If you can't enforce it, contract it: Enforceability in Policy-Driven (Linked) Data Markets

Simon Steyskal*, Sabrina Kirrane
Vienna University of Economics and Business, Vienna, Austria
[firstname.lastname]@wu.ac.at

ABSTRACT

The Web of Data refers to a network of data, which is published from various data sources, distributed across different machines, and possibly interconnected as Linked (Open) Data. We assume that in the near future these machines will not only publish and consume data, but will also perform transactions in digital data markets without human intervention. For these digital data markets to succeed, it is crucial that published data is accessed and used in a manner, which is compliant with restrictions or regulations that have been defined by data publishers. While it is fairly simple to express access policies using one of the numerous vocabularies available, the actual enforcement of those policies is rather difficult especially when taking dependencies among policies into account. In this paper, we demonstrate how ODRL can be used not only to represent access policies but also to specify access requests, offers and agreements, and propose an approach to generate on-the-fly contracts that govern all explicit and implicit non-enforceable policies.

Categories and Subject Descriptors

K.4.4 [**Computers and Society**]: Electronic Commerce - *Security*; K.6.5 [**Management of Computing and Information Systems**]: Security and Protection

General Terms

Data Markets, Policies, Enforceability, Policy-Driven Linked Data Markets, ODRL

1. INTRODUCTION

In recent years, we have seen the emergence of online service providers who trade, potentially one of the most valuable commodities for any business, `data`. The service offering, which is commonly known as a `data market`, caters not only for the buying and selling of raw data, but also offers value-added services derived from this data (e.g. data cleansing, integration, analytics and visualisation). According to a recent survey conducted by the European Research Center for Information Systems (ERCIS) [9], last year there was a slight decrease in the number of service providers offering access to raw data and an increase in the provision of high quality processed data. Here high quality processed data refers to data that is represented in a manner which supports data integration and analytics (i.e. accurate data represented in manner which is interoperable, flexible and extensible). Additionally the survey highlighted that the number of service providers that publish data using the Resource Description Framework (RDF) is significantly less than the number that publish data using the Extensible Markup Language (XML) or Comma-Separated Values (CSV) / Excel Spreadsheets (XLS). Given that interoperability, flexibility, and extensibility are cornerstones of the RDF data model and the fact that the number of Linked Open Data publishers is growing year-on-year, it begs the question *what are the current challenges for Linked Data Markets?*. Although there are a number of challenges with respect to data quality, data lifecycle management and quality of service, in this paper we focus specifically on the challenges that relate to access control and licensing.

Möller and Dodds [7], De Virgilio et al. [2] and Kim et al. [5] all propose systems that can potentially be used to realise the LDM vision, however very little consideration if any is given either to access control or machine readable licensing. A number of authors have looked into using the Open Digital Rights language (ODRL) to specify access constraints and licensing [1, 10, 8, 11], however they do not focus on the question of enforceability nor do they apply their work to LDMs. In order to fill this gap in this paper, we present our vision of a Policy-Driven (Linked) Data Market and discuss how our framework can be used to cater for both enforceable and non-enforceable ODRL policies. Our main contributions can be summarized as follows, we: (i) propose a workflow for PDLDM transactions and demonstrate how ODRL can be used not only to represent access policies but also to specify access requests, data offers and agreements; and (ii) present a framework which can be used to both enforce access restrictions (in the case of enforceable policies) and automatically generate license agreements (in the case of non-enforceable policies). The remainder of the paper is structured as follows:

*Simon Steyskal has been partially funded by the Vienna Science and Technology Fund (WWTF) through project ICT12-015 and by the Austrian Research Promotion Agency (FFG) grant 845638 (SHAPE).

We demonstrate how the ODRL can be used to express a variety of policies in *Section* 2. Our strategy for dealing with non-enforceable policies is presented in *Section* 3. We discuss related work in *Section* 4. Finally, we conclude and outline directions for future work in *Section* 5.

2. EXPRESSING (LINKED) DATA MARKET POLICIES IN ODRL

A Data Market is a platform where data and potentially value-added services derived from the data are bought and sold. Although data markets are not a new concept, with an ever increasing amount of data available (social data, sensor data, open data) and advances in information technology we are seeing more and more online marketplaces appear [9]. Data consumers can benefit from the high quality data, that is aggregated and presented in a consistent format, making it easier for then to find and use the data they require. On the other hand, data produces can outsource the cleansing, hosting and discoverability of their data. While, both parties can take advantage of value added services such as integration and analytics.

A Linked Data Market is a specific type of marketplace, which is built on top of the Linked Data Web (LDW) and adheres to the Linked Data principles. In this paper, we propose a Policy-Driven (Linked) Data Market (PDLDM) where data requests, data offers, access policies and agreements are encoded in machine readable policies. The various transactions required for contract negotiation are represented using the workflow illustrated in *Figure* 1, which consists of four major steps:

1. **Make a request.** A data transaction is initiated when a data consumer issues a request to the data market, which is subsequently forwarded to one or more data providers who can potentially service the request.

2. **Check applicable policies.** On receipt of the request the data provider retrieves the relevant access policies (relevance is determined based on the data requested and the credentials supplied by the data consumer).

3. **Compose and offer contract.** The data provider generates a machine readable contract (known as an offer), based on the explicit and implicit non-enforceable actions that are associated with the request. The auto-generated contract is subsequently offered to the data consumer.

4. **Accept contract.** If the data consumer agrees to the terms of the contract, an agreement between the data consumer and the data publisher is generated and persisted for accountability and compliance purposes.

The Open Digital Rights Language (ODRL) [4] is a comprehensive policy expression language that is suitable for expressing fine-grained access restrictions, access policies, as well as licensing information for Linked Data as shown in [1, 10].

An *ODRL Policy* is composed of a set of *ODRL Rules* and an *ODRL Conflict Resolution Strategy*, which is used by the enforcement mechanism to ensure that when conflicts among rules occur the system either grants access, denies access or generates an error. In the sample policies that follow we use an `odrl` prefix for `<http://w3.org/ns/odrl/2/>` and an `ex` prefix for `<http://www.example.com/>`. *Listing* 1 demonstrates how ODRL can be used to specify two policies, one that prohibits ex:provider1 to aggregate data from ex:dataset1 and another that permits ex:provider1 to read data from ex:dataset1.

Listing 1: Policy governing access to ex:dataset1

```
ex:storedPolicy1 a odrl:Set ;
        odrl:prohibition [ a odrl:Prohibition ;
                odrl:assigner ex:provider1 ;
                odrl:target ex:dataset1 ;
                odrl:action odrl:aggregate ] ;
ex:storedPolicy2 a odrl:Set ;
        odrl:permission [ a odrl:Permission ;
                odrl:assigner ex:provider1 ;
                odrl:target ex:dataset1 ;
                odrl:action odrl:read ] .
```

2.1 Selected ODRL Policy Types

In this paper, we go beyond simple access control policies and licenses and demonstrate how ODRL can be used to represent access requests, data offers and agreements. Although all types of policies share the same general structure (i.e. they all consist of a set of rules and a conflict resolution strategy) they differ in terms of their scope.

ODRL Request Policies contain rules that represent the terms of usage sought by a data consumer. The policy defined in *Listing* 2 can be used to specify that ex:consumer1 requests read access to ex:dataset1.

Listing 2: Request read access to ex:dataset1

```
ex:request a odrl:Request ;
        odrl:permission [ a odrl:Permission ;
                odrl:assignee ex:consumer1 ;
                odrl:target ex:dataset1 ;
                odrl:action odrl:read ] .
```

ODRL Offer Policies contain rules that propose terms of usage to data consumers. The policy defined in *Listing* 3 offers ex:consumer1 read access to ex:dataset1 if they agree to a contract that prohibits them from aggregating the data.

Listing 3: Offer a contract for ex:dataset1

```
ex:offer a odrl:Offer ;
        odrl:prohibition [ a odrl:Prohibition ;
                odrl:assigner ex:provider1 ;
                odrl:assignee ex:consumer1 ;
                odrl:target ex:dataset1 ;
                odrl:action odrl:aggregate ] .
```

ODRL Agreement Policies represent contracts between data producers and consumers that stipulate all terms of usage. The policy defined in *Listing* 4 states that ex:consumer1 has agreed to a contract that prohibits them from aggregating the data from ex:dataset1.

Listing 4: Construct an agreement for ex:dataset1

```
ex:agreement a odrl:Agreement ;
```

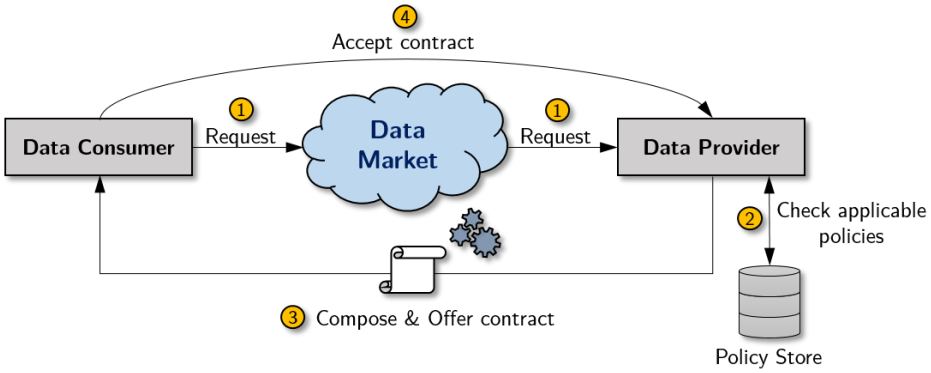

Figure 1: PDLDM workflow

```
odrl:prohibition [ a odrl:Prohibition ;
      odrl:assigner ex:provider1 ;
      odrl:assignee ex:consumer1 ;
      odrl:target ex:dataset1 ;
      odrl:action odrl:aggregrate ] ;
odrl:permission [ a odrl:Permission ;
      odrl:assigner ex:provider1 ;
      odrl:assignee ex:consumer1 ;
      odrl:target ex:dataset1 ;
      odrl:action odrl:read ] .
```

3. ENFORCING ODRL POLICIES

Not only in PDLDMs but also in other domains, policies and especially licenses are widely used to stipulate terms of usage for assets. From a data producer perspective, governance and ensuring compliance with non-enforceable policies is difficult and can result in litigation, which can be a lengthy and expensive processes. As such when it comes to PDLDMs, it is necessary to make the distinction between enforceable and non-enforceable policies and to propose a framework that is capable of handling both. Another consideration is the fact that data consumers might be less eager to conduct business with data providers that offer complex and verbose contracts (even if they able to comply with the verbose policies), as opposed to data providers that keep their contracts as concise as possible. As such, we propose an access control strategy, which on receipt of a request verifies that the requested access is allowed and auto-generates contracts for non-enforceable policies that are as concise as possible (i.e. minimal contracts).

3.1 Enforceability of ODRL Policies

A policy is enforceable if restrictions on actions defined in the policy can actually be controlled by a system. In the context of ODRL we define an *ODRL Action* to be controllable if its execution is permitted, or in the case where its execution is prohibited compliance with the prohibition can be controlled by the party who assigned the policy. Thus, a policy is defined to be enforceable, if all actions it aims to prohibit are not part of the set of uncontrollable actions.

3.2 Composition of Minimal Contracts

We propose an algorithm which auto-generates contracts for non-enforceable policies based on the workflow presented in *Section* 2. A data request which is submitted by a data consumer (Step 1) is matched against a set of stored policies based on the credentials of the requesting party and the actions relating to assets that they request (Step 2). This matching process does not only consider actions explicitly stated in the request but also those which are implicitly related to them and the relevant conflict resolution strategy. A contract which is composed and offered (Step 3) is represented as an *ODRL Offer Policy* and incorporates a set of requested permissions together with the terms of usage that are retrieved from the data provider's stored policies.

Algorithm 1: Minimal Contract Composition Algorithm

Input: A set of applicable ODRL Policies P according to a certain ODRL Request Policy π^R.
Output: A minimal ODRL Offer Policy π^O.

```
1  forall the policies π in P do
2      forall the permission rules δ in π do
3          add δ to the set of permission rules in π^O;
4          add all new uncontrollable actions to the set of
           uncontrollable actions;
5      end
6      forall the prohibition rules δ in π do
7          if prohibited action α is uncontrollable then
8              add δ to the set of prohibition rules in π^O;
9          end
10     end
11 end
```

Algorithm 1 (*minimal contract composition*) denotes the composition procedure that is used to generate minimal *ODRL Offer Policies*. The algorithm takes an *ODRL Request Policy* π^R and a respective set of applicable *ODRL Policies* P retrieved from the policy store as input and iterates over all policies in P.

- For each of the permission rules, the algorithm adds all actions that become uncontrollable once the permission has been granted to the overall set of uncontrollable actions of the policy, and adds the permission rule to the set of permission rules in π^O (line 1-5).
- For each of the prohibition rules, the algorithm checks whether the rule prohibits an action that is defined to be uncontrollable (line 6-7). If that is the case, the respective prohibition rule is added to the *ODRL Offer Policy* π^O (line 8).

The final *ODRL Offer Policy* π^O now consists of all requested permissions a data provider is able to grant as well as all non-enforceable prohibitions that are consequences of these permissions. The *ODRL Offer Policy* is subsequently offered to the data consumer that initiated the transaction (Step 4). If the data consumer agrees to the terms of the contract (i.e. accepts), an *ODRL Agreement Policy* is generated from the *ODRL Offer Policy* and persisted for accountability and compliance purposes (Step 5).

4. RELATED WORK

Möller and Dodds [7], De Virgilio et al. [2] and Kim et al. [5] all propose systems that can potentially be used to realise the LDM vision. Möller and Dodds [7] describe the Kasabi information marketplace which is built on Linked Data principles. Although data publishers are required to supply licensing metadata, the authors do not detail how access to data is restricted or how licenses are enforced. De Virgilio et al. [2] present Nyaya, a system which can be used to manage different Semantic Web datasets. The authors discuss how their system can support user defined constraints, however no specific consideration is given either to access policies or licenses. Kim et al. [5] present an architecture that can be used to support Linked Open Data as a Service (LODaaS) however, they do not mention either access control or licensing.

When it comes to access control for RDF, broadly speaking researchers have focused on representing existing access control models and standards using semantic technology; proposing new access control models suitable for open, heterogeneous and distributed environments; and devising languages and frameworks that can be used to facilitate access control specification and maintenance. Kirrane et al. [6] provide a comprehensive survey of existing access control proposals for RDF. To date no specific consideration has been given to enforceable versus non-enforceable policies. There has however been a number of digital rights management proposals that use ODRL to model their access control and licensing policies. Guth et al. [3] demonstrate how ODRL can be used to exchange access control information and present a framework, which can be used to enforce access control policies. Cabrio et al. [1] discuss how ODRL can be used to model licenses as opposed to access rights. Rodriguez-Doncel et al. [8] present a legal framework for publishing and consuming Linked Data and provide an overview of the existing vocabularies for rights and licensing represented using RDF. Villata and Gandon [11] present a framework which associates licensing terms with data and auto-generates an aggregated license.

In this paper, we go beyond existing proposals by demonstrating how ODRL can be used to represent not only access policies and licenses, but can also support contract negotiation in the form of data requests, data offers and data agreements. We subsequent present a framework, which is capable of dealing with both enforceable and non-enforceable policies.

5. CONCLUSIONS AND FUTURE WORK

A digital data market is an online marketplace where data and potentially value-added services such as data cleansing, integration, analytics and visualisation are bought and sold. A LDM is a specific type of marketplace, which is built on top of the LDW and adheres to the Linked Data principles. If LDMs are to succeed, it is crucial that data published is accessed and used in a manner, which is compliant with access restrictions and licenses. In this paper, we demonstrated how ODRL can be used to specify auto-generated contracts. We subsequently proposed a framework which can be used to both enforce access restrictions and automatically generate contractual agreements for non-enforceable policies. In future work, we will investigate the various mechanisms that can be used to ensure policy compliance and accountability. We also plan to extend the existing framework to support advanced contract composition and privacy protecting, using a combination of negotiation and reasoning techniques.

References

[1] Elena Cabrio, Alessio Palmero Aprosio, and Serena Villata. These are your rights. In *Proceedings of the 11th Extended Semantic Web Conference (ESWC)*, 2014.

[2] Roberto De Virgilio, Giorgio Orsi, Letizia Tanca, and Riccardo Torlone. Semantic Data Markets: a Flexible Environment for Knowledge Management. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, 2011.

[3] Susanne Guth, Gustaf Neumann, and Mark Strembeck. Experiences with the Enforcement of Access Rights Extracted from ODRL-based Digital Contracts. In *Proceedings of the 3rd ACM Workshop on Digital Rights Management*, DRM '03, 2003.

[4] Renato Iannella, Susanne Guth, Daniel Pähler, and Andreas Kasten. Odrl: Open digital rights language 2.1. W3C ODRL Community Group, 2012. `http://www.w3.org/community/odrl/`.

[5] Seonho Kim, Ivan Berlocher, and Tony Lee. RDF based Linked Open Data Management as a DaaS Platform. 2015.

[6] Sabrina Kirrane, Alessandra Mileo, and Stefan Decker. Access control and the resource description framework: A survey. *Technical Report*, 2015.

[7] Knud Möller and Leigh Dodds. The Kasabi Information Marketplace. In *21nd World Wide Web Conference, Lyon, France*, 2012.

[8] Victor Rodriguez-Doncel, Asunción Gómez-Pérez, and Nandana Mihindukulasooriya. Rights declaration in linked data. In *The 3rd International Workshop on Consuming Linked Data.* 2013.

[9] Florian Stahl, Fabian Schomm, and Gottfried Vossen. The Data Marketplace Survey Revisited. Technical report, Working Papers, ERCIS-European Research Center for Information Systems, 2014.

[10] Simon Steyskal and Axel Polleres. Defining expressive access policies for linked data using the ODRL ontology 2.0. In *Proceedings of the 10th International Conference on Semantic Systems, SEMANTICS 2014*, 2014.

[11] Serena Villata and Fabien Gandon. Licenses compatibility and composition in the web of data. In *The 2nd International Workshop on Consuming Linked Data.* 2012.