actor *C2D* represents the system itself, deployed and maintained in our university. The actor Community represents the academic community, composed by students, professors (researchers), staff, etc. As such, the community has the goals *Data obtained for E-learning*, *Data obtained for academic research* and *Data obtained for curriculum databases*, accomplished by the task *Search information in linked data* and depends on C2D for its publication and availability.

The central goal for C2D, therefore, is *Data Published in Linked Data*, divided in subgoals, according to the data that will be published: *Scores*, *Venues*, *Publications* and *Researchers*; and the goal *Use W3C standards to encode data*. The goal *Users not published in linked data* represents the fact that user data should not be published. The data is registered in the system by the tasks *Calculate researcher score*, *Manage venues*, *Manage publications*, *Manage researchers* and *Manage and authenticate users*.

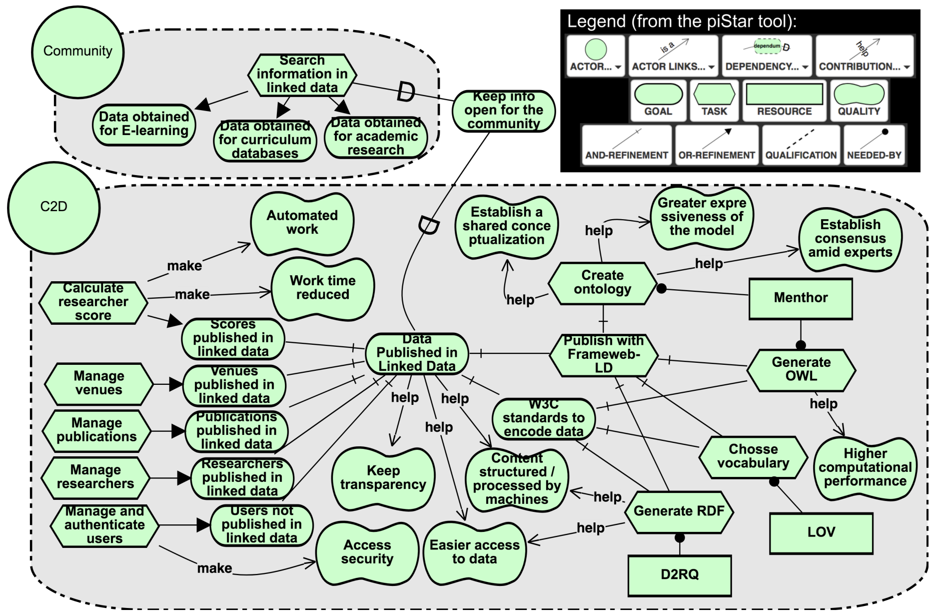

Fig. 3. iStar goal model for C2D, built with http://www.cin.ufpe.br/~jhcp/pistar/.

About the *qualities* of the system, the main goal *Data Published in Linked Data* helps C2D to *Keep transparency* because the data on researcher accreditation are open for the community to search; *Content structured and processed by machines* and *Easier access to data* are helped because the data is published in RDF format, allowing the possibility of a computational agent to process it. The task *Calculate researcher score* makes the *Automated work* and *Work time reduced* sofgoals, as it replaces the manual calculation of the scores. The qualities *Establish a shared conceptualization*, *Greater expressiveness of the model* and *Establish consensus amid experts* are helped by the task *Create ontology*. The task *Generate OWL* helps the quality *Higher computational performance*. Finally, the task *Manage and authenticate users* makes *Access security*.

Implementation is performed according to the FrameWeb-LD approach and the tasks that are necessary are: *Create ontology*, *Generate OWL*, *Chosse vocabulary* and *Generate RDF*. Resources being used, in this case, are: the *Menthor* editor (http://www.menthor.net/), *LOV (Linked Open Vocabularies)* (http://lov.okfn.org) and *D2RQ* respectively. Although these tasks are about the same in different projects (that use FrameWeb-LD), the resources used to accomplish them can vary from one project to another.

Once the goal model is produced, we move on to the activity *Develop Risk Model*. Based on the risk situations captured during *Early Requirements*, a risk model is created with the RiskML language, in order to capture the impact of the occurrence of risk events on goals. Figure 4 shows the risk model produced for C2D, related to the choice of Linked Data vocabulary.

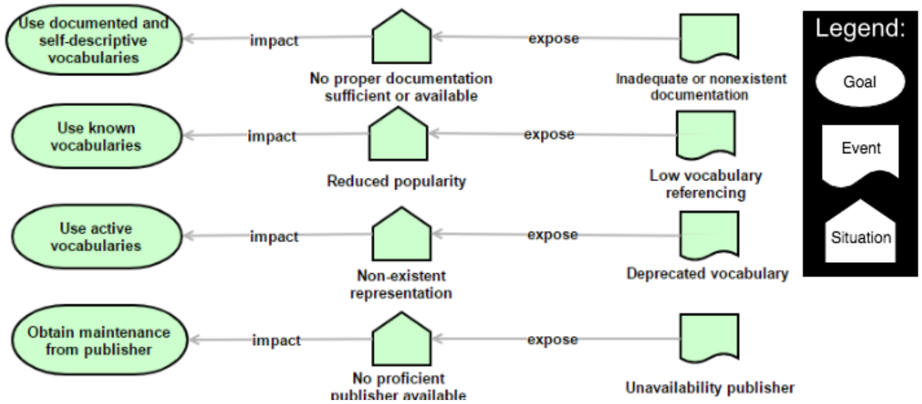

Fig. 4. RiskML model for C2D w.r.t. vocabulary adoption.

In this model, we can capture new goals related to the choice of vocabulary that are impacted by risk events. For instance, the goal *Use documented and self-descriptive vocabularies* is impacted by risk event *No proper documentation sufficient or available* exposed by risk situation *Inadequate or nonexistent documentation*. Other risks related to the creation and maintenance of vocabulary and ontology, dataset selection, trust and proof of information, publishing data and traditional Web risks can be modeled in separate models, not shown here.

Again based on RISCOSS, the last activity of this phase is *Integrate Goal Model with Risk Model*, aligning goals and risks. To this end, goals that were elicited during the construction of the RiskML model are added to the iStar model and are associated with existing goal model elements. At this point, elements from both models can be maintained, added or discarded in order to produce an integrated model.

Once the models are integrated, risk analysis can be performed as per [6]. The impact relation between a risk and a goal represents a negative effect when the event is likely and significant, increasing the evidence that the goal is not achieved. Such evidence is then propagated through the goal graph calculating, for each intentional element, if it is totally/partially satisfied/denied. We are then able to see how risks affect the strategic/high-level goals of each of the involved actors and prioritize our risk mitigation efforts based on this analysis.

In the case of C2D, elements from Figure 3 (goal model) and Figure 4 (risk model) were aligned, producing the integrated model shown in Figure 5. Due to space constraints, the figure concentrates on goals related to the publication of Linked Data only. We can thus see how risk propagation can impact the main goal *Data Published in Linked Data*, which interests the actors C2D and Community. The risk event *Non-existent representation* can impact the goal *Active Vocabularies Used* and, in this case, other vocabularies will have to be chosen, incurring in system maintenance. Further, the risk event *Reduced Popularity* impacts the goal *Know vocabularies used* if the adopted vocabulary is not well

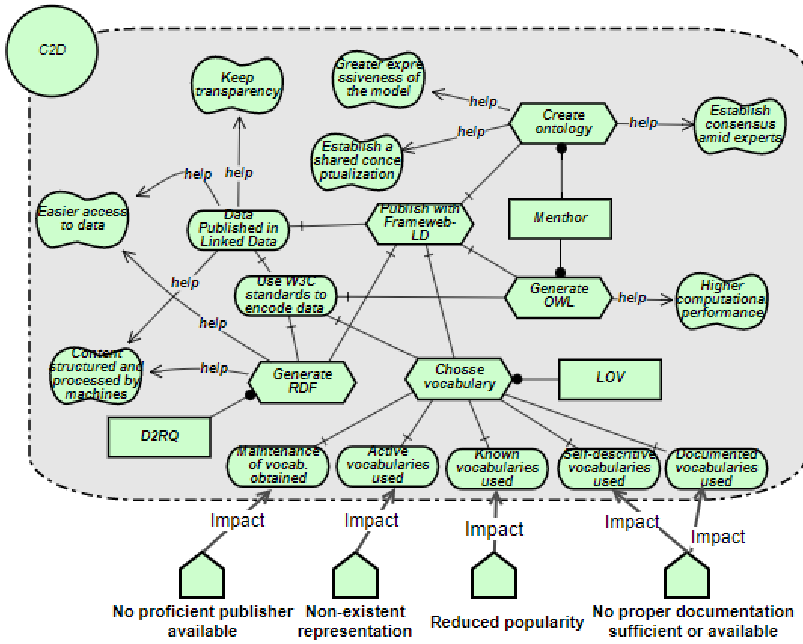

Fig. 5. Goal model for Linked Data implementation connected to RiskML risk events.

referenced by other datasets in the Linked Data community. The total or partial dissatisfaction of the goals *Use documented and self-descriptive vocabularies* and *Obtain maintenance from publisher* can hinder the process of vocabulary adoption and maintenance. In these cases, a benefit of risk analysis is to help in the choice of vocabularies.

3.3 Design

Based on tasks *Elicit Requirements*, *Develop Domain Model* and *Develop Goal Model*, design begins with the *Search for Candidate Vocabularies* for Linked Data publication. The activities in the previous phases help identify existing classes and relations and, based on them, we can search for vocabularies. For this activity, Hyland et al. [10] suggest Linked Data search engines such as Linked Open Vocabularies (LOV), Watson (http://watson.kmi.open.ac.uk/WatsonWUI/), Prefix.cc (http://prefix.cc/), Swoogle (http://swoogle.umbc.edu/2006/), Bioportal (biological domain, http://bioportal.bioontology.org/), etc.

According to [10], in the process of choosing a vocabulary we must take into account if the vocabularies are published by a trusted group or organization, if the vocabularies have permanent URIs, confirm a version control policy, choose documented vocabularies, choose self descriptive vocabularies, choose vocabularies described in more than one language, choose vocabularies used by other data sets and choose vocabularies that are available for access for a long or infinite time. These recommendation form a checklist developers should go through in order to determine the quality of each candidate vocabulary.

Regarding the choice of vocabularies, for our running example, we used with more emphasis the LOV search engine. To search for vocabulary classes for the *Researcher*, *Publication* and *Venue* domain classes, we searched the categories (tags) *People*, *Catalogs* and *Academy*. Analyzing results using the aforementioned recommendation checklist resulted in the choice of new vocabularies for C2D, namely *Schema.org*, *DBPedia*, *Bio*, *Bibtex* and *Bibo*. Analyzing links between vocabularies also helped in the discovery of new vocabularies to consider.

The checklist used in this process is shown in Table 1. Vocabulary attributes are presented in different rows, whereas the columns indicate if the vocabularies being checked meet the criteria (represented by a checkmark: ✓), do not meet the criteria (represented by an ×), or partially meet the criteria (represented by a plus/minus sign: ±). For each attribute, the data presented by LOV was analyzed, as well as the vocabularies' own documentation and their OWL schema.

Table 1. Vocabulary checklist for C2D.

#	**Attributes**	**Dbo**	**Schema**	**Bibo**	**Bio**	**Bibtex**
1	Published by a trusted organization	✓	✓	✓	✓	✓
2	Have permanent URIs	✓	✓	✓	✓	✓
3	Version control policy	✓	✓	✓	✓	✓
4	Documented vocabularies	✓	✓	✓	✓	✓
5	Self descriptive vocabularies	✓	✓	✓	✓	✓
6	Described in more than one language	✓	×	×	×	×
7	Used by other data sets	✓	✓	✓	✓	±
8	Available for access for a long time	✓	✓	✓	✓	✓

Linked Data search engines provide vocabulary information such as label, URI, namespace, description, creator, publisher, comment and language. Also, information such as vocabulary version history is important to measure the reliability of vocabulary regarding the level of updates that may represent the addition of new classes, properties and deprecated classes. Incoming links represents the popularity of vocabulary because it means that other projects are referencing it. Below, we further describe how each item of the checklist can be verified:

- Item 1: check if there is at least one creator, URI and namespace;
- Item 2: check if the URI is fixed and stable;
- Item 3: check if the vocabulary uses any sort of versioning system, e.g., are there previous versions with different numbering?
- Item 4: check if there is a website with documentation;
- Item 5: check the vocabulary OWL schema for triples that describe its classes and properties (e.g., comments or labels);
- Item 6: check the vocabulary OWL schema for strings in more than one language (in our example, Dbo was the only vocabulary that met this criterion);

– Item 7: check if the vocabulary has a substantial amount of incoming links (in our example, LOV indicated Bibtex had only a single incoming link, therefore we consider that it partially met this criterion);
– Item 8: check for how long the vocabulary has been maintained and if they are published in a stable domain.

The above checklist is, of course, not exhaustive and could be improved with further vocabularies and/or desired attributes to check, depending on the availability of resources involved in the software development project.

Once the vocabularies are chosen, we move on to *Create FrameWeb-LD Entity Model*. In this activity, we build an Entity Model as proposed by [5], by adding Linked Data mapping annotations to the domain model, based on FrameWeb-LD meta-model, to the vocabulary chosen in the previous activity. Figure 1, shown earlier, represents the model built for C2D, which is based on the model exemplified in [5], with new vocabulary classes added by the process suggested by *GRALD*, which are filled in blue (darker background).

For instance, for the domain class `Researcher` vocabularies `schema:Person` and `dbo:Person` were added; for `Publication`: `bibo:Article` and `bibtex:Article`; and for `Venue`: `schema:Organization` and `bio:Organization`.

4 Evaluation

The evaluation of this proposal was conducted by the first and second authors of this paper using Web-based Information Systems developed by students of the *Web Development & Semantic Web* course of our Postgraduate Program in Computer Science, all of which aim to publish Linked Data. We evaluated our proposal by creating goal and risk models for these systems and searching for vocabularies based on these models. Artifacts are available in a public source code repository: https://github.com/nemo-ufes/FrameWeb-GRALD.

Regarding tools, we used piStar (cf. Figure 3) to create goal models, we built a simple RiskML editor based on their metamodel (c.f. *Tool* folder in the repository) and we used the generic draw.io tool to produce the integrated models. Future work includes expanding the RiskML editor to support all models.

During evaluation, we particularly focused on three research questions about GRALD: **RQ1**: can it be applied to different systems and domains? **RQ2**: can it be applied to identify risks and new related GORE elements? **RQ3**: can it aid in the identification of vocabularies?

We applied GRALD to five different systems: RightPlace (a system that helps people find a place to live according to their preferences), Rural (management of rural properties), Semed (information system for a medical practice), TransparencyPortal (display government data for citizens) and TravelNM (storefront for a travel agency). By applying GRALD on these existing systems, we were able to identify their goals, tasks, resources and actors, and build their goal models. Moreover, we were able to identify risks related to linked data and other risks, producing their risk models.

For instance, in the case of TravelNM (whose goal model can be found under *goal-models/TravelNM/* in the repository), publication of Linked Data about cities and tour packages being offered is represented by the goal *Data published in linked data*. The data is stored in a triplestore using the resource *Stardog*, which is the triplestore chosen for this project. There is an *Agent* that represents websites that provide unified search engines for traveling services, such as Trivago or Expedia, with a goal *Process data published in LD*, processing the Linked Data published by *TravelNM*. Different from the goal model of C2D, the task *Generate RDF* is performed by the resource *DotNet RDF* (http://www.dotnetrdf.org/), the tasks *Create ontology and Generate OWL* can be performed by the resource *Protégé* (http://protege.stanford.edu/) and so on.

TravelNM's risk model is also available at the repository (under *risk-models/TravelNM/*), presenting four new goals and risk events, for example a goal related to the creation of vocabulary *Use cool URI*, that is impacted by a risk event *URI not provided in accordance to best practices* exposed by a risk situation *URI non-compliant with best practices*. The integrated model of risks and goals was also created (see *impact-models (Goals and Risks)* in the repository) according to the guidelines of our proposal, having new goals added to the model with their respective risks events. Regarding vocabulary search, for the class *City*, for instance, we found vocabularies such as `schema:City` and `dbo:City`, which were not originally found by the students when their assignment was produced.

Finally, we analyze the proposed research questions:

RQ1: Can GRALD be applied to different systems and domains? The systems in which GRALD was successfully applied during this evaluation involved many different domains, such as education, geographical, government, medical, etc., which indicates a positive answer to this RQ.

RQ2: Can GRALD be applied to identify risks and new related GORE elements? Applying GRALD to the aforementioned systems, although very simple and small, we were able to elicit and model risk elements, then augment the goal model with new elements (goals) related to these risks. Further risks could be found with the use of risk identification techniques that are out of the scope of this paper.

RQ3: Can GRALD aid in the identification of vocabularies? GRALD activities *Elicit Requirements*, *Develop Domain Model* and *Develop Goal Model* allowed us to model the classes of the system and clearly specify those that will have the published objects in Linked Data. The checklist used during *Design* aided us in the definition of at least two new (i.e., not previously found by the students) links to external vocabularies per class.

5 Related Work

There are many works published on Linked Data, but in our case we are particularly interested in publications that involve requirements elicitation, risks identification, risk modeling and goal modeling for the development of systems that publish or consume Linked Data. In our search, we had difficulty to find specific references related to above subjects, which seems to imply that this is an open area of research. In this context, this section refers to proposals on risk/goal modeling for software in general.

In [8], requirements analysis is performed with the iStar-based Tropos methodology in two phases: *Early Requirements*, which seeks to understand the organizational context where the system can work, and *Late Requirements*, which seeks to define functional (goals) and non-functional (softgoals) requirements for the system-to-be. The authors also propose reasoning with goal models using forward and backward reasoning. In our proposal we have requirements elicitation and risk identification performed in *Early Requirements* and the creation of the models for WIS that use Linked Data in *Late Requirements*.

Kenett et al. [11] propose capturing, filtering, analyzing and reasoning about risks, based on RISCOSS, using a three layered approach to risk management in FLOSS (*Free Libre Open Source Software*) projects. In the first layer, raw data is collected from FLOSS communities and projects; in the second layer risk indicators are defined and models are produced, in which the risks can be linked to the objectives; finally, in the third layer the risks indicators are converted in Business Risks and, linked with iStar, model business goals to see how risks impact them. In our case we are seeking to apply the RISCOSS approach with the focus on Linked Data.

In [18], the software risk management process is performed with steps Risk Identification, Risk Analysis, Plan and Tracking. In our work, we propose risk identification and modeling related to the development of information systems for the Web of Data.

6 Conclusions

In this paper, we presented GRALD, *Goal and Risk Analysis for Linked Data*, an approach based on RISCOSS, which applies Goal-Oriented Requirements Engineering (GORE) for the development of Web-based Information Systems that publish Linked Data, integrating goal models with risk models in order to perform risk analysis.

GORE is applied in order to help developers to analyze their system objectives, as well as the goals and actors related to the implementation of Linked Data, mapping the necessary resources and tasks to accomplish it. Moreover, performing risk analysis helps to analyze the impact of the occurrence of risk events on system/business goals, as well as to carry out the prevention/mitigation of these risks. Finally, GRALD assists developers in the choice of vocabularies based on the tasks performed in the phases of early and late requirements, having the search of such vocabularies accomplished using Linked Data search engines following guidelines from a checklist.

Our research proposal is a work in progress and with some limitations, which we intend to address in future work, such as (i) evaluate the proposal with more systems and practitioners, going through goal-oriented modeling and risk analysis; (ii) evaluate the scalability of our models; (iii) create a repository of risk indicators related to Linked Data; and (iv) develop a tool integrated with Linked Data search engines (e.g., LOV) to assist developers in the task of choosing vocabulary.

References

1. Asnar, Y., Giorgini, P., Mylopoulos, J.: Goal-driven risk assessment in requirements engineering. Requir. Eng. 16(2), 101–116 (Jun 2011)
2. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American 284(5), 34–43 (2001)
3. Bizer, C., Heath, T., Berners-Lee, T.: Linked data-the story so far. Int. J. Semant. Web Inf. Syst. 5(3), 1–22 (2009)
4. Bruwer, R., Rudman, R.: Web 3.0: governance, risks and safeguards. Journal of Applied Business Research 31(3), 1037 (2015)
5. Celino, D.R., Reis, L.V., Martins, B.F., Souza, V.E.S.: A Framework-based Approach for the Integration of Web-based Information Systems on the Semantic Web. In: Proc. of the 22nd Brazilian Symposium on Multimedia and the Web. pp. 231–238. ACM (nov 2016)
6. Costal, D., López, L., Morandini, M., Siena, A., Annosi, M.C., Gross, D., Méndez, L., Franch, X., Susi, A.: Aligning business goals and risks in oss adoption. In: International Conference on Conceptual Modeling. pp. 35–49. Springer (2015)
7. Eric, S.Y.: Social Modeling and i*. In: Conceptual Modeling: Foundations and Applications, pp. 99–121. Springer (2009)
8. Giorgini, P., Mylopoulos, J., Sebastiani, R.: Goal-oriented requirements analysis and reasoning in the tropos methodology. Eng. Appl. Artif. Intell. 18(2), 159–171 (2005)
9. Heath, T., Bizer, C.: Linked Data: Evolving the Web into a Global Data Space. Synthesis Lectures on the Semantic Web: Theory and Technology, Morgan & Claypool Publishers (2011)
10. Hyland, B., Atemezing, G., Villazón-Terrazas, B.: Best Practices for Publishing Linked Data, https://www.w3.org/TR/ld-bp/ (last access: May 26th, 2017) (2014)
11. Kenett, R.S., Franch, X., Susi, A., Galanis, N.: Adoption of free libre open source software (floss): A risk management perspective. In: 2014 IEEE 38th Annual Computer Software and Applications Conference. pp. 171–180 (July 2014)
12. van Lamsweerde, A., Letier, E.: Handling obstacles in goal-oriented requirements engineering. IEEE Trans. Softw. Eng. 26(10), 978–1005 (Oct 2000)
13. López, L.: Welcome to the RISCOSS Risk Modeling repository, https://github.com/RISCOSS/riscoss-risk-modeling/wiki (last access: May 26th, 2017) (2015)
14. López, L., Siena, A.: How to create a Risk Model, https://github.com/RISCOSS/riscoss-risk-modeling/wiki/How-to-create-a-Risk-Model (last access: May 25th, 2017) (2015)
15. López Cuesta, L., Franch Gutiérrez, J.: Applying business strategy models in organizations. In: Proceedings of the 7th International i* Workshop (2014)
16. Martins, B.F., Souza, V.E.S.: A Model-Driven Approach for the Design of Web Information Systems based on Frameworks. In: Proc. of the 21st Brazilian Symposium on Multimedia and the Web. pp. 41–48. ACM (2015)
17. Pohorec, S., Zorman, M., Kokol, P.: Analysis of approaches to structured data on the web. Comput. Stand. Interfaces 36(1), 256–262 (2013)
18. Westfall, L., Road, C.: Software Risk Management. Risk Manag. pp. 1–8 (2001)